

基于深度学习的行人重识别综述

杨 锋^{1,2*}, 许 玉¹, 尹梦晓^{1,2}, 符嘉成¹, 黄 冰¹, 梁芳焯¹

(1. 广西大学 计算机与电子信息学院, 南宁 530004; 2. 广西多媒体通信与网络技术重点实验室(广西大学), 南宁 530004)

(* 通信作者电子邮箱 yf@gxu.edu.cn)

摘 要: 行人重识别(Re-ID)是计算机视觉领域的热点问题,主要研究的是“如何关联位于不同物理位置的不同摄像机捕获到的特定人员的问题”。传统的行人Re-ID方法主要基于底层特征如局部描述符、颜色直方图和人体姿势的提取。近几年,针对行人遮挡和姿势不对齐等传统方法所遗留问题,业内提出了基于区域、注意力机制、姿势和生成对抗性网络(GAN)等深度学习的行人Re-ID方法,实验结果得到较明显的提高。故对深度学习在行人Re-ID中的研究进行了总结和分类,区别于以前的综述,将行人重识别方法分成四大类来讨论。首先,通过区域、注意力、姿势和GAN四类方法来综述基于深度学习的行人Re-ID方法;然后,分析这些方法在主流数据集上的mAP和Rank-1指标性能表现,结果显示基于深度学习的方法可以增强局部特征之间的联系并缩小域间隙,从而减少模型过拟合;最后,展望了行人Re-ID方法研究的发展方向。

关键词: 行人重识别;深度学习;生成对抗性网络;区域;注意力;姿势

中图分类号: TP391.4 **文献标志码:** A

Review on deep learning-based pedestrian re-identification

YANG Feng^{1,2*}, XU Yu¹, YIN Mengxiao^{1,2}, FU Jiacheng¹, HUANG Bing¹, LIANG Fangxuan¹

(1. School of Computer, Electronics and Information, Guangxi University, Nanning Guangxi 530004, China;

2. Guangxi Key Laboratory of Multimedia Communications Network Technology (Guangxi University), Nanning Guangxi 530004, China)

Abstract: Pedestrian Re-Identification (Re-ID) is a hot issue in the field of computer vision and mainly focuses on “how to relate to specific person captured by different cameras in different physical locations”. Traditional methods of Re-ID were mainly based on the extraction of low-level features, such as local descriptors, color histograms and human poses. In recent years, in view of the problems in traditional methods such as pedestrian occlusion and posture disalignment, pedestrian Re-ID methods based on deep learning such as region, attention mechanism, posture and Generative Adversarial Network (GAN) were proposed and the experimental results became significantly better than before. Therefore, the researches of deep learning in pedestrian Re-ID were summarized and classified, and different from the previous reviews, the pedestrian Re-ID methods were divided into four categories to discuss in this review. Firstly, the pedestrian Re-ID methods based on deep learning were summarized by following four methods: region, attention, posture, and GAN. Then the performances of mAP (mean Average Precision) and Rank-1 indicators of these methods on the mainstream datasets were analyzed. The results show that the deep learning-based methods can reduce the model overfitting by enhancing the connection between local features and narrowing domain gaps. Finally, the development direction of pedestrian Re-ID method research was forecasted.

Key words: Re-Identification (Re-ID); deep learning; Generative Adversarial Network (GAN); region; attention; pose

0 引言

行人重识别(Re-Identification, Re-ID)是典型的研究图像和视频的问题,可以理解为一个图像检索的子问题,其目的是对跨越摄像机领域中的目标行人进行检测和跟踪,可广泛用于智能视频监控、智能安保等领域。

行人重识别就是要对出现在某摄像机视域中的某一个目

标行人,在另外一个摄像机视域里快速准确地将这一个目标行人识别出来。想要在某个时间或者某个地点识别某个特定的人,对于计算机视觉是一项艰难的任务,关键在于要找到好的特征表示和度量学习方法来证明两张图像的相似性。

1961年Plantinga^[1]在思考心理学和行为学时提出了第一个行人重识别定义^[2]。随后多摄像机跟踪和检测目标技术也相继出现。1997年Huang等^[3]为了分析其他相机中观测到的

收稿日期:2019-10-10;修回日期:2019-12-16;录用日期:2019-12-16。

基金项目:国家自然科学基金资助项目(61861004, 61762007);广西自然科学基金资助项目(2017GXNSFAA198267, 2017GXNSFAA198269)。

作者简介:杨锋(1979—),男,广西玉林人,副教授,博士,CCF会员,主要研究方向:人工智能、网络信息安全、大数据与高性能计算、精准医学;许玉(1993—),女,广西百色人,硕士研究生,主要研究方向:深度学习、行人重识别;尹梦晓(1978—),女,河南南阳人,副教授,博士,CCF会员,主要研究方向:计算机图形学与虚拟现实、数字几何处理、图像与视频编辑、图论;符嘉成(1995—),男,广西柳州人,硕士研究生,主要研究方向:深度学习、医学图像配准;黄冰(1993—),女,广西南宁人,硕士研究生,主要研究方向:医学图像融合;梁芳焯(1994—),女,广西横县人,硕士研究生,主要研究方向:深度学习、图像分割。

视图的证据,以及估计一台相机中物体外观的后验,而提出了一种贝叶斯公式。2005年在多摄像机目标跟踪^[4]的文章中,来自阿姆斯特丹大学的Zajdel等^[5]首次提到了“行人重识别”这一概念。2014年图像分类^[6]中的深度学习^[7]被运用到行人重识别领域^[8],而2016年深度学习如雨后春笋般在行人重识别领域出现,一直流行至今。

行人重识别由两个部分组成^[9]:一部分是侧重于寻找和改进特征提取方法^[10-12],另一部分是找到用于特征比较的相似性度量^[13-17]或者是两者的组合^[18-20],其中文献^[13-15, 17, 19-20]包含了颜色直方图的变化,可以一定程度减轻客观条件变化的干扰,文献^[13-15, 19-20]包含了局部的二值模式,文献^[14]描述的是Gabor特征,文献^[12]基于局部补丁和文献^[10]基于的是显著颜色名称等方法。然而行人重识别还面临着许多复杂环境中非可控因素的问题,如姿势大幅度变化、视角的变化、光线变化、遮挡和图像的分辨率低下等问题。

1)姿势的大幅度变化如图1(a)所示,这就要求行人检测框能够灵活地变化,从而更准确地检测出行人。2)视角变化如图1(b)所示,因为多台摄像机拍摄角度的不同,最后出现一个人有多个视角的情况。3)光线变化如图1(c)所示,拍摄的过程中,可能因为天气等因素的影响,导致视图颜色和对比度的差别,这种颜色和对比度的差别很容易混淆目标。4)遮挡问题如图1(d)所示,行人明显存在部分遮挡的情况,被遮挡的行人往往会导致漏检或者误检。5)如图1(e)是摄像机分辨率低,无法匹配脸部,这种情况下通常提取身体上的某些局部特征来代替。



