JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2022
07
VOL.27

ISSN1006-8961
CN11-3758/TB

舰船目标检测 *P2078*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第27卷第7期（总第315期）
2022年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@aircas.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2022年7月 第27卷 第7期（总第315期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

深度学习背景下视觉显著性物体检测综述(第2112页)

提取全局语义信息的场景图生成算法(第2214页)

融合弱监督目标定位的细粒度小样本学习(第2226页)

Volume 27, Number 7
Published July 16, 2022

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Review of deep learning based salient object detection(P2112)

Global semantic information extraction based scene graph generation algorithm(P2214)

Weakly-supervised object localization based fine-grained few-shot learning(P2226)

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2022)07-2263-11

论文引用格式:Xu S, Zheng F, Tang J and Bao W X. 2022. Dual branch feature fusion network based gait recognition algorithm. Journal of Image and Graphics,27(07):2263-2273(徐硕,郑锋,唐俊,鲍文霞. 2022. 双分支特征融合网络的步态识别算法. 中国图象图形学报,27(07):2263-2273)

[DOI:10.11834/jig.200730]

双分支特征融合网络的步态识别算法

徐硕[1],郑锋[2],唐俊[1*],鲍文霞[1]

1. 安徽大学电子信息工程学院,合肥　230601; 2. 南方科技大学工学院,深圳　518055

摘　要:**目的**　在步态识别算法中,基于外观的方法准确率高且易于实施,但对外观变化敏感;基于模型的方法对外观变化更加鲁棒,但建模困难且准确率较低。为了使步态识别算法在获得高准确率的同时对外观变化具有更好的鲁棒性,提出了一种双分支网络融合外观特征和姿态特征,以结合两种方法的优点。**方法**　双分支网络模型包含外观和姿态两条分支,外观分支采用 GaitSet 网络从轮廓图像中提取外观特征;姿态分支采用 5 层卷积网络从姿态骨架中提取姿态特征。在此基础上构建特征融合模块,融合外观特征和姿态特征,并引入通道注意力机制实现任意尺寸的特征融合,设计的模块结构使其能够在融合过程中抑制特征中的噪声。最后将融合后的步态特征应用于识别行人身份。**结果**　实验在 CASIA-B(Institute of Automation, Chinese Academy of Sciences, Gait Dataset B)数据集上通过跨视角和不同行走状态两种实验设置与目前主流的步态识别算法进行对比,并以 Rank-1 准确率作为评价指标。在跨视角实验设置的 MT(medium-sample training)划分中,该算法在 3 种行走状态下的准确率分别为 93.4%、84.8% 和 70.9%,相比性能第 2 的算法分别提升了 1.4%、0.5% 和 8.4%;在不同行走状态实验设置中,该算法在两种行走状态下的准确率分别为 94.9% 和 90.0%,获得了最佳性能。**结论**　在能够同时获取外观数据和姿态数据的场景下,该算法能够有效地融合外观信息和姿态信息,在获得更丰富的步态特征的同时降低了外观变化对步态特征的影响,提高了步态识别的性能。

关键词:生物特征识别;步态识别;特征融合;双分支网络;SE 模块;人体姿态估计;步态轮廓图像

Dual branch feature fusion network based gait recognition algorithm

Xu Shuo[1], Zheng Feng[2], Tang Jun[1*], Bao Wenxia[1]

1. School of Electronics and Information Engineering, Anhui University, Hefei 230601, China;

2. College of Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Abstract: **Objective** Gait is a kind of human walking pattern, which is one of the key biometric features for person identification. As a non-contact and long-distance recognition way to capture human identity information, gait recognition has been developed in video surveillance and public security. Gait recognition algorithms can be segmented into two mainstreams like appearance-based methods and the model-based methods. The appearance-based methods extract gait from a sequence of silhouette images in common. However, the appearance-based methods are basically affected by appearance changes like non-rigid clothing deformation and background clutters. Different from the appearance-based methods, the model-based methods commonly leverage body structure or motion prior to model gait pattern and more robust to appearance variations.

