doi:10.3969/j.issn.1001-4616.2018.04.015

基于深度学习的行人重识别研究综述

朱　繁,王洪元,张　继

(常州大学信息科学与工程学院,江苏 常州 213164)

[**摘要**]　由于视角、背景、光照条件和相互遮挡等因素的变化,行人重识别是一个具有挑战性的问题. 近年来,许多研究者将深度学习的方法引入到行人重识别研究中,并获得了较好的重识别结果. 本文介绍了基于深度学习的行人重识别的主要研究方法(局部特征学习、距离度量学习、基于视频序列学习和生成对抗网络),并介绍目前常用的用于深度学习的行人重识别数据集(DukeMTMC-reID、CUHK03 和 Market1501)及其存在的问题,同时,对行人重识别提出了自己的理解和观点. 最后指出了未来可能的研究方向.

[**关键词**]　深度学习,行人重识别,局部特征学习,距离度量学习

[**中图分类号**]TP391　[**文献标志码**]A　[**文章编号**]1001-4616(2018)04-0093-09

A Survey of Person Re-identification Based on Deep Learning

Zhu Fan,Wang Hongyuan,Zhang Ji

(School of Information Science and Engineering,Changzhou University,Changzhou 213164,China)

Abstract:Due to changes in perspectives,backgrounds,lighting conditions,and mutual occlusion,person re-identification is still a challenging issue. In recent years,many researchers have introduced deep learning methods into person re-identification research and obtained better re-identification results. This paper introduces the main research methods of person re-identification based on deep learning(local feature learning,distance metric learning,video sequence learning,and generation of confrontation networks),and introduces commonly used person re-identification data sets for deep learning (DukeMTMC-reID,CUHK03,and Market1501) and their existing problems. At the same time,it puts forward their own understanding and viewpoints on person re-identification,and finally points out possible future research directions.

Key words:deep learning,person re-identification,local feature learning,distance metric learning

行人重识别(Person Re-identification)也称行人再识别,是在行人检测的基础上利用计算机视觉技术判断图像或者视频序列中是否存在特定行人的技术,也就是用一个摄像头下的照片去判断其他摄像头下是否再次出现了这个人. 而深度学习是一个复杂的机器学习算法,在学习样本数据的内在规律和表示层次的过程中获得的信息对诸如文字、图像和声音等数据的解释有很大的帮助. 它的最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据.

从计算机视觉的角度来看,重识别最具有挑战的问题是如何正确地匹配在集中的外观变化下同一个人的两张图像,如:(1)目标遮挡导致部分特征的丢失;(2)不同的视觉,光照等条件导致同一目标的特征差异;(3)不同目标衣服颜色近似、特征近似导致区分度下降. 基于这 3 个问题,解决方案主要为:(1)提取更适合表征人体的特征;(2)选择合适的距离度量函数;(3)通过训练的方法进行参数训练或者空间映射使得类内距离更小、类间距离更大.

传统的行人重识别研究方法主要是从特征提取和距离度量学习两个方面进行研究. 由于数据集的不足,以及传统重识别方法的缺陷等问题,越来越多的研究者将深度学习方法应用于行人重识别上. 行人重识别常用的深度学习方法通常有 3 个步骤,即首先在训练集上训练一个分类网络,然后,在网络收敛之后,用它的全连接层的输出作为其特征表达. 最后,对所有的图像特征,计算其欧氏距离,并判断相似性. 本文

收稿日期:2018-08-15.

基金项目:国家自然科学基金(61572085).

通讯联系人:王洪元,博士,教授,研究方向:计算机视觉. E-mail:hywang@ cczu.edu.cn

对具有代表性的方法进行评述,并针对当前在行人重识别中使用的数据集提出一些存在的问题,最后对未来的发展趋势作了总结.

1　主要方法

目前基于深度学习的行人重识别方法根据深层神经网络的使用情况可分为特征学习和距离度量学习两类. 特征学习网络的目标是学习一种鲁棒性和辨别性的特征表示行人图像. 距离度量学习旨在减少包含同一个人的图像的描述符之间的距离. 根据对数据集处理方式的不同,这里列举了基于视频序列学习与生成对抗网络(GAN)两种方法进行阐述.

1.1　基于局部特征学习

基于局部特征的学习主要是解决全局特征难学习、特征提取效率低的问题,常用的方法是局部表示法. 局部表示法通常是通过将人的边界框划分为单元格来计算的,例如将图像分割为横条纹或网格,并在单元格上提取深层特征[1-2]. 这些解决方案是基于人的姿态和人体在包围盒中的空间分布相似的假设. 例如,在实际情况中,会检测到边界框,而不是手工标记,因此人类可能处于不同的位置,或者人类的姿势是不同的. 换句话说,空间分区与人体各部分不协调. 基于局部特征的行人重识别是提取输入图片的局部特征,也就是对于不同部件学习不同的特征,之后再将其进行串联. 对于行人匹配,计算各个部分的表示,然后聚合在相应部分之间计算的相似性. 常用的提取局部特征的思路主要有图片切块、利用骨架关键点定位以及姿态矫正等等.