图1 复杂的环境增加行人重识别的困难

Fig. 1 Complex environments increase the difficulty for pedestrian re-identification

传统方法主要是致力于低级特征,如局部描述符、颜色直方图和人体姿势^[21-22]等。近几年的大多数文章倾向于深度学习的方法,而深度学习的方法侧重于注意力机制以及姿势等,并且该类方法在行人重识别中也取得阶段性的成果^[23-26]。主要是因为深度学习提供了更加高效的方法,用于处理因客观条件引起的目标变化^[6]问题,同时深度学习具有丰富的架构和较强的学习能力,可以通过分层非线性映射学习有效的目标识别功能。而现有的基于深度学习的行人重识别算法大多采用卷积神经网络(Convolutional Neural Network, CNN)结构来进行实验,卷积神经网络结构有较好的特征提取能力,但是对于视频帧中的时间连续性难以实现建模。故近期又针对时间连续性建模提出了循环神经网络(Recurrent Neural Network, RNN)架构来提高识别效率。

1 传统方法和深度学习方法相关工作

1.1 传统方法中的特征提取和匹配

行人重识别技术的两个重要组成部分:特征提取和特征匹配,两者之间相辅相成,从而形成整个识别过程,如图2所示。

1) 特征提取。

特征提取总的来说可分为三类:基于底层的特征表达^[27];基于中高层特征的学习^[28],包括显著性学习^[11]、镜像特征表达^[29]、属性学习^[30-31]等;还有基于混合模型^[32]特征表达。

2) 特征匹配。

特征匹配是指将视图提取出来的特征与目标图片进行匹配,进一步确定是否是同一目标。

2010年Prosser等^[33]提出了一种关于支持向量排序模型的匹配算法,把行人相似性问题的计算转化成行人相对关系的排序问题;Pedagadi等^[34]提出利用行人局部线性判别分析的方法来衡量两行人的相似性。同时国内外的研究者还提出了许多其他的匹配方法,例如:基于稀疏成对约束的度量学习模型^[35]、基于聚类采样的组合模板匹配模型^[36]、基于显著性的匹配方法^[37]等。上述方法中大多数是基于有监督学习的,但是在现实摄像机获取的数据中也会存在大量的无标签的数据样本,这些无标签的数据具有一定的价值,可以加以利用。故近期大量基于无监督学习的相似性度量模型逐渐引起研究者的关注。

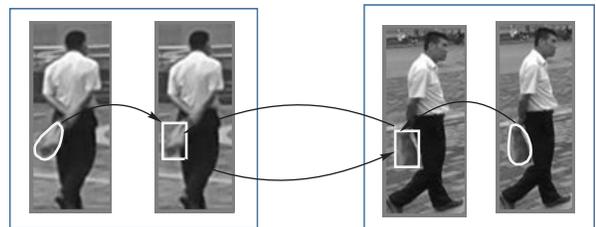


图2 传统方法中的显著性颜色匹配示例

Fig. 2 Example of significant color matching in traditional methods

1.2 卷积神经网络用于提高行人重识别效率

如图3所示,是一个简单的基于深度学习的行人重识别的过程。和传统的行人重识别方法有所不同,基于深度学习的行人重识别不仅使识别过程中各个部分都具有学习的能力,还使得各个部件之间互相连通,不再是单个独立的过程,这样一来各部分的工作效率也会得到提高。图3参照了文献^[8]中的模型,在输入样本之后,首先利用CNN对样本进行特征提取。图3中的卷积神经网络包含两个卷积层、两个最大池化层和一个完全连接层。完全连接层是卷积神经网络的最后一层,它对前几层提取的信息进行特征融合和输出,并把输出的信息传递给连接函数,利用连接函数来评估两个样本之间的关系,而成本函数用于将样本之间的关系转化成成本。根据成本函数的输出结果来判断是否是同一个人。

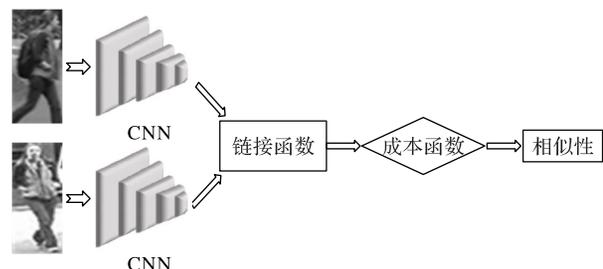


图3 基于深度学习的行人重识别过程

Fig. 3 Process of pedestrian re-identification based on deep learning

2 四类方法在行人重识别中的应用

本章总结了近几年深度学习在行人重识别领域的发展,将围绕区域(region)、注意力(attention)、姿势(pose)和生成对

抗性网络(Generative Adversarial Network, GAN)四类方法来讲述,因为四类方法针对的问题和解决问题的思路各有偏重,所以将方法分成四类有助于对未来的研究。

2.1 基于区域的方法

基于区域(region)的方法^[38]是众多行人识别匹配方法中的一个小分支,该方法在传统手工方法的基础上有一定的进展。其内容包括将行人分为水平条纹^[39]或者分成几个部位来进行识别,从而可以更好地提取颜色和纹理特征。许多论文也采用了类似的水平条纹分区^[40-42],还有一些模型采用局部特征的方法^[43]。在这些方法中,关键还是要使模型具有学习的能力。因为识别检测的过程千变万化,要增加识别过程的灵活性就要增强模型的学习能力。从基于区域的角度出发,可以大体分成四个部分,方法总结见表1。

2.1.1 区域分区方法中的水平条纹分区

由于图像的像素不稳定,故需要一个有效的表示来对行人进行特征提取和匹配。2014年Yi等^[8]提出了深度度量学习方法(Deep Metric Learning, DML),此方法是受到神经网络启发,用暹罗卷积神经网络(Siamese Convolutional Neural Network, SCNN)将目标行人图像强行分成3个大的部分来评估两图像的相似性,此方法从理论上说,应该是归类于基于部件的方法,但是深入观察发现与基于条纹的方法也有相似之处。其原理为,例如两张图片分别是A和B,那么 $S=DML(A, B)=DML(X(A), Y(B))$,其中X和Y表示DML的两个子网络。

在此之前也有一些较好的度量学习方法,比如2012年Köstinger等^[13]验证了KISSME(Keep It Simple and Straightforward METric)方法比同类方法效果略优;Zheng等^[44]提出相对距离比较(Relative Distance Comparison, RDC)模型,并在3个公开的数据集上进行了实验,实验结果表明RDC明显优于当下流行的其他模型。多次实验还发现新的RDC模型对外观和视觉效果更好,使模型对客观因素更具鲁棒性,同时降低了模型的过度拟合性。在后续工作中Tao等^[45-46]又在Zheng等^[44]的基础上进行了优化和扩展。