收稿日期:2020-12-14;**修回日期**:2021-04-13;**预印本日期**:2021-04-20

*通信作者:唐俊　tangjunahu@163.com

基金项目:国家自然科学基金项目(61772032);国家重点研发计划资助(SQ2018YFC080102);安徽省重点研发计划资助(202004a7020050)

Supported by: National Natural Science Foundation of China (61772032); National Key R&D Program of China (SQ2018YFC080102); Anhui Provincial Key Research and Development Project(202004a07020050)

Actually, it is challenged to identify a universal model for gait description, and the previous pre-defined models can be constrained in certain scenarios. Recent model-based methods are focused on deep learning-based pose estimation to model key-points of human body. But the estimated pose model constrains the redundant noises in subject to pose estimators and occlusion. In summary, the appearance-based methods are based visual features description while the model-based methods tend to describe a semantic level-based motion and structure. We aim to design a novel approach for gait recognition beyond the existed two methods mentioned above and improve gait recognition ability via the added appearance features and pose features. **Method** we design a dual-branch network for gait recognition. The input data are fed into a dual-branch network to extract appearance features and pose features each. Then, the two kinds of features are merged into the final gait features in the context of feature fusion module. In detail, we adopt an optimal network GaitSet as the appearance branch to extract appearance features from silhouette images and design a two-stream convolutional neural network (CNN) to extract pose features from pose key-points based on the position information and motion information. Meanwhile, a squeeze-and-excitation feature fusion module (SEFM) is designed to merge two kinds of features via the weights of two kinds of features learning. In the squeeze step, appearance feature maps and pose feature maps are integrated via pooling, concatenation, and projection. In the excitation step, we obtain the weighted feature maps of appearance and pose via projection and Hadamard product. The two kinds of feature maps are down-sampled and concatenated into the final gait feature in accordance with adaptive weighting. To verify the appearance features and pose features, we design two variants of SEFM in related to SEFM-A and SEFM-P further. The SEFM module merges appearance features and pose features in mutual; the SEFM-A module merges pose features into appearance features and appearance features remain unchanged; the SEFM-P module merges appearance features into pose features and no pose features changed. Our algorithm is based on Pytorch and the evaluation is carried out on database CASIA (Institute of Automation, Chinese Academy of Sciences) Gait Dataset B (CASIA-B). We adopt the AlphaPose algorithm to extract pose key-points from origin RGB videos, and use silhouette images obtained. In each iteration of the training process, we randomly select 16 subjects and select 8 random samples of each subject further. Every sample of them contains a sub-sequence of 30 frames. Consequently, each batch has 3 840 image-skeleton pairs. We adopt the Adam optimizer to optimize the network for 60 000 iterations. The initial learning rate is set to 0.000 2 for the pose branch, and 0.000 1 for the appearance branch and the SEFM, and then the learning rate is cut10 times at the 45 000-th iteration. **Result** We first verify the effectiveness of the dual-branch network and feature fusion modules. Our demonstration illustrates that our dual-branch network can enhance performance and there is a clear complementary effect between appearance features and pose features. The Rank-1 accuracies of five feature fusion modules like SEFM, SEFM-A, SEFM-P, Concatenation, and multi-modal transfer module (MMTM) are 83.5%, 81.9%, 93.4%, 92.6% and 79.5%, respectively. These results demonstrate that appearance features are more discriminative because there are noises existed in pose features. Our SEFM-P is capable to merge two features in the feature fusion procedure via noises suppression. Then, we compare our methods to advanced gait recognition methods like CNNs, event-based gait recognition (EV-Gait), GaitSet, and PoseGait. We conduct the experiments with two protocols and evaluate the rank-1 accuracy of three walking scenarios in the context of normal walking, bag-carrying, and coat-wearing. Our method archives the best performance in all experimental protocols. Our three scenarios-based rank-1 accuracies are reached 93.4%, 84.8%, and 70.9% in protocol 1. The results of protocol 2 are obtained by 95.7%, 87.8%, 77.0%, respectively. Comparing to the second-best method of GaitSet, the rank-1 accuracies in the context of coat-wearing walking scenario are improved by 8.4% and 6.6%. **Conclusion** We harness a novel gait recognition network based on the fusions of appearance features and pose features. Our analyzed results demonstrated that our method can develop two kinds of features and the appearance variations is more robust, especially for clothing changes scenario.

