

DOI: 10.13382/j.jemi.B2307031

基于姿势估计和特征融合的行人重识别算法^{*}

姬晓飞¹ 赵 帅¹ 宋京浩¹ 崔 童²

(1. 沈阳航空航天大学自动化学院 沈阳 110136; 2. 沈阳航空航天大学人工智能学院 沈阳 110136)

摘 要:行人重识别在交通管理、寻找走失人口等范畴用途较广。现有算法难以处理人体姿势改变、遮挡和特征不对齐的问题,提出一种姿势引导和特征融合的行人重识别算法。所提出的算法包括 3 个分支,包括全局分支、基于姿势估计引导的全局分支、局部对齐分支。全局分支提取行人的全局特征,可以捕捉行人的粗粒度信息以及整体的上下文关系。基于姿势估计引导的全局分支利用姿势估计网络引导模型关注行人的全局可见区域,降低遮挡物对行人识别的干扰。局部对齐分支利用姿势估计算法构成对齐的局部特征,同时区分可见的局部区域,以降低遮挡以及姿势变化的影响。通过多分支结构,将局部特征和全局特征融合,以加强特征的多元化,增强模型的鲁棒性。最终,利用交叉熵和软边界三元损失进行模型训练。Market-1501 和 DukeMTMC-ReID 数据集上的测试结果效验了所提算法的可行性,其间,DukeMTMC-ReID 数据集的 Rank-1、mAP 各达成了 91.2%、81.8%,具有较佳的实用性。

关键词: 行人重识别;深度学习;多损失函数;行人遮挡

中图分类号: TP 391 **文献标识码:** A **国家标准学科分类代码:** 510.40

Person re-identification algorithm based on pose estimationand feature fusion

Ji Xiaofei¹ Zhao Shuai¹ Song Jinghao¹ Cui Tong²

(1. School of Automation, Shenyang Aerospace University, Shenyang 110136, China;
2. School of Artificial Intelligence, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: Person re-identification is highly used in the areas of traffic management, searching for lost people, etc. It is hard for existing algorithms to deal with the problem of human pose change, occlusion and feature misalignment, and a pose-guided and feature-fused pedestrian re-recognition algorithm is proposed. The proposed algorithm includes three branches, including global branch, global branch based on pose estimation guidance, and local alignment branch. The global branch extracts the global features of pedestrians and captures the coarse-grained information of pedestrians. The global branch based on posture estimation guidance uses the posture estimation network guidance model to focus on the global visible area of pedestrians and reduce the interference of occlusion to pedestrian recognition. Local alignment branch utilizes a pose estimation algorithm to establish aligned local features while distinguishing visible local regions to reduce occlusion as well as the influence of postural changes. Through a multi-branch structure, integrated local characteristics with global ones to augment feature diversity is achieved and enhanced model robustness. Finally, network training is conducted using cross-entropy and triplet loss functions. The viability of the proposed algorithm is validated by the test results on Market-1501 and DukeMTMC-ReID datasets, during which the Rank-1 and mAP of the DukeMTMC-ReID dataset reached 91.2% and 81.8%, respectively, which has a better practicality.

Keywords: person re-identification; deep learning; multi-loss functions; person occlusion

0 引言

行人重识别(person re-identification, ReID)也称行人再识别,旨在通过计算机视觉技术在不重迭的监控图像中进行行人匹配的方法^[1-3]。行人重识别技术于智能交通管理、寻找丢失人口等范畴显现了巨大的发展潜力,吸引了世界各地的诸多研究人员的强烈重视。如今这已是计算机视觉的众所关注的研究重点之一。然而,在现实情况中,受摄像机设置、姿势、光照条件、以及遮挡等要素的综合作用,同一行人在采集的不同视频中会展现出显著的外观差别,使得行人重识别的研究依然充满挑战性。因此,提取一个鲁棒性强、具有辨识性的行人特征对行人重识别任务越来越重要。

在实际应用中,完整的行人重识别系统包括行人检测、行人重识别、目标跟踪3个流程。首先需要对输入数据,通过行人检测技术提取包含行人前景的图像,以降低原始监控视频中杂乱的背景信息的影响。因此,行人检测产生的检测框的准确性影响着行人重识别的准确度。如果检测框不准确,裁剪的行人图像可能会包括更多的杂乱的背景信息;以及由于行人的姿势多变,对不精确检测框中的行人直接切分提取特征,会造成特征的不对齐,增加行人重识别的难度。