深度度量学习(DML)较前面几个方法优势在于模型中学习的多通道滤波器可以同时检测纹理和颜色特征,明显比传统方法中的简单融合更合理,也使得识别过程更具灵活性。

2.1.2 基于区域分区的局部特征方法

早期的一些研究会假设几何和光照等因素的变化,并且可以从样本中学习模型参数。例如2003年Porikli等^[47]预设变换是单峰的。2008年,Prosser等^[48]预设光度变换是双向累积亮度传递函数。而2014年Li等^[18]提出一种新的滤波器配对神经网络(Filter Pairing Neural Network, FPNN)模型来进行行人重识别,所提出的滤波器配对神经网络,利用滤波器和最大分组层来学习光度变换。此模型较深度度量学习更细致,把行人分成更多小块来计算相似性,在深度神经网络下统一完成未对准、遮挡和光度变化等问题的处理,使识别通道的各个部件联合起来工作,提高了识别的灵活性。许多研究想通过改进原来的方法来提高行人的识别效率,然而不少识别方法中的视觉特征表示是手动的,但手动的方法容易失去空间信息导致辨别率下降,使用基于贴片的局部^[49-51]特征可以适当地提高辨别率,但是没有根本地解决问题,要更好解决问题还是要求模型有从数据中学习的能力。

2.1.3 局部分区和全局特征协同工作

前人利用深度学习的优势,使得行人重识别技术逐渐完善,现阶段还是要着力于如何降低外界因素对行人重识别的影响和加强局部特征之间的联系^[51-53]。2017年Yao等^[54]提出部分损失网络(Part Loss Network, PL-Net)的深度学习方法,并指出只看全局特征是局限的,所以把目标人物分成几个部分来进行表示,但没有考虑几个局部特征之间的联系和不对齐,依然没有根本解决问题。同年一些文章考虑到以上问题,使用姿势估计结果对两行人图像进行对齐^[51-53],此过程需要大量额外的监督和姿势预测过程,实现较复杂,故2017年Zhang等^[55]基于部件的方法提出了对齐识别(Aligned ReID)模型,实现了局部特征和全局特征协同工作。

如图4,给定两幅图的局部特征,其中实线表示对最短路贡献较大,虚线次之。两张图片的特征分别为 $F=\{f_1, f_2, \dots, f_H\}$ 和 $G=\{g_1, g_2, \dots, g_H\}$,通过元素转化将距离归一化为 $[0, 1)$:

$$d_{i,j} = \frac{e^{\|f_i - g_j\|_2} - 1}{e^{\|f_i - g_j\|_2} + 1}; i, j \in (1, 2, \dots, H) \quad (1)$$

其中 $d_{i,j}$ 中的 i 表示图4中左边行人图像的第 i 水平区间和图4中右边行人图像的第 j 水平区间之间的距离,通过这些距离得到距离矩阵 D ,文献^[55]中将两图像之间的局部距离定义为矩阵 D 从 $(1, 1)$ 到 (H, H) 的最短总路径。像这样通过计算两局部特征之间的最短路径来实现对齐和匹配的方法,不需要额外的监督,在一定程度上减轻了工作,但是从实验结果来看,对于行人姿势大幅度变化和遮挡的问题还是没有得以最佳的解决。

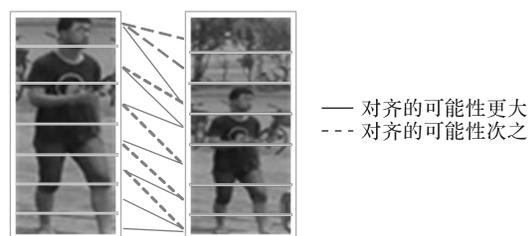


图4 基于条纹区域的对齐

Fig. 4 Alignment based on stripe regions

2.1.4 基于区域的网格分区方法

对遮挡问题处理较好的方法是,2018年He等^[56]针对遮挡的问题提出部分行人重识别方法。文献首先是利用完全卷积神经网络来生成大小固定的空间特征映射,其次为了匹配一对不同大小的图像,提出了一种深度空间特征重建方法,来避免显式对齐。此方法不要求特征对齐,在数据集Market1501上Rank-1指标提高到了83.58%。

基于区域的方法,如2015年Ahmed等^[9]的研究,如图5,输入两张图片进行匹配,输入的一张图片能够在另一张图片区域中找到匹配,思路是将两张行人图像划分成多个区域来进行实验。与以前方法不同的是此架构在模型中新增了一层卷积层,所以文献^[9]有三个卷积层,网络更深,在一些小的数据集上得出较好的结果,模型过度拟合现象得到一定的改善,但是匹配不适应大的场景变化。

2016年Wu等^[57]提出一种基于深度卷积神经网络的行人模型PersonNet(Person re-identification with Deep Convolutional Neural Network),在较大的数据集CUHK03和Market-1501上

进行了训练,实验结果得到小部分改善。2016年Zhang等^[58]提出的端到端的深度对应结构(Deep Correspondence Structure Learning, DCSL),端到端^[2]的方法本质上就是将行人重识别中独立的模块融合在一起,当从输入端输入一张图片时,输出端得到一个预测的结果去跟真实的结果对比。如果出现误差,可以返回到模型中去调整这个误差,使之更接近预期的结果。这种方法可以解决部分跨摄像机的目标识别问题。将网络结构转化成金字塔类型,在本层找不到合适的匹配时,转到上一层去寻找匹配,以基于语义的视角来思考问题。这种复杂的方法对于场景变化和未对准具有稳健性。