图片切块是一种很常见的提取局部特征方式. 如图 1 所示,图片被水平等分为若干份之后,被分割好的若干块图像块按照顺序送到一个长短时记忆网络(long short term memory network,LSTM)[2],最后的特征融合了所有图像块的局部特征. 但是这种方式的缺点在于对图像对齐的要求比较高,如果两幅图像没有上下对齐,那么很可能出现头和上身对比的现象,反而使得模型判断错误. 为了避免上述问题,Zhao 等人[3]提出一种新的卷积神经网络,称为主轴网(Spindle Net),它是基于人体区域引导的多阶段特征分解和树结构特征融合策略的一种新型卷积神经网络. 这是卷积神经网络(convolutional neural network,CNN)框架中首次考虑人体结构信息以促进特征学习. 如图 2 所示,Spindle Net 利用了 14 个人体关键点来提取局部特征,直臂、两条腿),再将感兴趣区域和原始图片进入同一个 CNN 网络提取特征,最终得到一个融合全局特征和多个尺度局部特征的行人重识别特征. 该方法基于区域特征,可以更好地表示大量的细节信息,有利于识别具有细微差异的个体.

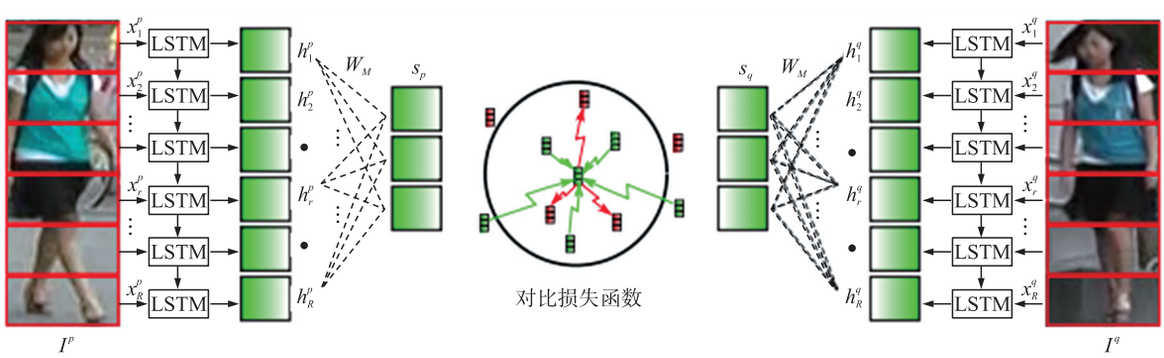

图 1　LSTM 体系结构

Fig. 1　LSTM architecture

Wei 等人[4]提出了一种全局局部对齐描述符(global-local-alignment descriptor,GLAD)来克服姿态变化和错位问题. 与 Spindle Net 类似,GLAD 利用提取的 4 个人体关键点把图片分为 3 个粗略的部分区域(头部、上半身和下半身). 在描述符学习模块中提出了有 4 个子网络组成的 CNN,如图 3 所示,这些子网络共享卷积层并被设计为分别在 3 个部分区域和全局图像上学习描述符. 在训练阶段,共享卷积层可以通过多个不同身体区域的学习任务进行高效优化,以避免过度训练. 在网络训练之后,将 3 个部分的区域和全局图像馈入神经网络中,以提取 4 个描述符,最终连接为 GLAD. 最后提取的特征融合了全局和局部的特征. 该网络利用全局平均池化(global average pooling,GAP)来提取各自的特征. 和 Spindle Net 不同的是,

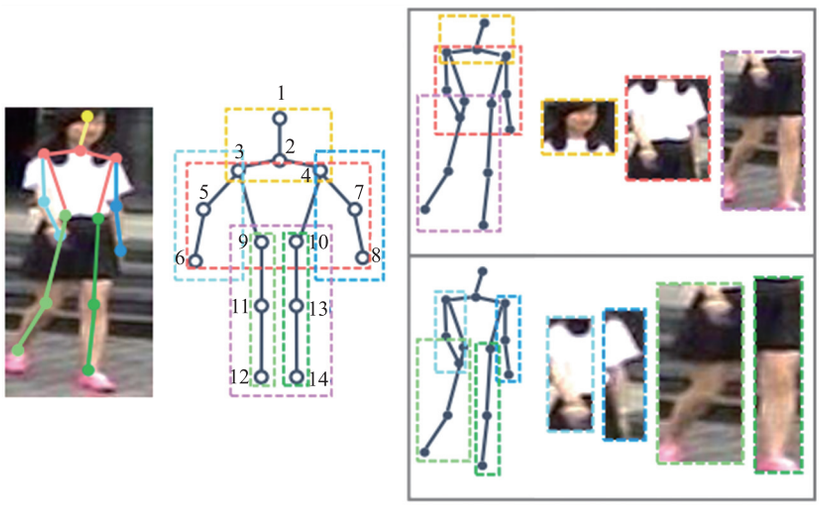

图 2　14 个人体关键点图

Fig. 2　14 Human key map

4 个输入图片各自计算对应的损失,而不是融合为一个特征计算一个总的损失. 该方法较好地利用了人体的局部和全局的信息,既加快了重识别的速度,又不丢失准确度,因此,这项工作有望应用于真实的行人重识别任务场景中.