传统的行人重识别涵盖特征采集和度量学习两个过程,将二者分开进行,二者之间缺乏联系。其中更关键的是特征采集,即捕捉人体具有代表性的特征描述^[4]。该类方法首先通过手工提取低级视觉特征,如梯度直方图^[5]、局部最大出现描述符^[6]描述行人的特征,再次利用度量学习方法,如欧氏距离度量、余弦距离度量方法进行图片的相似度计算^[7],最后进行相似度排序。传统方法的准确性受限于手工特征的设计,特别在复杂现实场景中,手工设计的特征的适应性不佳。此外,传统度量学习方法在处理海量数据时速度慢,限制了实际应用的可能性。

近几年,相随深度学习的快速演进,行人重识别范畴斩获了斐然的研究突破。深度学习的行人重识别策略首要运用卷积神经网络(convolutional neural networks, CNN)直接捕捉行人的外貌特征,获得行人的有效表示,继而计算特征描述之间的相似度^[8]。基于深度学习的方法不需要手工设计特征,CNN可以自适应的捕捉人体特征,学习由浅层到深层的特征,具有更好的泛化能力。并且深度学习的方法将特征提取和度量学习结合到一起,优化网络模型,减少了二者之间的分离性,实现了端到端的行人重识别一体化模型,降低了行人重识别任务的复杂性。截止当前,主要的行人重识别方法专注于直接提取人的整体信息,核心问题是如何获得更富表现力的特

征表达。张晓涵^[9]在卷积过程中利用不同空洞率的卷积核,之后应用注意机制以捕获重要特征,随后将不同分支的特征融合,构建人体的全局特征的表示形式。Chen等^[10]提出将高阶注意模块插入在基线网络中间,关注最显著的全局信息,捕捉图像中行人之间的微弱不同,并且将人的注意力识别问题重新考虑为零镜头学习问题,提出混合高阶注意网络,将全局信息编码到更高的空间中进行分类,以显式方式进一步增强注意力来提升特征的辨别能力和丰富度。许多行人重识别方法关注抽取网络的高层特征,但忽视了中层特征,降低了识别的准确性。因此,Liu等^[11]采用混合池化策略从基线的后两层提取多尺度特征,并采用自注意模块使网络更加关注于非局部特征,捕获更具有判别性的特征。以全局特征为基础的行人重识别方法在处理速度上具有优势,主要学习于行人的整体外观信息。然而,这类方法在处理图像的空间特征时存在制约,易受遮挡、不同姿势、复杂背景等多因素的影响,致使识别的不成功。

针对以上困难,研究人员提出以局部特征的方式达成行人重识别工作。Sun等^[12]提出了PCB网络,将CNN生成的特征图等距划分为6个水平区域,捕获行人的局部特征,单独利用各个局部特征实现行人身份的预测识别。但是,这类传统的分割方法未完全考虑到特征对齐和各个局部特征彼此之间的语义关系,因而致使识别准确性不佳。Zhao等^[13]提出了Spindle Net,通过人体部位的指导对特征图进行分割,之后捕捉不同部位的特征并融合。该算法较好地处理了局部特征的对齐困难。Kalayeh等^[14]提出了SPReID,借助人体的分割模型捕捉局部特征,该方法首先提取前景和几许不同部位的掩膜,之后从掩膜区域学习特征,获得每个区域的局部特征,最后进行特征融合。Huang等^[15]提出了EA-Net,通过姿势估计和语义分割网络对特征图进行分割,提取局部特征并使特征对齐,增强了模型的泛化能力,为解决遮挡问题提供了思路。运用姿势估计和人体分割方法需求更多的标记数据,而且受制于预训练模型性能。与此同时,另外的处理步骤加大了网络的计算量,也增多了干扰信息。学习局部特征有助于捕获丰富的细节特征,在一定程度上处理了遮挡问题,但这种方法忽视了整体信息对局部学习的潜在影响以及切分位置的语义信息。而且当检测框不精确时,直接对图像进行硬性分割获取局部特征,会造成局部特征的错位和不对齐。

针对上述问题,Wang等^[16]提出了一种多粒度网络,全局特征综合多尺度局部特征共同训练网络,涵盖1个全局支路和2个局部支路,通过将特征图水平分割为多个条带,同时调整各局部支路的条纹数,以捕捉多粒度的部分特征描述。在此基础上,Hou等^[17]提出了一种部分上下文感知关联特征的多粒度网络,包括4个分支:一个