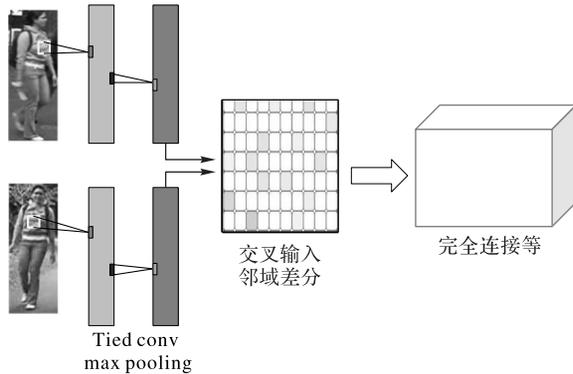


图5 基于网格区域结构的行人重识别方法

Fig. 5 Pedestrian re-identification method based on grid region structure

2.2 基于注意力机制的方法

注意力机制方法就是选择一个最具代表性的局部特征来进行下一步的跟踪,如图6。本节分析和总结近期受注意力启发的一些模型,分为三部分,如表1。

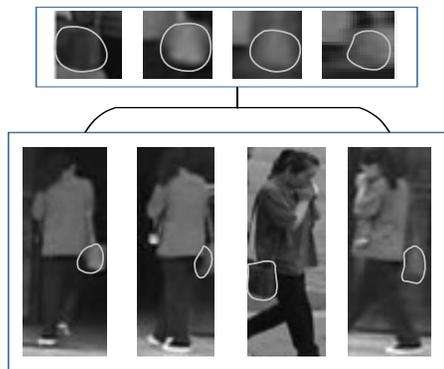


图6 基于注意力机制的行人重识别方法

Fig. 6 Pedestrian re-identification method based on attention mechanism

2.2.1 细粒度特征识别的出现

行人重识别是具有挑战性的,人们也在不断地从各个方面寻求和改进能够解决问题的方法。比如说文献[18,21,35,59]是从外观进行改进的方法,现阶段深度学习成了大多数外观特征表示的首选方法。如2.1节方法中所述,可以把行人图像以条纹^[60]或者部件等形式来划分,这些都是默认行人姿势和边界框的空间分布是相似的。可是在实际情况中,行人姿势复杂,划分的边界框不一定能够适应所有的行人,实现起来和理想状态总有一定的差距,就算用复杂的匹配技术也不一定能够得到理想的结果。针对此问题,2017年Zhao等^[61]受到注意力模型的启发提出了部分对齐(Part-Aligned)方法,主

要步骤是,把在匹配的过程中有区别的局部进行检测,先计算局部的表示,最后汇总各个部分之间的相似性,提出的局部对齐特征表示提取器,也是一个深度神经网络,是由完全卷积神经网络(Full Convolutional Neural Network, FCNN)组成,实验得出较好的结果。

在行人识别的过程中,卷积网络在辨别行人特征方面表现出让人满意的结果,也是当下主流的方法。然而通过研究者们分析得出,虽然在卷积神经网络的贡献下,结果有明显的提高,但还是存在一些识别死角。例如对细粒度综合特征的行人识别,没有得到充分的解决。故2017年Liu等^[62]基于注意力神经网络,提出一种对细粒度特征识别有优势的网络模型HPnet(HydraPlus-Net)。它主要是通过将多层的注意力图以多方向地反馈到不同的层,从而达到识别目的。

2.2.2 基于局部斑块和显著性的方法

之前提到遮挡、背景杂波和行人姿势的大幅度形变等客观因素引起的问题依然存在。在检测过程中,为改善实验效果,研究者们针对这些客观因素引起的复杂问题提出许多行之有效的办法,如自动检测的行人检测框,此类基于注意力识别的方法模型,可以解决一定的遮挡和背景杂波等问题。

还有一些思路是想通过选择注意力行人检测框来解决遮挡等问题,主流的方法是局部斑块校准匹配和显著性加权^[63-65]。如2013年Zhao等^[37]提出了一种显著性学习行人识别方法,将行人显著性特征加入补丁匹配中,使得模型能够有效地找到有区别且可靠的补丁匹配特征,实验在数据集ETHZ和数据集VIPeR上得到相应的验证。

进一步的研究是2018年,在原来研究成果的基础上,受注意力^[66-68]神经网络启发,Li等^[49]提出了注意力融合卷积神经网络(Harmonious Attention CNN, HA-CNN)模型,用于硬区域和软像素联合学习,同时优化特征表示,从而实现优化图像未对准情况的行人识别。

在最近的研究中就遮挡的问题又有了新的成就。2018年Li等^[67]提出了一种新的时空注意力模型,此模型可以自动地发现行人身体的各个特殊部位,利用多个空间注意力模型提取每个图像帧的有用部位信息,并且通过时间注意力模型整合输出,允许从所有帧中提取可利用的部位信息,对遮挡和错位等问题有一定的改善,在三大数据集上进行评估,实验结果超过了部分先进的方法。还有一些基于片段相似性聚合和协同注意力片段嵌入的视频行人识别^[68]也取得了较好的实验结果。

2.3 基于姿势分类方法中的迁移和嵌入

如2.2节所述,在行人图像拍摄的过程中,难免存在一些由客观因素引起的复杂视图变化等问题。如图7所示,最早的基于姿势的研究是Spindel Net^[51]方法,其次是Wei等^[53]对姿势进行预估。而近期基于姿势的行人识别方法,在该问题上也取得了阶段性的成果。2017年Su等^[69]为解决此类问题,提出了一种姿势驱动的深度卷积(Pose-driven Deep Convolution, PDC)模型,此模型改进了常规的特征提取和匹配模型。首先是特征嵌入子网(Feature Embedding sub-Net, FEN),实现的功能是,利用人体姿势信息,将身体全局图像转换为包含规范化部分区域的图像,再通过全局特征和不同局部特征相结合学习强大的特征表示。为了进行特征匹配还设计了特征加权子网(Feature Weighted sub-Net, PWN)。实验操作中首先是将姿势的信息嵌入到结构化网络当中,然后生成modified结构图像,最后是对这个结构图像进行识别。实验在3个常用的数据集上得到验证,故得出结论考虑额外姿

势线索可以提高 Rank-1 准确率。

2018 年 Liu 等^[70]提出一种 Data Augment 的方法。主要是通过引入存在遮挡的姿势的样本库,再结合 GAN 生成多姿态的样本标签,用于辅助后期的训练。但是训练 GAN 生成的样本数据集因辨识度不高对训练模型帮助不大,文献中还提出了一种姿态迁移的行人重识别框架来改善实验结果。



图 7 行人姿势的变化

Fig. 7 Change of pedestrian posture

有些研究强调姿势和嵌入(Pose+Embedding),如 2018 年 Sarfraz 等^[71]就是基于姿势和嵌入提出一种行之有效姿势敏感嵌入(Pose-Sensitive Embedding, PSE)的方法。此方法结合人细致(如人的关节等)和粗略的姿势(全局的人像)信息来学习嵌入。又基于计算距离的排序技术^[72-76],提出和改进一种无监督的自动排序框架来提高识别性能,并在几个大的数据集上进行了实验,取得具有竞争力的实验效果。

2.4 基于生成对抗性网络的方法

在以上行人重识别工作中,利用深度学习很好地解决遮挡和不对齐等问题。但是通过对各个方法在数据集上的表现进行分析时发现,部分方法在某个数据集上表现较好,却在另外一个数据集上训练可能就没有那么好的实验结果。像这样因为训练数据集的改变引起的实验效果变化,很可能是模型