Key words: biometric recognition; gait recognition; feature fusion; two-branch network; squeeze-and-excitation module; human body pose estimation; gait silhouette images

0 引 言

步态识别旨在通过行人行走的模式判断其身份,具有远距离、易采集、不易模仿和伪装等优点(贲晛烨 等,2012),在视频监控、公共安全等领域有着广阔的应用前景。步态识别方法分为基于外观的方法和基于模型的方法。基于外观的方法使用轮廓

图像作为输入数据并从中提取步态特征。轮廓图像的优点是易于获取和处理,且去除了背景和人体纹理信息等干扰因素,更专注于步态。基于模型的方法先通过输入数据对步态进行建模,如关节角度、运动轨迹和姿态等,再通过模型提取步态特征。

目前,步态识别尤其是基于深度学习的方法大多采用基于外观的方法。基于外观的方法一般面临两个难点,一是轮廓图像在不同视角下差异较大;二是轮廓图像在不同的行走状态下差异较大,如携带背包、衣物变化等干扰。对于视角的差异,Yu 等人(2019)利用生成对抗网络将各个视角的轮廓图像都转换至相同的视角,再提取特征以提高对视角变化的鲁棒性。Wang 等人(2019)利用动态视觉传感器获取数据并通过运动一致性去除噪声,再使用卷积神经网络提取步态特征。Zhang 等人(2019a)结合同步态和跨步态两种网络的优势,设计了一种联合网络提取步态特征。Ben 等人(2019a,2020)针对跨视角问题提出了一系列步态识别算法,如提出一种通用张量表示框架从跨视角步态张量数据中进行耦合度量学习、提出耦合双线性判别投影算法以跨视角对齐步态图像以及提出了耦合区块对齐算法。对于行走状态的差异,大多数方法都不做针对性处理,仅靠卷积神经网络自身提取尽可能与行走状态无关的步态特征。由于输入数据为轮廓图像序列,因此时序特征的融合也是研究的关注点。Wu 等人(2017)利用传统算法,先将轮廓图像序列融合为一幅特征模板图像,再将该模板送入卷积网络提取特征,并针对不同场景设计了不同的图像预处理方法和网络结构,该方法的优点是计算简单,仅需要从一幅步态模板图像中提取特征。Wolf 等人(2016)利用 3D 卷积网络从轮廓图像序列中提取时空特征,能够更充分地融合时序上的信息,但计算量较大且 3D 卷积网络对序列的长度有限制。Chao 等人(2019)设计新的卷积网络从轮廓图像序列的每幅图像中提取空间特征,再通过时序池化和特征融合的方法得到时空步态特征,该方法网络结构简单且融合效果良好。在此基础上,Chao 等人(2021)提出新的特征融合方法并改进训练策略,进一步提高算法的识别准确率;Zhang 等人(2019b)利用自动编码器将序列中的步态特征解耦为外观特征和姿态特征,再用长短时记忆网络融合时序信息生成步态特征。

与基于外观的方法相比,基于模型的方法较少。由于早期方法没有建立合适的模型,导致识别准确率与基于外观的方法相比有很大差距。另一方面早期方法构建的模型仅适用于严格条件限制下的场景,因此泛化性能较差。随着姿态估计算法的进步,可以利用已有的姿态估计器获取较为准确的姿态信息,这为基于模型的方法提供了新的思路。Liao 等人(2017)利用姿态估计方法提取 2D 姿态关键点,并使用卷积网络提取步态特征;之后进一步使用 3D 姿态关键点(Liao 等, 2020),提高了算法对视角场景下的准确率,同时结合人体姿态先验如上下肢的运动关系、运动轨迹等提取步态特征。此外,也有部分工作利用深度相机、雷达等获取 3D 姿态骨架并从中提取步态特征(Kastaniotis 等,2015;Sadeghzadehyazdi 等,2019)。相比之下,虽然这些方法获取的姿态信息更加准确,但需要额外的硬件设备,不便于实际应用。