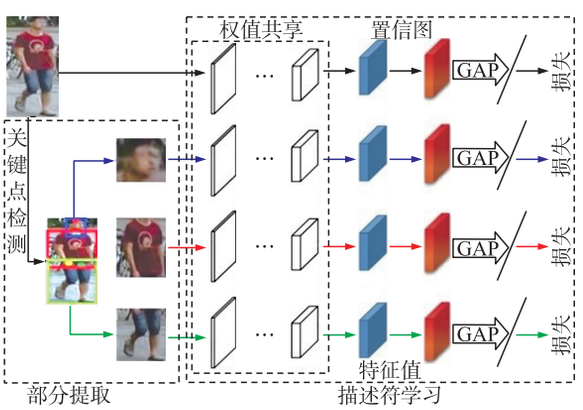

图 3　GLAD 提取框架图

Fig. 3　Framework of GLAD extraction

1.2　基于距离度量学习

距离度量学习旨在减少包含同一个人的图像的描述符之间的距离,一个好的距离度量对于它的成功是至关重要的,因为高维的视觉特征通常不会捕获样本方差下的不变因素. 典型情况下,使用数千个维度的特征表示行人的外观,由于难以收集匹配的训练图像,因此只有数百个训练样本可用. 但训练样本的数量远小于特征维数,导致现有的方法面临经典的小样本量(small sample size,SSS)问题,这不得不求助于维度降低技术或矩阵正则化,但这会导致区分能力的丧失. 距离度量学习的一个关键挑战是小样本容量(small sample size,SSS)问题. 因为样本容量比特征维度小得多(通常是数量级),度量学习方法受到了 SSS 问题的影响,它们在本质上是为了尽量减少类内的方差(距离),同时最大化类间的方差(距离).

文献[5]提出通过在训练数据的一个有区别的零空间中匹配行人来克服重复距离度量学习中的 SSS 问题. 在这个零空间中,如图 4 所示,在一个固定的维度下,同一个人的图像被折叠成一个点,从而最小化了类的离散化,并同时最大化了类间的相对距离.

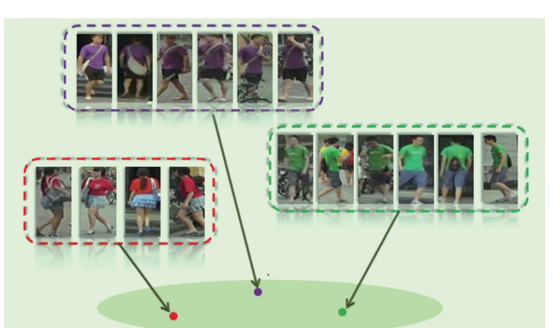

图 4　同身份图像投影到辨别零空间的点

Fig. 4　Training images of same identity are projected to a single point in a learned discriminative null space

度量学习可分类为监督学习与非监督学习,全局学习[6]与局部学习等. 度量学习方法的目的是找到一个从特征空间到另一个距离空间的映射函数[7]. 在深度学习中,常用的度量学习损失方法是三元组损失. 三元组损失函数的思想是要求类内特征距离小于类间特征距离. 如图 5 所示,通过三元组损失(三元:Positive、Negative、Anchor)的学习后使得 Anchor 元和 Positive 元之间的距离最小,而和 Negative 元之间的距离最大.

文献[8]使用改进的三元组损失的执行端到端

的深度度量学习. 三重损耗优化了嵌入空间,使具有相同身份的数据点比具有不同身份的数据点更接近. 文献[9]中提出了一种改进的三元组损失函数,改进后的损失函数如式(1)所示,在原来的三元组损失函数上添加了新的损失函数增强约束,目标样本和正样本之间的距离小于一个阈值 Γ_2,并且阈值 Γ_2 小于 Γ_1. 这个改进的损失函数进一步拉近了同一个行人之间的距离,并拉远了不同行人之间的距离. 改进后的三重损失函数公式为:

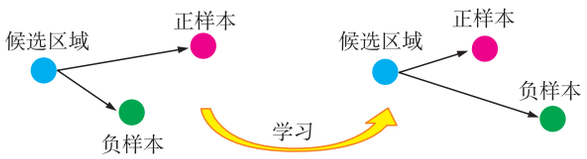

图 5　三元组损失

Fig. 5　Triple loss

$$L(I,w)=\frac{1}{N}\sum_{i=1}^{N}\left(\max\left\{d^n(I_i^0,I_i^+,I_i^-,w),\Gamma_1\right\}+\beta\max\left\{d^p(I_i^0,I_i^+,w),\Gamma_2\right\}\right), \tag{1}$$

$$d^n(I_i^0,I_i^+,I_i^-,w)=d(\phi_w(I_i^0),\phi_w(I_i^+))-d(\phi_w(I_i^0),\phi_w(I_i^-))\leqslant\Gamma_1, \tag{2}$$

$$d^p(I_i^0,I_i^+,w)=d(\phi_w)I_i^0(,\phi_w(I_i^+))\leqslant\Gamma_2, \tag{3}$$

式中,N 是训练样本的数量,β 是平衡类内与类间约束的权重. 距离函数 $d(\cdots)$ 为 $L2$ 范数距离,即:

$$d(\phi_w(I_i^0),\phi_w(I_i^+))=\|\phi_w(I_i^0)-\phi_w(I_i^+)\|^2. \tag{4}$$

文献[10-11]均采用了不同的方式改善了三元组损失函数,其实验结果均有显著的提升. 因此,三元组损失函数可以广泛地应用于基于距离度量学习方面的研究.

1.3　基于视频序列学习

由于单帧图像信息的有限性,基于视频的方法不仅考虑了图像的内容信息,还考虑了帧与帧之间的运动信息等[12-13]. 因为额外的时空信息和更多的外观线索可以用来大大提高匹配性能. 然而目前现有的基于视频的行人重识别方法都等同处理所有帧,忽略了由物体遮挡和运动引起的质量差异,这在实际监控场景中是普遍现象. 基于视频序列的方法主要思想是利用 CNN 来提取空间特征的同时利用递归循环网络(recurrent neural networks,RNN)来提取时序特征. 如图 6 所示,网络输入为图像序列[14]. 每张图像都经过一个共享的 CNN 提取出图像空间内容特征,之后这些特征向量被输入到一个 RNN 网络去提取最终的特征. 最终的特征融合了单帧图像的内容特征和帧与帧之间的运动特征,最后用来训练网络. 该方法最主要的贡献在于第一次将深度学习应用于视频再识别问题,自动学会提取与重识别相关的时空特征这一特点和手工特征的方法相比有很大的区别.

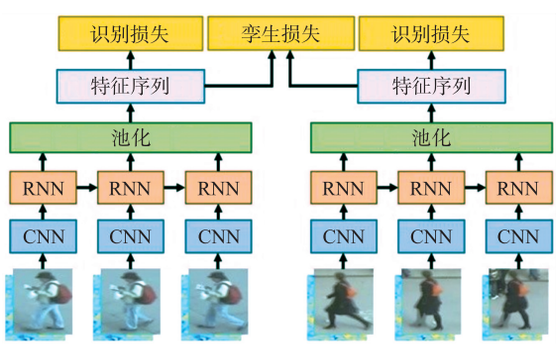

图 6　基于视频的重识别系统

Fig. 6　Video-based re-identification system

与上述方法相比较,Huang 等人[15]提出了一种通过自我步调权重(self paced weighting,SPW)的基于视频的行人重定向方法,在视频序列分割部分,视频序列通过检测 SSM 信号的固定点(图 7 中序列稳定性测量中的峰值点,其中每个点对应于人的局部稳定状态)将其划分为一系列子序列. 然后采用自步态离群算法(SPOD)对子序列噪声度进行评估,子序列产生噪声的可能性越大,距离越远. 最后采用加权多对距离度量学习方法来测量两人图像序列的距离.