的过度拟合导致的。解决此类问题,可以从扩大数据集和增加数据多样性方面着手。

故 2017 年 Zheng 等^[15]利用生成对抗性网络来生成更多模拟的数据,从实验结果来看,模型对数据集的过度依赖确实得到部分的优化,但是数据集生成的样本存在辨识度不高和有些复杂的图像出现了扭曲等问题,与真实图像差距较大。

针对以上问题,Liu 等^[70]提出了一种姿态迁移的行人重识别框架。在传统 GAN 判别器的基础上增加一个引导子网络,使得网络能够生成更具代表性的行人重识别样本。更进一步的研究是 2018 年 Zhong 等^[77]提出的 CamStyle 方法,主要是利用 CycleGAN^[78]和标签平滑正则化(Label Smooth Regularization, LSR)来改善模型在训练中生成扭曲图像和产生噪声的问题,改进幅度可达到 17% 左右,但是对生成的数据集进行分析时发现,该论文中的方法只是学习了摄像机的风格(色调),没有对行人的姿势进行处理。

对行人姿势处理较好的是 Zhu 等^[79]提出的姿势转移对抗性网络,将给定的行人姿势转换成目标行人的姿势,从而实现生成新的行人姿势。转换姿势的还有 Song 等^[80]的研究。

传统的 GAN 模型生成的模拟数据集之间仍然存在差异性,导致不同的数据集在相同的模型上训练出来的结果稍有偏差,也叫域间隙。2018 年 Deng 等^[81]在论文中提到用迁移学习的方法来解决此类问题,在无标签的数据集 A 和有标签的数据集 B 上进行训练,最后在数据集 A 的测试数据集上测试。2018 年 Wei 等^[82]也是针对域间隙的问题提出了人员转移生成对抗性网络(Person Transfer GAN, PTGAN),来弥补领域差距;同时还提供了一个叫作 MSMT17 的新数据集,此数据集增加许多重要的功能,并且提供许多较复杂的图像。结合此数据集进行实验,结果表明迁移生成对抗性网络确实缩小了域间隙。

表 1 四种类型方法的特点

Tab. 1 Characteristics of four categories of methods

种类	方法	特点
区域	DeepMetric ^[8]	基于区域的方法将行人图像分成水平条纹或者几个均匀的部分,可以有效地观察各个分区的差异值;可优化内部一致性,以便更好地提取颜色和纹理特征;结合卷积神经网络使模型具有学习的能力;对于部分遮挡和姿势小规模的变化具有鲁棒性;但是这类方法对于遮挡和姿势不对齐等问题还存在大的挑战,对于行人图像姿势的大幅度变化还有待提高
	Deep Reid ^[18]	
	PL-Net ^[54]	
	Aligned Reid ^[55]	
	IDLA ^[9]	
	DCSL ^[58]	
注意力机制	PersonNet ^[57]	针对行人的脸部是模糊的,可以考虑从行人身体上其他比较特别的局部特征入手。此类方法对于行人的遮挡和分辨率不高等问题有一定的帮助,但是在识别的过程中有可能丢失更重要的数据信息,导致识别效率不高
	Part-Aligned ^[61]	
	HydraPlus-Net ^[62]	
姿势	HA-CNN ^[49]	基于姿势的方法大多是从全局和局部特征出发来缓解姿势变化,利用全局图像的完整性与不同的局部相结合来学习特征表示,此类方法一定程度上可以解决因姿势变化引起的行人特征形变问题
	PDC ^[69]	
生成对抗性网络	PSE ^[71]	此类方法大多数是通过改进 GAN 来增强数据集,从而缓解因域间隙引起的模型过度拟合性问题。一般会结合上述类型的方法,一起完成识别过程。从实验结果来看确实部分增强了数据集,但是整体的识别效率还有待提高
	DCGAN ^[15]	
	SPGAN ^[79]	
	Pose+transfer ^[70]	
	IDE*+CamStyle+RE ^[77]	
	PTGAN ^[82]	

3 在数据集上的性能分析

数据集在行人重识别中扮演着重要的角色。在行人重识别方法研究的发展过程中,涉及到多种多样的数据集,而本文选出几个比较具有代表性的进行分析。常用到的三大数据集

Market-1501、CUHK03 和 DukeMTMC-reID。另外还有一个 VIPeR 数据集,该数据集在实验中也经常用到,但是图像较少,一般是结合三大数据集进行实验。先简单介绍几个数据集,然后对上一章节讨论的方法在这 3 个数据集上的实验结

果进行分析比较。

3.1 数据集的介绍

Market-1501^[38]数据集的采集地点是清华大学,数据集图像如图8(a)所示,整个数据集共有32668张图像,图像由6个不同的摄像头拍摄。共包含1501个行人。其中:训练集有751人,12936张图像;测试集有750人,19732张图像。该数据集采用可形变数据模型(Deformable Part Model, DPM)进行行人检测,并且数据集中的每个行人都有多张图像。故每个身份都有多个查询和身份的确认。相关的文献有[38,49,55,61,69,71]。

DukeMTMC-ReID^[50]数据集包含8个摄像头,在杜兰克大学校园里采集。数据集图像如图8(b)所示,同样该数据集提供有用于测试的测试集和用于训练的训练集。训练集中702个行人,对应16522张图像,测试集中也有702个行人,对应17611张图像。平均每个身份有23.5张图像,同时提供的行人属性是三大数据集中最大的一个。

CUHK03^[18]数据集中包含1360个行人和13164张图像。该数据集在香港中文大学采集,图像如图8(c)所示,有两个摄像头拍摄。在早一些工作中还发布了CUHK01^[83]和CUHK02^[14]两个数据集,CUHK03在CUHK01和CUHK02上做了相应的扩充。

VIPeR数据集^[84]有632个行人和1264张图像,由两个摄像头拍摄而成,平均每人有2张图像。主要的特征是光照、颜色和视角的多样性。数据集图像如图8(d)所示。

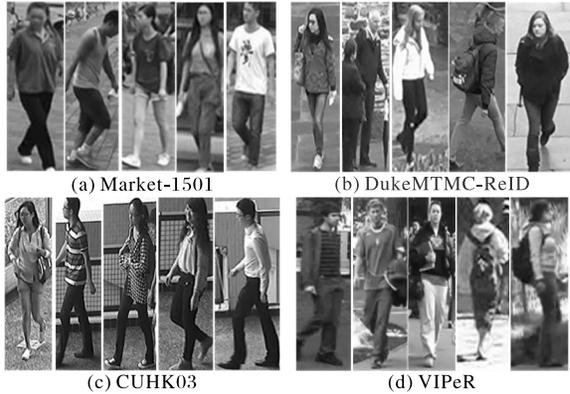


图8 四个数据集的行人图像示例

Fig. 8 Examples of pedestrian images in four datasets

3.2 衡量实验结果的指标

mAP(mean Average Precision)^[85]和Rank-1是常用来衡量搜索能力的两个指标,文献[49,55,62]中还用到累计匹配特征(Cumulative Match Characteristic, CMC)曲线,CMC表示查询标识出现的不同候选列表中的概率^[2]。通常在摄像机存在