综上所述,基于外观的方法和基于模型的方法各有优缺点。基于外观的方法效果较好且步骤简单,但易受外界因素干扰,如人体外观变化、视角变化等,这些因素导致轮廓图像改变,进而影响识别准确率;基于模型的方法对外观变化更加鲁棒,但在建模过程中丢弃了外观信息,导致可用信息减少,而且姿态准确性受姿态估计算法的限制,识别准确率与基于外观的方法仍有一定差距。基于此,融合外观与模型两种方法有助于进一步提高步态识别的准确率。目前,与此相似的研究是多模态特征融合(Vaezi Joze 等,2020),由于外观和模型两种方法的网络结构和特征维度有较大差异,常见的特征融合方法如特征相加、张量积等并不能直接用于步态识别算法。针对以上问题,本文设计了一种基于特征融合的步态识别算法,使用双分支网络,两条分支分别用于提取外观特征和姿态特征,最后利用特征融合模块,将两种特征自适应地融合以发挥两种特征间的互补性。本文算法的创新之处在于:1)设计了一种双分支卷积神经网络,利用外观和姿态两种数据分别提取外观特征和姿态特征,并进一步融合以得到更准确更鲁棒的步态特征,从而达到更高的准确率;2)设计了一种新的特征融合模块,能够有效利用外观特征和姿态特征间的互补性,且适用于两种特征维度差距较大的情形;3)在 CASIA-B(Institute of Automation, Chinese Academy of Sciences, Gait Dataset B)数据集上与主流方法进行实验对比,验证了算法的有效性。

1 本文网络模型

本文算法的总体结构如图 1 所示,输入数据为原始 RGB 视频序列,通过背景分割算法得到轮廓图像序列,通过姿态估计算法获得姿态关键点序列。轮廓图像数据通过外观分支网络得到外观特征,姿态关键点数据通过姿态分支网络得到姿态特征。之后两种特征通过特征融合模块得到最后的步态特征。在测试时将步态识别看做检索问题,根据样本间步态特征的距离判断其相似性。

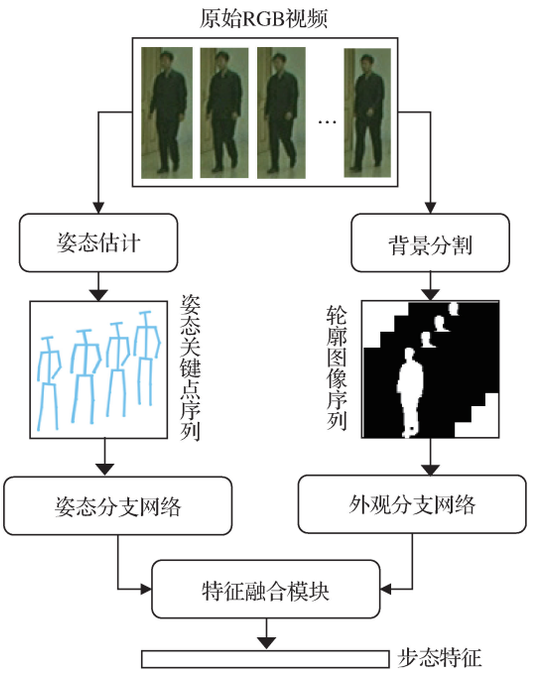


图 1 本文网络框架图

Fig. 1 The pipeline of the our network

1.1 数据预处理

由于原始视频中行人行走过程与相机距离在不断变化,一个序列中轮廓图像的分辨率和姿态骨架的尺寸也在改变,为了避免数据的影响,需要先对输入数据标准化。令原始 RGB 视频序列为 $\boldsymbol{I}=\{\boldsymbol{I}_1,\boldsymbol{I}_2,\cdots,\boldsymbol{I}_T\}$,$\boldsymbol{I}$ 表示单幅 RGB 图像,T 表示序列中的图像数量。视频序列 $\boldsymbol{I}$ 经过 AlphaPose 姿态提取器(Fang 等,2017)得到姿态骨架序列 $\{\boldsymbol{p}_1^N,\boldsymbol{p}_2^N,\cdots,\boldsymbol{p}_T^N\}$,$\boldsymbol{p}$ 表示单个姿态骨架,N 表示每个姿态骨架包含的关键点的数量;经过背景分割算法得到轮廓图像序列 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T\}$,$\boldsymbol{x}$ 表示单幅轮廓图像。