顶部消除支路、一个全局支路和两个局部支路;顶部消除支路丢弃顶部激活特征,使网络学习到更多更强的特征;两个局部支路根据人体的结构特征,将图像分别均匀划分为3条和6条,得到多粒度的局部特征表示。孙劲光等^[18]采用ResNeSt-50网络提取行人的多尺度特征,随后将这些特征传递到加权特征金字塔网络,以实现特征融合,在此生成了多尺度特征,涵盖了高层语义信息和低层轮廓特征;最后利用生成的多尺度特征构建全局和局部特征,实现身份的预测。综上所述,全局特征注重获取人体的整体外观,局部特征则关注细节信息,二者相互协同,有助于网络获取更丰富的信息,然而这些方法并不能彻底克服人体遮挡和特征不对齐的困难。

全局特征可以获得人体的粗粒度的外部轮廓信息。并且其处理速度高效。而局部特征有助于获取人体的细节信息,但采用基于姿势估计以及人体分割的方法提取局部特征时,需要引入额外模型,这会提高网络的复杂性并导致训练时间增加。结合全局和局部特征的方案,增多了模型的分支数,类似于引入另外模型,都使网络更繁琐,并增加了训练的难度。

鉴于以上存在的挑战,本文提出了一种姿势引导结合特征融合的行人重识别算法。为解决模型的复杂度较高的挑战,提出以全尺度网络(omni-scale network, OSNet)^[19]作主干网络,该网络将传统的卷积操作分解为点卷积和深度卷积,有效地降低了模型的参数量,并且OSNet网络较简单,可以降低模型的计算量,提升网络的计算效率。本文所提出的算法主要包括3个分支。全局分支通过OSNet网络提取行人的全局特征图,提取行人的整体信息。基于姿势估计引导的全局分支将OSNet和ResNet网络相结合,增加行人的深层特征信息的捕获能力,且利用姿势估计算法引导关注行人的可见深层特征,降低遮挡物对模型的噪声干扰。局部对齐分支利用姿势估计网络提取人体的可见区域特征,以进一步解决行人遮挡问题。同时,为了解决由于姿势多变引起的局部特征不对齐问题,引入局部对齐方法,以确保各个局部特征对齐。采用了全局综合局部特征的策略达成行人重识别,促使全局特征与局部特征的相互弥补。本文对所提方法在公开的Market1501和DukeMTMC-ReID数据集上测试了其可行性。实验数据显示,该算法性能出色,有效地提高行人重识别的准确性。

1 行人重识别算法的总体设计

本文以全尺度网络作为骨干网络捕捉人体的特征描述,其涵盖了五个卷积块conv1、conv2、conv3、conv4以及conv5,使用前4个卷积块作为本文的backbone。本文的网络模型主要包括3个分支,从上至下依次为全局分支、

基于姿势估计引导的全局分支、局部对齐分支。如图1所示。

1) 全局分支:首要将输入图片经过主干网络,以进行特征提取。其次经OSNet的conv5块处理,最终经过广义平均池化(Generalized Meam, GEM)操作。如图1的分支1,全局分支获得粗粒度的特征信息,以及行人的整体信息和轮廓信息。

2) 基于姿势估计引导的全局分支:输入图片首先经过backbone进行特征的提取,之后通过ResNet-50的第四五个卷积块,输出行人的特征图。之后借助姿势估计网络对特征图切分,提取可见的特征,之后通过GEM操作处理。姿态估计选择OpenPose人体姿态识别网络,它表现出卓越的鲁棒性和实时性。如图1中的分支2,可以获得行人的可见特征信息,并捕捉更复杂的抽象特征,去除噪声的干扰。

3) 局部对齐分支:通过姿态估计网络对backbone输出的特征图执行分割操作,生成6个基于姿态引导的局部特征和1个可见全局特征,如图1中的分支3。这一分支旨在处理局部特征对齐问题,增强网络的识别性能。之后利用姿势估计算法获得的关键点信息对人体具有相同语义的部位进行对齐。最后对3个分支输出的特征向量通过降维模块进行降维,计算各个分支的损失。

2 网络结构及实现

2.1 OSNet 网络

当前,行人重识别范畴的大多数模型选择ResNet网络捕捉特征,但其网络参数量较大,计算复杂度高。本文采用全尺度网络作为主干网络,该网络是专为行人重识别任务构建的。其允许网络捕捉多尺度的特征,消除了网络只能捕捉到一个固定感受野的尺度特征的局限。OSNet由5个卷积块和2个过渡层构成。网络结构如表1所示。