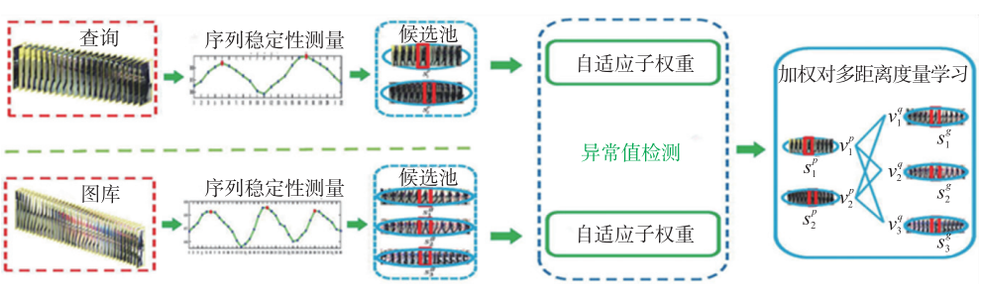

图 7　SPW 方法图

Fig. 7　SPW method diagram

Liu 等人[16]提出了一种累计运动背景网络(accumulative motion context network,AMOC),在 AMOC 体系结构中设计了一个新的运动网络来执行端到端运动上下文信息学习任务. 如图 8 所示,在该体系结构中,两个连续帧之间的空间位置上有两个空间网络(Spat Nets 网络和 Moti Nets 网络)分别从原始视频帧和时间特征中学习空间特征表示. 图像序列的每一帧图像都被输入到 Spat Nets 来提取图像的全局内容特征. 而相邻的两帧将会送到 Moti Nets 来提取光流图特征. 之后空间特征和光流特征融合后输入到一个 RNN 来提取时序特征. 通过 AMOC 网络,每个图像序列都能被提取出一个融合了内容信息、运动信息的特征. 网络采用了分类损失和对比损失来训练模型. 融合了运动信息的序列图像特征能够提高行人重识别的准确度. 其中摄像机 A 网络和摄像机 B 网络的参数是共享的. 为了端到端地训练这个网络,文献[17]采用了包括对比损失和分类损失在内的多任务丢失功能. 对比损失决定两个序列是否描述同一个人,而分类损失预测序列中人的身份.

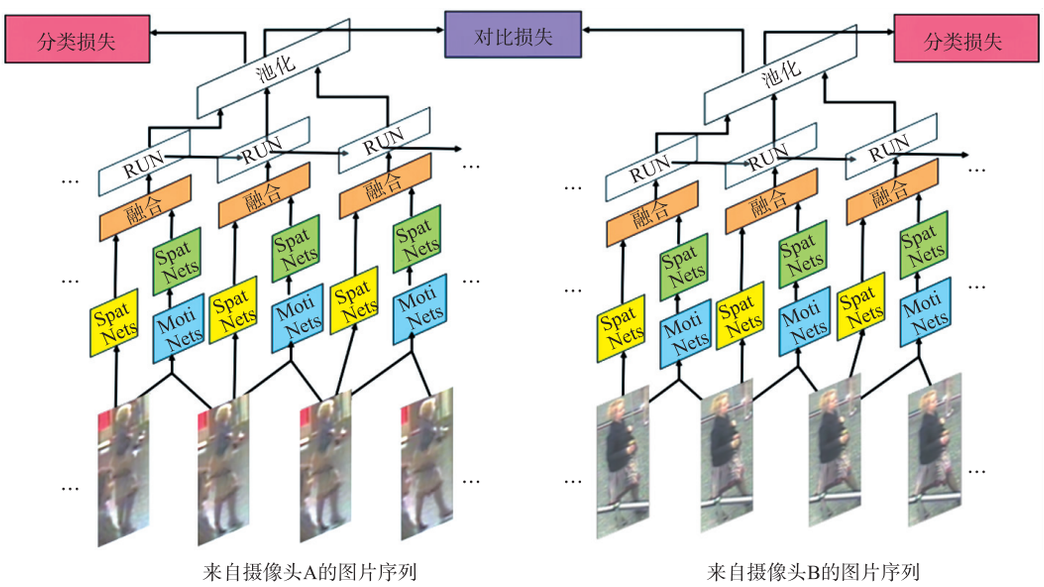

图 8　AMOC 网络结构图

Fig. 8　AMOC network structure

如图 9 所示,AMOC 的空间网络细节图是由 3 个卷积层和 3 个混合层组成,在每个卷积层之后都有一个 tanh 非线性层差值,并且在最后的 max-pool 层的顶部有一个完全连接的层. 底部是 5×5 的立方体的为卷积核,而底部是 2×2 的立方体的则为汇聚内核. Song 等人[17]指出当单帧图像遇到遮挡等情况的时候,可以用多帧的其他信息来弥补,直接诱导网络去对图片进行一个质量判断,从而降低质量差的帧的重要度. 这几种方法均是利用序列中相邻帧之间的遮挡信息来度量其序列的稳定性,有效地利用其稳定性变化实现视频序列的子序列提取,从而使得同一子序列中的图像具有相同的状态.

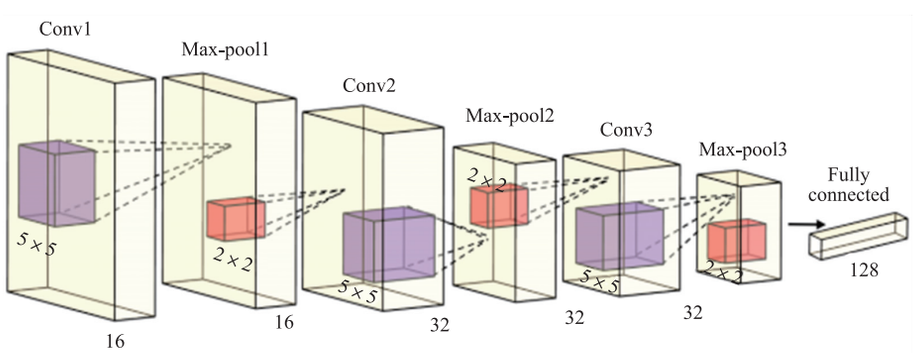

图 9　空间网络细节图

Fig. 9　Space network detail diagram

1.4　生成对抗网络(GAN)

行人重识别的一个非常大的问题就是数据获取困难,Zheng 等人[18]采用深度卷积生成对抗性网络(GAN)作为样本生成,使用了最原始的数据集而不收集额外的数据,提出标签平滑的方法将一个统一的标签分配给无标记的图像,生成的图像作为训练数据加入到训练之中. 但其 GAN 是随机的. Zhong 等人[19]的优势在于 GAN 是可控制的.