一个基本事实的情况下,使用CMC比较准确。但是对于像Market-1501这样来自多摄像机的多个基本事实的数据集,CMC的判断能力存在一定的局限性。相比之下,mAP更适用于当数据集中存在多个基本事实的评估^[38]。以下对论文中关于mAP和Rank-1这两个指标的实验结果进行了分析总结。

mAP即各类别平均精度(Average Precision, AP)的平均值,AP是PR(Precision-Recall)曲线下的面积,其中:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

其中:TP(True Positives)表示IoU>0.5(交并比(Intersection over Union, IoU))的检测框的数量(同一个真实标签(Ground Truth)只计算一次);FP(False Positives)是表示IoU≤0.5的检测框数量,也可以说是检测到同一个真实标签的多余检测框的数量;FN(False Negatives)表示没有检测到真实标签检测框的数量。

要得出mAP的数值就要绘制出PR曲线,再计算PR曲线下的面积得出平均精度AP。由此看来关键在于如何对PR曲线进行采样。2010年改变了AP测量的计算细节^[86],是为了提高mAP指标的精确度。在VOC2007以前,可以选取当Recall≥0,0.1,0.2,⋯,1共11个点时的精度(Precision)最大值,2007年AP采用的是11个精度的平均值,公式如下:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} p_{inter}(r) \quad (4)$$

其中 $p_{inter}(r)$ 表示精度插值,公式如式(5),精度值取所有Recall>r中的最大值:

$$p_{inter}(r) = \max p(\tilde{r}) \quad (5)$$

\tilde{r} 的范围是 $\tilde{r} \geq r$,这里的 $\max p(\tilde{r})$ 是召回率 \tilde{r} 的最大测量精度。

在VOC2010及以后,需要针对每一个不同的Recall值(包括0和1),选取其大于等于这些Recall值时的Precision最大值,然后计算PR曲线下的面积作为AP值。Rank-1表示检测结果中最靠前的一张图像是正确结果的概率。

3.3 实验结果的对比

就mAP表现出来的值来看,如表2所示,表现较好的有PSE+ECN方法,该方法考虑到行人粗略的和细致的姿势信息来嵌入学习,在几个数据集上的表现较均衡。单看CUHK03数据集上的表现最好的是Zhao等^[61]提出的Part-Aligned方法,但是在Market-1501上的表现一般,此论文中强调先进行部分对齐,再把部分进行融合。

表2 各种方法在数据集上的mAP值

Tab. 2 mAP values of various methods on datasets

种类	方法	各数据集 mAP/%		
		CUHK03	Market-1501	DukeMTMC-ReID
区域	AlignedReID	—	79.30	—
注意力机制	Part-Aligned	90.9	63.40	—
	HA-CNN	—	75.70	63.80
姿势	PDC	—	63.41	—
	PSE+ECN	—	84.00	79.80
生成对抗性网络	Pose+transfer(D, Tri)	—	68.92	56.91
	IDE*+CamStyle+RE	—	71.55	57.61

注: D表示DenseNet-169, Tri表示三重损失, IDE*是对IDE的改进, RE是Random Erasing。

如表3所示,通过对Rank-1值的分析,在4个大的类型中基于部件的AlignedReID在数据集CUHK03上的表现略优,此方法是Zhang等^[55]在2017年提出的。主要是在模型中增加了动态匹配的过程,效果表现较好,但是对行人姿势大幅度改变和遮挡的检测效果还有待提高。表中还呈现了一些较好的方法,如Liu等^[62]提出的HydraPlus-Net在数据集CUHK03和VIPeR上也得到较好的结果,但是在Market-1501上的表现并不突出。包括Liu等^[70]提出Pose+transfer方法,在数据集CUHK03上的表现与其他两个数据集差距较大。

表3 各种方法在数据集上的Rank-1值

Tab. 3 Rank-1 values of various methods on datasets

种类	方法	各数据集 Rank-1/%			
		CUHK03	Market-1501	VIPeR	DukeMTMC-reID
区域	DeepReID	20.7	—	—	—
	DML	—	—	28.23	—
	AlignedReID	92.4	91.80	—	—
	IDLA	54.7	—	—	-
	DCSL	80.2	—	—	—
	PersonNet	64.8	37.20	—	—
注意力机制	Part-Aligned	85.4	81.00	—	-
	HydraPlus-Net	91.8	76.90	56.60	-
	HA-CNN	—	91.20	—	80.5
姿势	PDC	88.7	84.10	51.27	-
	PSE+ECN	—	90.30	—	85.20
生成对抗性网络	Pose-transfer(D, Tri)	—	87.65	—	78.52
	IDE*+CamStyle+RE	—	89.49	—	78.32

注:D表示DenseNet-169, Tri表示三重损失, IDE*是对IDE的改进, RE是Random Erasing。

4 结语

行人重识别技术在几个大规模的数据集上取得了阶段性的成果,但是仍然还有一些实际性的问题需要解决。比如视频分辨率较低、行人被遮挡、姿势变化和复杂环境中的非可控因素等问题。此外,对基于深度学习的行人重识别方法,在无标注或者少标注情形下的学习问题更加突出。

深度学习模型通常需要含有较多信息和较大的数据集来进行实验,为了得到更多的数据,主流方法是利用生成对抗性网络来实现。从第3章的实验结果来看,以上部分方法初步取得了较好的结果,但该类方法在增强数据集时所生成的图片姿势改变不大、结构单一、分辨率不高、存在模糊现象等。在后续工作中可以考虑如何在保持图像像素不变的前提下生成一些姿势大幅度变化的行人图像,比如一个行人图片在像素保持不变的情况下从站着的姿势生成坐着的姿势,同时还可以考虑对生成的行人图像中的行人增加帽子、眼镜、书包等配饰来增强数据集。

此外,深度神经网络具有的丰富的架构使得算法精度得到提升,但往往也伴随着复杂的计算和更新模型时繁琐的数据参数变更,势必会影响训练的时效性。在以后的工作中,更希望能找到一种高速并且有效的算法来弥补以上的不足。

参考文献 (References)

[1] PLANTINGA A. Things and persons[J]. Review of Metaphysics, 1961, 14(3):493-519.
 [2] ZHENG L, YANG Y, HAUPTMANN A G. Person reidentification: past, present and future[EB/OL]. [2018-10-10]. <https://arxiv.org/pdf/1610.02984.pdf>.