表1 OSNet 网络结构
Table 1 OSNet network structure

stage	output	OSNet
Conv1	128×64, 64	7×7 Conv, stride=2
	64×32, 64	3×3 max pool, stride=2
Conv2	64×32, 256	Bottleneck × 2
transition	64×32, 256	1×1 Conv
	32×16, 256	2×2 average pool, stride=2
Conv3	32×16, 384	Bottleneck × 2
Transition	32×16, 384	1×1 Conv
	16×8, 384	2×2 average pool, stride=2
Conv4	16×8, 512	Bottleneck × 2
Conv5	16×8, 512	1×1 Conv
gap	1×1, 512	global average pool
fc	1×1, 512	fc

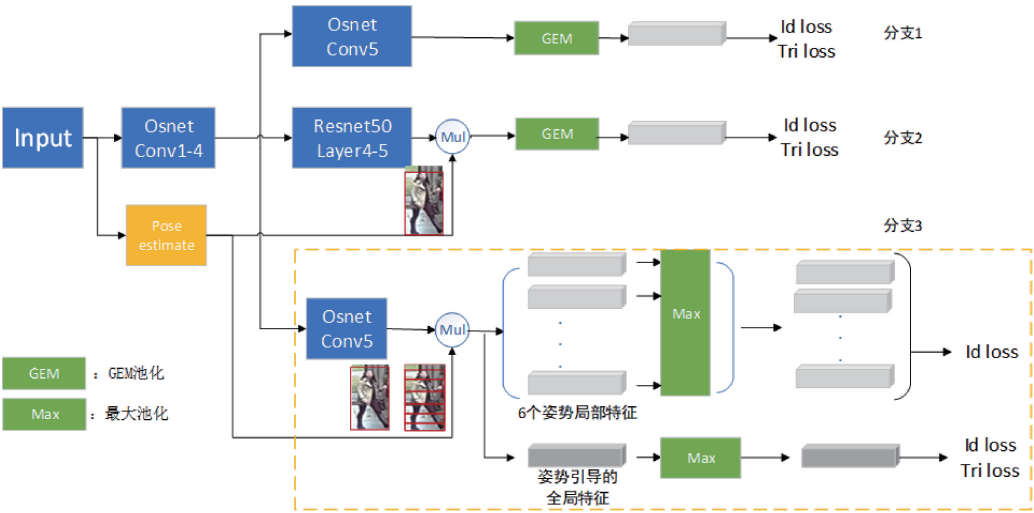


图 1 算法整体框图

Fig. 1 Overall algorithm block diagram

本文将 OSNet 网络与 ResNet-50 网络的参数量、计算量以及时间进行了对比,如表 2 所示。OSNet 网络的参数量相当于 ResNet-50 网络的 1/10,降低了网络过拟合的可能性,计算量接近于 ResNet-50 的 1/3。OSNet 网络的训练时间和推理时间相较于 ResNet 网络都有所提升,其中推理时间提升了 3.78 s。因此综合考虑,选择 OSNet 网络作为本文的主干网络。

表 2 参数量、计算量、时间的对比

Table 2 Comparison of Parameters, FLOPs and time			
网络	OSNet	ResNet-50	
参数量/($\times 10^6$)	2.17	23.51	
计算量/($\times 10^9$)	1.55	4.13	
训练时间(ms/张)	2.54	2.66	
测试时间(s/batch)	19.92	23.70	

2.2 全局分支

输入图片首先经过 backbone 后获得尺寸为 $24 \times 8 \times 512$ 的特征图,之后 OSNet 网络的 conv5 卷积块,得到大小为 $24 \times 8 \times 512$ 的特征图。再次通过 GEM 生成 $1 \times 1 \times 512$ 的全局特征向量 g ,提取较显著的特征。再次经一个线性层、BN 层、ReLU 层组成的降维模块,输出 512 维的特征向量。最终,计算此特征向量的损失。

2.3 基于姿势估计引导的全局分支

以全局特征为依托的行人重识别方法主要聚焦于人的整体外观信息和轮廓的学习。然而,当面对行人被遮挡时,直接从整张图像中提取特征信息容易受到噪声的影响,从而致使网络性能和识别准确性的减弱。因此,本文提出基于姿势估计引导的全局分支,如图 1 中的分支 2。OSNet 网络提取的全局特征较浅,提取到的语义信息不够丰富,限制了重识别的准确性。是以,本文提出使 OSNet 和 ResNet-50 相结合,使得网络捕捉到更深层的语