ReID 中的一个问题就是不同的摄像头存在着偏差,这个偏差可能来自光线、角度等各个因素. 为了克服这个问题,文献[19]使用 GAN 将一个摄像头的图片转移到另外一个摄像头. 而 GAN 是可控的,也就是说 ID 是明确的. 因此将标签平滑正则化(label smooth regularization,LSR)应用于风格转移的图像上,以使其标签更加柔和. 而另一个问题就是数据集存在偏差,也就是缺少交叉视角配对的训练数据以及在大姿态变化的情况下学习并识别身份敏感和视图不变特征. Wei 等人[20]使用 GAN 把一个数据集的行人迁移到另外一个数据集. GAN 通过设计损失函数实现这个迁移,即一个是前景的绝对误差损失,一个是正常的判别器损失. 判别器损失是用来判断生成的图属于哪个域,前景的损失是为了保证行人前景尽可能逼真不变. 对于重识别中姿态不同的问题,Qian 等[21]人提出了一种新颖的深度重构框架,该框架的关键是一个深度人物头像生成模型,该模型是基于专门针对行人重识别的姿态归一化而设计的 GAN. GAN 造出了一系列标准的姿态图片. 如图 10 所示,事先定义了一组 8 个规范姿势,每一张图片都生成这样标准的 8 个姿势,那么姿势不同的问题就得以解决. 最终用这些图片的特征进行一个平均池化得到最终的特征,最终的特征融合了各个姿势的信息,很好地解决了姿势偏差问题. GAN 的方法好在都是为了从某一个角度解决行人重识别的问题,并且可以利用一张图片生成其他不同的图片,很好地进行了数据集的扩充.

图 10　标准的姿势图

Fig. 10　Standard pose map

1.5　国内外一些其他的深度学习行人重识别方法

He 等人[22]利用完全卷积网络(FCN)来生成特定尺寸的空间特征图,使得像素级特征是一致的. 为了匹配一对不同尺寸的人物图像以及避免明确的对齐,进一步开发了一种称为深空间特征重构(DSR)的新方法. Sun 等人[23]提出用奇异向量分解(SVD)来优化深度表示学习过程,在约束和松弛迭代(RRI)训练方案中,能够迭代地将 CNN 训练中的正交性约束整合起来,产生 SVDNet. Zhong 等人[24]提出了基于 k 阶导数编码的方式,对需要检测的 gallery 中的图片进行重排序,使得识别结果有所提升. Zheng 等人[25]提出随机擦除数据增强的方法,对行人重识别数据集做了不同程度的遮挡,从而提高了网络的泛化能力. 文献[26-28]均采用无监督学习的方式进行学习.

2　数据集

近些年来,随着深度学习发展,数据集的规模越来越大,现有行人重识别数据集都比较大. 常用的数据集如表 1 所示. 目前基于深度学习的行人重识别常用 3 个大型的数据集:DukeMTMC-reID、Market1501、CUHK03. 它们都有超过 1 000 的 ID 和 10 000 个边界框,并且这些数据集都提供了大量的数据来训练深度学习模型.

2.1　DukeMTMC-reID 数据集

数据集 DukeMTMC-reID 图像来自 8 个不同摄像头. 该数据集提供训练集和测试集. 训练集包含 702 个行人的 16 522 张图像,测试集包含剩余 702 个行人的 17 661 张图像的搜索库(gallery). 在测试集中采样了每个 ID 的每个摄像头下的一张图片作为查询图像(query),总共 2 228 个查询图像. 是目前最大的行人重识别数据集,并且该数据集提供了行人属性(性别/长短袖/是否背包等)的标注.

表 1　常用的行人重识别数据集

Table 1　Common used Re-ID datasets

数据集	时间	行人数	相机数	图片数	标签方法
VIPeR	2007	632	2	1 264	Hand
GRID	2009	1 025	8	1 275	Hand
CAVIAR4ReID	2011	72	2	1 220	Hand
3DPeS	2011	192	8	1 011	Hand
PRID2011	2011	934	2	24 541	Hand
CUHK01	2012	971	2	3 884	Hand
CUHK03	2014	1 467	10	13 164	Hand/DPM
Market1501	2015	1 501	6	32 217	Hand/DPM
MARS	2016	1 261	6	1 191 003	DPM+GMMCP
DukeMTMC-reID	2017	1 812	8	36 441	Hand

2.2　Market1501 数据集

数据集 Market-1501 图像来自 6 个不同的摄像头,其中有一个摄像头为低像素. 同时该数据集提供训练集和测试集. 训练集包含 12 936 张图像,测试集包含 19 732 张图像(总共 1 501 个人的 32 668 张图像). 图像由检测器自动检测并切割,包含一些检测误差(接近实际使用情况). 训练数据中一共有 751 人,测试集中有 750 人. 所以在训练集中,平均每类(每个人)有 17.2 张训练数据.