由表2和表3的实验结果可以看出,在没有使用生成对抗性网络的方法模型上,大多数模型或多或少地存在过度拟合。解决该问题的主流方法是利用GAN^[87]生成模拟的图像来增强数据集。从实验效果来看,采用无监督和弱监督的方法,可以有效地解决样本不足的问题。基于这些无监督和弱监督的方法,生成对抗性网络可以生成迷惑机器的负样本,从而增强分类器的判别能力。效果较好的是Zhong等^[77]提出的CamStyle方法,该方法考虑到了模型生成模拟的负样本的同时还会产生噪声。

[3] HUANG T, RUSSELL S. Object identification in a Bayesian context [C]// Proceedings of the 15th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1997: 1276-1282.
 [4] HAMEETE P, LEYSEN S, VAN DER LAAN T, et al. Intelligent multi-camera video surveillance[J]. International Journal on Information Technologies and Security, 2012, 4(4):51-62.
 [5] ZAJDEL W, ZIVKOVIC Z, KROSE B J A. Keeping track of humans: have I seen this person before?[C]// Proceedings of the 2005 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2005: 2081-2086.
 [6] SUN Y, CHEN Y, WANG X, et al. Deep learning face representation by joint identification-verification[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 1988-1996.
 [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2012: 1097-1105.
 [8] YI D, LEI Z, LIAO S, et al. Deep metric learning for person re-identification [C]// Proceedings of the 22nd International Conference on Pattern Recognition. Piscataway: IEEE, 2014: 34-39.
 [9] AHMED E, JONES M, MARKS T K. An improved deep learning architecture for person re-identification [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3908-3916.
 [10] YANG Y, YANG J, YAN J, et al. Salient color names for person

- re-identification [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8689. Cham: Springer, 2014: 536-551.
- [11] ZHAO R, OUYANG W, WANG X. Learning mid-level filters for person re-identification [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 144-151.
- [12] ZHANG Z, CHEN Y, SALIGRAMA V. A novel visual word co-occurrence model for person re-identification [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8927. Cham: Springer, 2014: 122-133.
- [13] KÖSTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2288-2295.
- [14] LI W, WANG X. Locally aligned feature transforms across views [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3594-3601.
- [15] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3754-3762.
- [16] LI Z, CHANG S, LIANG F, et al. Learning locally-adaptive decision functions for person verification [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3610-3617.
- [17] ZHAO R, OUYANG W, WANG X. Person re-identification by saliency matching [C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 2528-2535.
- [18] LI W, ZHAO R, XIAO T, et al. DeepReID: deep filter pairing neural network for person re-identification [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 152-159.
- [19] KHAMIS S, KUO C H, SINGH V K, et al. Joint learning for attribute-consistent person re-identification [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8927. Cham: Springer, 2014: 134-146.
- [20] XIONG F, GOU M, CAMPS O, et al. Person re-identification using kernel-based metric learning methods [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8695. Cham: Springer, 2014: 1-16.
- [21] FARENZENA M, BAZZANI L, PERINA A, et al. Person re-identification by symmetry-driven accumulation of local features [C]// Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 2360-2367.
- [22] HAMDOUN O, MOUTARDE F, STANCIULESCU B, et al. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences [C]// Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras. Piscataway: IEEE, 2008: 1-6.
- [23] MATSUKAWA T, SUZUKI E. Person re-identification using CNN features learned from combination of attributes [C]// Proceedings of the 23rd International Conference on Pattern Recognition. Piscataway: IEEE, 2017: 2428-2433.
- [24] VARIOR R R, HALOI M, WANG G. Gated Siamese convolutional neural network architecture for human re-identification [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9912. Cham: Springer, 2016: 791-808.
- [25] CHENG D, GONG Y, ZHOU S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1335-1344.
- [26] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning [J]. Pattern Recognition, 2019, 95: 151-161.
- [27] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2005: 886-893.
- [28] 郑伟诗, 吴岸聪. 非对称行人重识别: 跨摄像机持续行人追踪 [J]. 中国科学: 信息科学, 2018, 48(5): 545-563. (ZHENG W S, WU A C. Asymmetric person re-identification: cross-view person tracking in a large camera network [J]. SCIENTIA SINICA Informationis, 2018, 48(5): 545-563.)
- [29] CHEN Y, ZHENG W, LAI J. Mirror representation for modeling view-specific transform in person re-identification [C]// Proceedings of the 24th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2015: 3402-3408.
- [30] SU C, YANG F, ZHANG S, et al. Multi-task learning with low rank attribute embedding for person re-identification [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3739-3747.
- [31] SU C, ZHANG S, XING J, et al. Deep attributes driven multi-camera person re-identification [C]// Proceedings of the 2016 European Conference on Computer Vision. Cham: Springer, 2016: 475-491.
- [32] 苏松志, 李绍滋, 陈淑媛, 等. 行人检测技术综述 [J]. 电子学报, 2012, 40(4): 814-820. (SU S Z, LI S Z, CHEN S Y, et al. A survey on pedestrian detection [J]. Acta Electronica Sinica, 2012, 40(4): 814-820.)
- [33] PROSSER B J, ZHENG W, GONG S, et al. Person re-identification by support vector ranking [C]// Proceedings of the 2010 British Machine Vision Conference. Durham: BMVA, 2010: No. 21.
- [34] PEDAGADI S, ORWELL J, VELASTIN S, et al. Local Fisher discriminant analysis for pedestrian re-identification [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3318-3325.
- [35] MIGNON A, JURIE F. PCCA: a new approach for distance learning from sparse pairwise constraints [C]// Proceedings of the 2012 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2666-2672.
- [36] XU Y, LIN L, ZHENG W, et al. Human re-identification by matching compositional template with cluster sampling [C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 3152-3159.
- [37] ZHAO R, OUYANG W, WANG X. Unsupervised saliency learning for person re-identification [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3586-3593.
- [38] SUN Y, XU Q, LI Y, et al. Perceive where to focus: learning visibility-aware part-level features for partial person re-identification