义信息以及更具多样性的全局特征信息,增强网络重识别的能力。具体的实施步骤是将 OSNet 网络的第五个卷积层替换为 ResNet-50 的第四五个卷积层。Backbone 生成的特征图尺寸为 $24 \times 8 \times 512$,经过 ResNet-50 的 Layer4、Layer5 卷积块,设置 Layer4、Layer5 的步长都为 1,扩张特征图的感受野范畴。Backbone 输出的特征图经过 ResNet-50 的 Layer4、Layer5 模块,生成 $24 \times 8 \times 2048$ 的特征图。之后利用姿势估计网络对获得的特征图进行切割,得到行人未被遮挡的特征图,通过 GEM 池化方式得到一个 2048 维的特征向量。再次通过降维块,使 2048 维的特征向量降维为 512 维的特征向量。之后计算其损失。

2.4 局部对齐分支

监控摄像头拍摄到的环境复杂度高,采用局部特征对人体进行表示可以减少行人姿态、外观变化的影响。之前研究人员提出 PCB 网络^[12],其等距切分人体的特征图以捕捉局部特征,但等距划分的特征存在特征不对齐和人体被遮挡的问题。为了解决这个问题,本文提出局部对齐分支,其根据人体的结构,对人体进行横向切分,提取具有判别性的特征。局部对齐分支首先接收来自 backbone 网络的特征图,尺寸为 $24 \times 8 \times 512$,之后通过 OSNet 网络的 Conv5 卷积块得到 $24 \times 8 \times 512$ 。

局部对齐分支详细的运行步骤如次:通过 OpenPose 算法识别人体的关键点。当关键点置信度大于一定阈值时,认为该关键点是可见的即该部位是可见的;小于阈值时,认为该关键点是不可见的即该部位是被遮挡的。利用姿态估计算法得到的关键点对行人图像进行分割,分割为 6 个局部特征引入姿态估计算法较好地解决了和 1 个可见全局特征,之后进行最大池化操作。6 个局部特

征利用交叉熵损失进行约束,因为三元组损失更聚焦于图像的整体表示,更适用于全局特征的学习,对于局部特征,三元组损失可能会拉大同一类样本之间的距离。可见全局特征利用交叉熵损失和三元组损失联合学习。

对于局部特征不对齐问题,促使池化布满在行人语义相同的区域,即查询对象和图库图像的对应的局部区域。距离的计算采用余弦距离。如果查询对象和图库图像的人体区域全部是可见的,对应位置的特征进行余弦距离的计算,即6个局部区域对应6个局部区域,如图2所示。如果查询对象的某一区域不可见,则在计算距离时,令查询对象和图库图像的对应区域的距离计算为0,即在计算距离的时候去除被遮挡的区域,如图2所示。



图2 局部距离计算方法

Fig. 2 Local distance calculation method

2.5 损失函数设计

为了增强模型的识别力,采用交叉熵和软边界三元组损失对网络进行约束。交叉熵损失有助于增强模型的特征学习能力,常用于处理分类任务。而行人重识别大多被归纳成分类任务。在本文中,交叉熵损失定义为:

$$l_{\text{CrossEntropy}} = - \sum_{i=1}^N q_i \log_2(P_i) \begin{cases} q_i = 0, n \neq i \\ q_i = 1, n = i \end{cases} \quad (1)$$

式中: N 代表训练数据中行人总数; n 代表行人的标签; P_i 代表正确匹配图片行人ID同标签的可能性。

软边界三元组损失公式如:

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

其中, $x_a^{(i)}$ 代表自固定样本捕获的特征向量; $x_p^{(i)}$ 代表自正样本捕获的特征向量; $x_n^{(i)}$ 代表自负样本捕获的特征向量; α 为边缘超参数。

3 实验结果

3.1 数据集

为了验证本文方法的效果,使用 ReID 常用的公开数

据集 Market-150^[20]、DukeMTMC-reID^[21]进行实验。

1) Market-1501 数据集

Market-1501^[20]由6部摄像机拍摄于夏天,共涵盖1501个不同行人,训练集涵盖了751个行人的12936张图像,图库涵盖750个行人的19732张图像。训练集和图库的图像采用DPM检测器自动检测边界,因此,较接近实际应用中的场景。此外,通过手工标记的方式,从测试集中选择了3368张图像,形成了查询集。与此同时,测试集中的一些图片存在一定范围的背景干扰或行人局部区域遮挡,如图3所示。



图3 Market-1501 数据集

Fig. 3 Market-1501 dataset

2) DukeMTMC-ReID 数据集

DukeMTMC-ReID^[21]经由8部高清摄像机拍摄,图像更清晰,共涵盖1812位不同行人,训练集涵盖702个行人的16522张图片,图库涵盖702个行人的17661张图片,查询集涵盖2228张图像。除此之外,该数据集拍摄环境更复杂,图像包括的背景信息更杂乱,且图像之间的相似性较大,还包含了多样的姿势变化以及包含更多行人被遮挡的图像,使得数据集更具有挑战性,如图4所示。