2.3　CUHK03 数据集

数据集 CUHK03 图像来自 2 个不同摄像头. 该数据集提供机器检测和手工检测两个数据集. 其中检测数据集包含一些检测误差,更接近实际情况. 平均每个人有 9.6 张训练数据. 其测试协议分为两种:一种是 single-shot setting 协议,具体来说是将随机选出 100 个行人作为测试集,1 160 个行人作为训练集,100 个行人作为验证集(总共 1 360 个行人). 另一种类似于 Market-1501,它将数据集分为包含 767 个行人的训练集和包含 700 个行人的测试集,在测试阶段,随机选择一张图像作为 query,剩下的作为 gallery,这样对于每个行人都有多个 ground truth 在 gallery 中.

在基于局部特征学习的行人重识别系统中,应用了深度学习的方法,Zhao 等人[3]提出的模型在数据集 CUHK03 上的准确率达到 88.5%,比最好的方法[29]高出 10.1%,同时该模型在数据集 Market1501 上可以达到 76.9%,比最好的方法[30]高出 11.0%.

在基于距离度量学习的行人重识别系统中,文献[10]通过改进三元组损失函数在 Caffe 框架上进行实验,在数据集 VIPeR 和 CUHK03 上 rank-1 的性能都要高于 DGD[31].

在基于视频序列学习的行人重识别系统中,数据集 PRID2011 和 iLDS-VID 使用最为广泛. 文献[15]提出的模型在数据集 PRID2011 和 iLDS-VID 上 rank-1 的准确率分别达到了 83.5% 和 69.3%,比 Zhou 等[32]提出利用深度神经网络将特征学习和度量学习统一在一个框架下的方法的准确率分别高出 4.1% 和 14.1%.

在基于生成对抗网络的行人重识别系统中,Qian 等人[21]提出了一个针对行人重识别的姿态归一化而设计的 GAN 的框架,在数据集 CUHK03 和 DUkeMTMC-reID 上 rank-1 的准确率分别达到 79.76% 和 73.58%,这个结果比一些在有监督的学习环境下接受训练的模型的结果有明显的提高.

虽然目前的行人重识别数据集的性能令人满意,但仍然存在一些重识别应用方面的问题. 一个是现有的公共数据集与实际场景中收集的数据不同. 例如,当前的数据集要么包含有限数据的标识,要么在受约束的环境下进行. 目前最大的 DukeMTMC-reID 包含少于 2 000 个身份,并提供简单的标注信息. 这些限制简化了行人重识别任务,有助于实现较高的准确性. 另外一个问题就是,在不同的行人重识别数据集之间存在领域差异,即对不同行人重识别数据集的训练和测试导致性能的严重下降.

3　总结

本文主要介绍了基于深度学习的行人重识别的几种重要方法(基于局部特征学习、基于距离度量学习、基于视频序列学习和生成对抗网络). 基于局部特征学习的方法很好地利用了人体的全局和局部信息,但这些局部特征对齐方法都需要一个额外的骨架关键点或者姿态估计的模型,而训练一个可以达到实

用程度的模型需要搜集足够多的训练数据. 对于这一问题可以利用生成对抗网络技术,而GAN技术的强大之处就在于可以很好地扩充数据集. 基于距离度量学习的方法主要在于损失函数的研究,尤其是三重函数,可以根据自己研究的需要在三重损失函数上做出改动. 而基于视频序列学习的方法就是需要利用好帧与帧之间的信息,提取出一个融合了内容信息、运动信息的特征,进而提高行人重识别的准确度.

4 问题及未来研究趋势

通过上述分析,可以看出基于深度学习的行人重识别的研究取得了一定的成果,研究日益趋向成熟. 但是行人重识别的研究在实际应用中依然很难取得很好的结果. 主要问题和未来的研究方向如下:

(1)在研究过程中会出现这样一种现象,实验模型在有的数据集上性能表现很好,但换到另一个难度更大的数据集上,性能就会大幅下降,也就是领域迁移的问题. 因此设计一个具有鲁棒性并能很好地解决领域迁移问题的模型,仍需进一步的研究.

(2)回顾现有的基于视频的方法,可以发现在图像分类和实例检索的过程中,行人重识别还很难达到准确和高效. 因此将检测过程、特征提取和特征学习联合起来实现端到端的行人重识别也是一个亟待解决的问题.

(3)实际应用中的行人重识别系统,大部分行人图像均无标签信息,利用人工进行标签标记将会浪费大量的人力物力,尤其是基于深度学习的方法,需要大量的数据集,怎样利用少量标签信息进行半监督或者无监督的行人重识别依旧需要进一步的研究.