- [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 393-402.
- [39] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11208. Cham: Springer, 2018: 501-518.
- [40] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 2197-2206.
- [41] MA A J, YUEN P C, LI J. Domain transfer support vector ranking for person re-identification without target camera label information[C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 3567-3574.
- [42] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]// Proceedings of the 2018 ACM Multimedia Conference. New York: ACM, 2018: 274-282.
- [43] GRAY D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]// Proceedings of the 2008 European Conference on Computer Vision, LNCS 5302. Berlin: Springer, 2008: 262-275.
- [44] ZHENG W, GONG S, XIANG T. Reidentification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3):653-668.
- [45] TAO D, JIN L, WANG Y, et al. Person re-identification by regularized smoothing KISS metric learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(10): 1675-1685.
- [46] TAO D, JIN L, WANG Y, et al. Person reidentification by minimum classification error-based KISS metric learning[J]. IEEE Transactions on Cybernetics, 2015, 45(2): 242-252.
- [47] PORIKLI F. Inter-camera color calibration by correlation model function[C]// Proceedings of the 2003 International Conference on Image Processing. Piscataway: IEEE, 2003: II-133.
- [48] PROSSER B, GONG S, XIANG T. Multi-camera matching using bi-directional cumulative brightness transfer functions[C]// Proceedings of the 2008 British Machine Vision Conference. Durham: BMVA, 2008: No. 64.
- [49] LI W, ZHU X, GONG S. Harmonious attention network for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2285-2294.
- [50] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 17-35.
- [51] 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 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 907-915.
- [52] 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.
- [53] WEI L, ZHANG S, YAO H, et al. GLAD: global-local-alignment descriptor for pedestrian retrieval[C]// Proceedings of the 25th ACM Multimedia Conference. New York: ACM, 2017: 420-428.
- [54] 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.
- [55] ZHANG X, LUO H, FAN X, et al. AlignedReID: surpassing human-level performance in person re-identification [EB/OL]. [2018-11-08]. <https://arxiv.org/pdf/1711.08184.pdf>.
- [56] HE L, LIANG J, LI H, et al. Deep spatial feature reconstruction for partial person re-identification: alignment-free approach[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7073-7082.
- [57] WU L, SHEN C, VAN DEN HENGEL A. PersonNet: person re-identification with deep convolutional neural networks [EB/OL]. [2019-01-10]. <https://arxiv.org/pdf/1601.07255.pdf>.
- [58] ZHANG Y, LI X, ZHAO L, et al. Semantics-aware deep correspondence structure learning for robust person re-identification[C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2016: 3545-3551.
- [59] 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, 2016: 1363-1372.
- [60] VARIOR R R, SHUAI B, LU J, et al. A Siamese long short-term memory architecture for human re-identification[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9911. Cham: Springer, 2016: 135-153.
- [61] ZHAO L, LI X, ZHUANG Y, et al. Deeply-learned part-aligned representations for person re-identification[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3239-3248.
- [62] LIU X, ZHAO H, TIAN M, et al. HydraPlus-Net: attentive deep features for pedestrian analysis[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 350-359.
- [63] SHEN Y, LIN W, YAN J, et al. Person re-identification with correspondence structure learning[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3200-3208.
- [64] ZHENG W, LI X, XIANG T, et al. Partial person re-identification[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 4678-4686.
- [65] WANG H, GONG S, XIANG T. Unsupervised learning of generative topic saliency for person re-identification[C]// Proceedings of the 2014 British Machine Vision Conference. Durham: BMVA, 2014: No. 19.
- [66] XU J, ZHAO R, ZHU F, et al. Attention-aware compositional network for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2119-2128.
- [67] LI S, BAK S, CARR P, et al. Diversity regularized spatiotemporal attention for video-based person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 369-378.
- [68] CHEN D, LI H, XIAO T, et al. Video person re-identification

- with competitive snippet-similarity aggregation and co-attentive snippet embedding[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 1169-1178.
- [69] SU C, LI J, ZHANG S, et al. Pose-driven deep convolutional model for person re-identification [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3980-3989.
- [70] LIU J, NI B, YAN Y, et al. Pose transferrable person re-identification [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4099-4108.
- [71] SARFRAZ M S, SCHUMANN A, EBERLE A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 420-429.
- [72] MA A J, LI P. Query based adaptive re-ranking for person re-identification[C]// Proceedings of the 2014 Asian Conference on Computer Vision, LNCS 9007. Cham: Springer, 2014: 397-412.
- [73] GARCÍA J, MARTINEL N, MICHELONI C, et al. Person re-identification ranking optimisation by discriminant context information analysis[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1305-1313.
- [74] LENG Q, HU R, LIANG C, et al. Person re-identification with content and context re-ranking[J]. *Multimedia Tools and Applications*, 2015, 74(17): 6989-7014.
- [75] YE M, LIANG C, YU Y, et al. Person re-identification via ranking aggregation of similarity pulling and dissimilarity pushing [J]. *IEEE Transactions on Multimedia*, 2016, 18(12): 2553-2566.
- [76] ZHONG Z, ZHENG L, CAO D, et al. Re-ranking person re-identification with k-reciprocal encoding[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 3652-3661.
- [77] ZHONG Z, ZHENG L, ZHENG Z, et al. Camera style adaptation for person re-identification [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 5157-5166.
- [78] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2242-2251.
- [79] ZHU Z, HUANG T, SHI B, et al. Progressive pose attention transfer for person image generation[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2342-2351.
- [80] SONG S, ZHANG W, LIU J, et al. Unsupervised person image generation with semantic parsing transformation [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2352-2361.
- [81] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 994-1003.
- [82] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 79-88.
- [83] LI W, ZHAO R, WANG X. Human re-identification with transferred metric learning[C]// Proceedings of the 2012 Asian Conference on Computer Vision, LNCS 7724. Berlin: Springer, 2012: 31-44.
- [84] GRAY D, BRENNAN S, TAO H. Evaluating appearance models for recognition, reacquisition, and tracking [C]// Proceedings of the 2007 IEEE International Workshop on Performance Evaluation for Tracking and Surveillance. Piscataway: IEEE, 2007: 1-7.
- [85] EVERINGHAM M, WINN J. The PASCAL Visual Object Classes challenge 2012 VOC2012) development kit [EB/OL]. [2019-03-20]. http://host.robots.ox.ac.uk/pascal/VOC/voc2012/devkit_doc.pdf.
- [86] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The PASCAL Visual Object Classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [87] PUMAROLA A, AGUDO A, SANFELIU A, et al. Unsupervised person image synthesis in arbitrary poses [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8620-8628.

This work was partially supported by the National Natural Science Foundation of China (61861004, 61762007), the Natural Science Foundation of Guangxi (2017GXNSFAA198267, 2017GXNSFAA198269).

YANG Feng, born in 1979, Ph. D., associate professor. His research interests include artificial intelligence, network information security, big data and high-performance computing, precision medicine.

XU Yu, born in 1993, M. S. candidate. Her research interests include deep learning, pedestrian re-identification.

YIN Mengxiao, born in 1978, Ph. D., associate professor. Her research interests include computer graphics and virtual reality, digital geometry processing, image and video editing, graph theory.

FU Jiacheng, born in 1995, M. S. candidate. His research interests include deep learning, medical image registration.

HUANG Bing, born in 1993, M. S. candidate. Her research interests include medical image fusion.

LIANG Fangxuan, born in 1994, M. S. candidate. Her research interests include deep learning, image segmentation.