图4 DukeMTMC-ReID 数据集

Fig. 4 DukeMTMC-ReID dataset

3.2 实验设置和评价指标

1) 实验设置

本实验实现基础为 Python 语言和 PyTorch 框架。运行在装备一块 NVIDIA GeForce RTX3060 的服务器上,操作系统是 Windows10。输入图像的大小统一标准化为 384×128,并在训练应用随机水平翻转增强。优化方式选择了 AdamW 方法。batchsize 设置为 32。初始学习率分别调整为 3.5×10^{-4} 和 4×10^{-4} ,backbone 网络的学习率调整为 3.5×10^{-4} ,其他支路调整为 4×10^{-4} 。训练时一共迭代 150 次,到迭代到 60 次的时候,分别下降为 3.5×10^{-5} 和 4×10^{-5} 。最后,迭代到 100 个 epoch 的时候,学习率分别下降为 3.5×10^{-6} 和 4×10^{-6} 。

2)评价指标

采用 Rank-1 和 mAP 评价模型性能。Rank-1 是指搜索结果排名第一的图像的行人与查询图像是一个人的概率。mAP 是一个综合指标,考量了排序列表中匹配正确图像的平均排名。与 Rank-1 对比,mAP 在评估 ReID 算法性能时更具全面性。计算公式为:

$$mAP = \frac{\sum_{k=0}^c AP_k}{C}$$

(3)

其中, $\sum_{k=0}^c AP_k$ 表示将每一个类型的平均精度 AP 合并,C 表示类别数总和。

3.3 消融实验

为研究多分支网络的各支路对网络模型整体性能的作用,在 DukeMTMC 数据集上进行了验证测试。表 3 展示了不同分支组合下模型的整体性能水平。如表 3 结果所示,在 DukeMTMC 数据集,多分支组成较于单分支结构表现出明显的性能改善,Rank-1 和 mAP 分别提升了 2%和 7.7%。根据表 3 结果,在双支路结构中,姿势引导的全局分支与局部对齐分支融合,效果最为显著。在基于姿势引导的全局分支和局部对齐分支结合的基础上,引入了全局分支,使得识别的准确率进一步得到了提升,得到了最优地测试结果。基于姿势引导的全局分支提取可见的人体的深层全局信息,获得较深的语义信息,增加特征的多样性,并降低了遮挡物的干扰,解决人体遮挡的问题;局部对齐分支专注于细粒度特征的提取,有效解决了遮挡问题,并采用特征对齐技术,降低局部特征不对齐的阻碍,增强网络的特征提取能力;在局部对齐分支和基于姿势引导的全局分支的基础上,全局分支可以提取图像的整体轮廓信息,令网络关注到人体的细节信息以及整体的上下文信息,捕捉到更充足有效的特征描述。

实验数据证明,多支路的联合是有效的,不仅学习到了整体的轮廓信息和局部的细粒度特征信息,还去除了遮挡物的影响,提升了网络的识别能力和准确度。多分支网络结构相与协同,增强对复杂特征的捕捉能力,提高了模型的鲁棒性。

表 3 各个分支对网络性能的对比

Table 3 Comparison of network performance between different branches

方法	DukeMTMC	
	Rank-1/%	mAP/%
全局分支	89.2	74.1
基于姿势引导的全局分支	89.5	74.5
局部对齐分支	89.8	76.5
基于姿势引导的全局分支+局部对齐分支	90.8	81.6
全局分支+局部对齐分支	90.5	81.1
本文	91.2	81.8

3.4 与其他方法对比

如表 4 所给出,在两个数据集上把本文所提出的解决方法与其他办法加以比较。其中,Luo^[22]提出了一个强基线网络,即 BoT(bag of tricks)网络,集合了众多行之有效的训练方法。OSNet 是一个为行人重识别构建的轻量化网络^[19]。Zheng 等^[23]提出了 Pyramid,是一个金字塔模型,主要捕捉多个粒度的特征信息。Cheng 等^[24]提出了 SCR (spatial and channel partition representation Network),是全局、局部特征共同训练的网络模型。