[参考文献]

[1] BAI X,YANG M K,HUANG T T,et al. Deep-person:learning discriminative deep features for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1711.10658.pdf.

[2] VARIOR R R,SHUAI B,LU J W,et al. A siamese long short-term memory architecture for human re-identification[C]// Proceedings of the European Conference on Computer Vision. Cham:Springer,2016:135-153.

[3] ZHAO H Y,TIAN M Q,SUN S Y,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. Hawaii,2017:1077-1085.

[4] WEI L H,ZHANG S L,YAO H T,et al. Glad:global-local-alignment descriptor for pedestrian retrieval[C]//Proceedings of the 2017 ACM on Multimedia Conference. California,2017:420-428.

[5] ZHANG L,XIANG T,GONG S G. Learning a discriminative null space for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,2016:1239-1248.

[6] ZHENG L,YANG Y,HAUPTMANN A G. Person re-identification:past,present and future[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1610.02984.pdf.

[7] ZHOU S,WANG J J,WANG J Y,et al. Point to set similarity based deep feature learning for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Hawaii,2017:3741-3750.

[8] HERMANS A,BEYER L,LEIBE B. In defense of the triplet loss for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1703.07737.pdf.

[9] CHENG D,GONG Y H,ZHOU S P,et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,2016:1335-1344.

[10] CHEN W H,CHEN X T,ZHANG J G,et al. Beyond triplet loss:a deep quadruplet network for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii,2017:403-412.

[11] XIAO Q Q,LUO H,ZHANG C. Margin sample mining loss:a deep learning based method for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1710.00478.pdf.

[12] MCLAUGHLIN N,RINCON J M D,MILLER P. Recurrent convolutional network for video-based person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,2016:1325-1334.

[13] ZHANG D Y,WU W X,CHENG H,et al. Image-to-video person re-identification with temporally memorized similarity learning[J]. IEEE transactions on circuits & systems for video technology,2017,PP(99):1-1.

[14] MCLAUGHLIN N,RINCON J M D,MILLER P. Recurrent convolutional network for video-based person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,2016:1325-1334.

[15] HUANG W J,LIANG C,YU Y,et al. Video-based person re-identification via self paced weighting[C]//Proceedings of the Thirty-Second Conference on Artificial Intelligence. Louisiana,2018:2273-2280.

[16] LIU H,JIE Z Q,JAYASHREE K,et al. Video-based person re-identification with accumulative motion context[J]. IEEE transactions on circuits & systems for video technology,2017,PP(99):1-1.

[17] SONG G L,LENG B,LIU Y,et al. Region-based quality estimation network for large-scale person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1711.08766.pdf.

[18] ZHENG Z D,ZHENG L,YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1701.07717.pdf.

[19] ZHONG Z,ZHENG L,ZHENG Z D,et al. Camera style adaptation for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1711.10295.pdf.

[20] WEI L H,ZHANG S L,GAO W,et al. Person transfer GAN to bridge domain gap for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1711.08565.pdf.

[21] QIAN X L,FU Y W,WANG W,et al. Pose-normalized image generation for person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1712.02225.pdf.

[22] HE L X,LIANG J,LI H Q,et al. Deep Spatial feature reconstruction for partial person re-identification:alignment-free approach [DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1801.00881.pdf.

[23] SUN Y F,ZHENG L,DENG W J,et al. Svdnet for pedestrian retrieval[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1703.05693.pdf.

[24] ZHONG Z, ZHENG L, CAO D, et al. Re-ranking person re-identification with *k*-reciprocal encoding [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Hawaii:IEEE,2017:3652-3661.

[25] ZHONG Z,ZHENG L,KANG G L,et al. Random erasing data augmentation[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1708.04896.pdf.

[26] SARFRAZ M S,SCHUMANN A,EBERLE A,et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1711.10378.pdf.

[27] FAN H H,ZHENG L,YANG Y. Unsupervised person re-identification:clustering and fine-tuning[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1705.10444.pdf.

[28] BAK S,CARR P,LALONDE J F,et al. Domain adaptation through synthesis for unsupervised person re-identification[DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1804.10094.pdf.

[29] JOSE C,FLEURET F. Scalable metric learning via weighted approximate rank component analysis[C]//European Conference on Computer Vision. Cham:Springer,2016:875-890.

[30] VARIOR R R,HALOI M,WANG G. Gated siamese convolutional neural network architecture for human re-identification[C]//European Conference on Computer Vision. Cham:Springer,2016:791-808.

[31] XIAO T,LI H S,OUYANG W L,et al. Learning deep feature representations with domain guided dropout for person re-identification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,2016:1249-1258.

[32] ZHOU Z,HUANG Y,WANG W,et al. See the forest for the trees:joint spatial and temporal recurrent neural networks for video-based person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii,2017:6776-6785.

[责任编辑:顾晓天]