表 4 与 Market-1501 和 DukeMTMC 上与最先进方法对比

Table 4 Comparison with state-of-the-art methods on Market-1501 and DukeMTMC

Method	Market-1501		DukeMTMC	
	Rank-1/%	mAP/%	Rank-1/%	mAP/%
BoT ^[22]	94.5	85.9	86.4	76.4
OSNet ^[19]	94.8	84.9	88.6	73.5
Pyramid ^[23]	95.7	88.2	89.0	79.0
SCR ^[24]	95.7	89.0	91.1	81.4
本文	96.1	90.0	91.2	81.8

如表 4 显示,本文方法在 Market-1501 数据集,相比与 OSNet^[19]的 Rank-1、mAP,本文方法中的 Rank-1 增加了 1.3%,mAP 增加了 5.1%。对于 DukeMTMC 数据集,与 OSNet 的 Rank-1、mAP 对比,本文的方法取得了更好的表现,本方法的 Rank-1 增加了 2.6%,mAP 增加了 8.3%。

与采用金字塔模型的 Pyramid 网络^[23]相比,本文的方法在 Market-1501 上取得了更好的性能,相比于 Pyramid 的 Rank-1、mAP,本方法的 Rank-1 增加了 0.4%,mAP 增加了 1.8%。在 DukeMTMC 上,相对于 Pyramid 的 Rank-1、mAP,本方法的 Rank-1 增加了 2.2%,mAP 增加了 2.8%。

相对于全局联合局部特征共同训练的 SCR 模型^[24],本文用的方法于 Market-1501 表现得更出色,本方法的 Rank-1 增加了 0.4 个%,mAP 增加了 1%。对于 DukeMTMC 数据集上,相较于 SCR 网络的 Rank-1,本文

的方法表现出略微的提升,本办法将 Rank-1 增加了 0.1%,mAP 增加了 0.4%。综合表 4 中的数据,本文方法相较于其他最先进的方法,具有一定的竞争优势。

4 结 论

本文提出了一种基于姿势估计和特征融合的行人重识别方法,旨在解决行人遮挡以和特征不对齐的挑战。该方法以 OSNet 为主干网络,包括 3 个核心部分,分别是全局分支、姿势引导的全局分支和局部对齐分支。全局分支捕捉人的粗粒度特征以及整体轮廓。基于姿势估计引导的全局分支提取可见的人体特征信息,其可以获得更深层的语义信息和多样性的可见人体全局信息,降低了遮挡物对网络进行识别的干扰;局部对齐分支依据姿势估计算法对人体进行切割,学习更细粒度的信息,并且进一步去除了遮挡物对人体的影响,去除了噪声对网络的干扰,并对局部特征进行特征对齐,降低了局部特征不对齐造成的识别错误率。使用多分支的网络可以使网络学习到更加多样性的信息,获取互补特征。在两个公开的数据集的测试结果论证了,本文提出的基于姿势引导和特征融合的行人重识别算法表现了出色的识别能力,可以与当前领域最先进的方法进行媲美。综上所述,基于姿势估计网络引导的方法虽然有效的提高了行人重识别的准确性,解决了行人遮挡、姿势改变的问题。但是引入了额外的模型,增加了网络的收敛难度以及网络中的噪声信息且限制了在实际应用中的可能性。因此,下一阶段的研究将继续探索相关资料,开发一种鲁棒性强、高效的行人重识别算法,能够在不引入额外模型的前提下,有效地解决遮挡和姿势变化的问题。

参考文献

- [1] 杨永胜, 邓森磊, 李磊, 等. 基于深度学习的行人重识别综述[J]. 计算机工程与应用, 2022, 58(9): 51-66.
YANG Y SH, DENG M L, LI L, et al. A review of pedestrian re-identification based on deep learning[J]. Computer Engineering and Applications, 2022, 58(9): 51-66.
- [2] 张勃兴, 马敬奇, 张寿明, 等. 利用全局与局部关联特征的行人重识别方法[J]. 电子测量与仪器学报, 2022, 36(6): 205-212.
Zhang B X, Ma J Q, Zhang SH M, et al. A pedestrian reidentification method using global and local association features [J]. Journal of Electronic Measurement and Instrument, 2022, 36 (6): 205-212.
- [3] 刘慧, 梁东升, 张雷, 等. 全局注意力机制与关系网络驱动的行人重识别[J]. 中国科技论文, 2023, 18(7):

759-765, 785.

- LIU H, LIANG D SH, ZHANG L, et al. Pedestrian re-identification driven by global attention mechanism and relation network [J]. China Science and Technology Paper, 2023, 18(7): 759-765, 785.
- [4] 潘凤, 王杰, 张艳莎, 等. 基于双分支特征拼接的行人重识别[J]. 计算机与现代化, 2023(5): 93-99.
PAN F, WANG J, ZHANG Y SH, et al. Pedestrian re-identification based on dual-branch feature concatenation[J]. Computer and Modernization, 2023(5): 93-99.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, 1: 886-893.
- [6] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2197-2206.
- [7] KOESTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012: 2288-2295.
- [8] 罗浩. 深度学习时代的行人重识别技术[J]. 人工智能, 2019(2): 40-49.
LUO H. Pedestrian re-identification technology in the era of deep learning [J]. Artificial Intelligence, 2019(2): 40-49.
- [9] 张晓涵. 基于全局特征改进的行人重识别. 计算机系统应用, 2022, 31(5): 298-303.
ZHANG X H. Pedestrian re-identification improved with global features [J]. Computer Systems & Applications, 2022, 31(5): 298-303.
- [10] Chen B, Deng W, Hu J. Mixed high-order attention network for person re-identification [J]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 371-381.
- [11] LIU H Y. Multi-level feature fusion and multi-loss learning for person re-identification [J]. Signal Processing. Image Communication: A Publication of the European Association for Signal Processing, 2021, 94(1): 116197.
- [12] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) [J]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 501-518.

- [13] ZHAO H, TIAN M, SUN S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017. DOI: 10.1109/CVPR.2017.103.
- [14] KALAYEH M M, BASARAN E, GÖKMEN M, et al. Human semantic parsing for person re-identification[J]. Proceedings of 2018IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018. 1062-1071.
- [15] HUANG H, YANG W, CHEN X, et al. EANet: Enhancing alignment for cross-domain person re-identification [J]. 2018, DOI: 10.48550/arXiv.1812.1139.
- [16] WANG G S, YUAN Y F, CHEN X, et al. Learning discriminative features with multiple granularities for person reidentification[J]. In: Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM, 2018:274-282.
- [17] HOU Y, CHEN C, LI Y. Efficient multi-granularity network based on local context-aware correlation feature for person re-identification[C]. 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021:553-556.
- [18] 孙劲光, 吴明岩. 基于多尺度加权特征融合的行人重识别方法研究[J]. 信号处理, 2022, 38(10): 2201-2210.
SUN J G, WU M Y. Research on pedestrian re-identification method based on multi-scale weighted feature fusion[J]. Signal Processing, 2022, 38(10): 2201-2210.
- [19] ZHOU K, YANG Y, CAVALLARO A, et al. Omni-scale feature learning for person re-identification [J]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3702-3712.
- [20] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1116-1124.
- [21] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi camera tracking[C]. Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II, 2016: 17-35.
- [22] LUO H. Bags of tricks and a strong baseline for deep

person re-identification [J]. IEEE, 2019. DOI: 10.1109/CVPRW.2019.00190.

- [23] ZHENG F, DENG C, SUN X, et al. Pyramidal person re-identification via multi-loss dynamic training [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8514- 8522.
- [24] CHENG R, WANG L, WEI M. Learning discriminative and generalizable features with multi-branch for person re-identification [J]. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2022, 42(6): 5987-6001.

作者简介



姬晓飞, 博士, 副教授, 主要研究方向为视频分析与处理、模式识别理论等。

E-mail: jixiaofei7804@126.com

Ji Xiaofei, Ph. D., associate professor and M. Sc. supervisor at Shenyang Aerospace University. Her main research interests include

video analysis and pattern recognition theory, etc.



赵帅, 2020 年于沈阳大学获得学士学位, 现为沈阳航空航天大学在读硕士研究生, 主要研究方向为图像处理和模式识别。

E-mail: zhaoshuai1234500@163.com

Zhao Shuai received a B. Sc. degree from Shenyang University in 2020. She is now a M. Sc. student at Shenyang Aerospace University. Her main research interests include image processing and pattern recognition.



宋京浩, 2022 年于沈阳航空航天大学获得学士学位, 现为沈阳航空航天大学在读硕士研究生, 主要研究方向为图像处理和模式识别。

E-mail: sjh1059725521@163.com

Song Jinghao received a B. Sc. degree from Shenyang Aerospace University in 2022. He is now a M. Sc. candidate at Shenyang Aerospace University. His main research interests include image processing and pattern recognition.



崔童, 博士, 研究兴趣包括散射环境下的图像恢复、计算机视觉和机器学习。

E-mail: ct61ct61@126.com

Cui Tong received the Ph. D. in Pattern Recognition and Intelligent Systems from the University of Chinese Academy of Sciences. His main research interests include image restoration in scattering environments, computer vision, and machine learning.