对于姿态数据,将一个序列中第 t 帧的姿态骨架记为 $\boldsymbol{p}_t^N=\{\boldsymbol{v}_t^i\mid i=1,2,\cdots,N\}$,其中 $\boldsymbol{v}_t^i$ 表示第 t 帧中第 i 个关键点的坐标。在姿态骨架中,颈部关键点和臀部关键点相对稳定,因此,采用颈部关键点和臀部关键点间的距离作为标准距离,标准化过程可表示为

$$\bar{\boldsymbol{p}}=\frac{\boldsymbol{p}-\boldsymbol{v}^{\text{neck}}}{dist(\boldsymbol{v}^{\text{neck}},\boldsymbol{v}^{\text{hip}})} \tag{1}$$

式中,$\boldsymbol{p}$ 表示原始姿态骨架,$\bar{\boldsymbol{p}}$ 表示标准化后的姿态骨架,$\boldsymbol{v}^{\text{neck}}$ 和 $\boldsymbol{v}^{\text{hip}}$ 分别表示颈部关键点和臀部关键点的坐标,$dist(\cdot)$ 表示欧氏距离函数。

对于轮廓图像,首先按照边界去除无用的背景,只保留包含轮廓图像的区域;再将轮廓图像保持宽高比例不变的同时将高度缩放到 64 像素,并将图像的宽度左右平均填充到 44 像素,最后对所有轮廓图像进行标准化处理。

1.2 外观分支网络

本文算法采用 GaitSet(Chao 等,2019)作为外观分支网络,结构如图 2 所示。网络主干由 6 层卷积层组成,输入轮廓图像序列后得到特征图序列。该网络的分支用于融合网络浅层与深层的特征,图 2 中的池化操作表示在时序上对序列进行池化,⊕表示对应元素相加,因此分支网络的另一作用是融合时序信息。经过主干与分支两条路径后得到两种特征图,为了便于后期的特征融合,添加了拼接操作将两种特征图在通道维度上拼接,经过池化层下采样和全连接层上采样后得到外观特征。

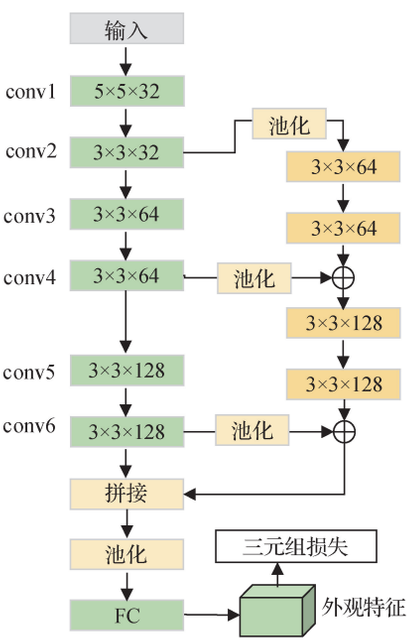


图 2 外观分支网络结构

Fig. 2 Architecture of the appearance branch network

1.3　姿态分支网络

受 HCN(hierarchical co-occurrence network)(Li 等,2018)启发,算法设计了一种用于处理姿态关键点的卷积网络,如图 3 所示。该网络由 5 层卷积层组成,其中前 3 层分为位置分支和运动分支两条网络;位置分支从姿态骨架序列 $\boldsymbol{P}=\{\boldsymbol{p}_1^N,\boldsymbol{p}_2^N,\cdots,\boldsymbol{p}_T^N\}$ 中提取特征,运动分支从骨架运动序列 $\boldsymbol{M}=\{\boldsymbol{p}_1^N-\boldsymbol{p}_2^N,\cdots,\boldsymbol{p}_T^N-\boldsymbol{p}_{T-1}^N\}$ 提取特征。两条分支的结构相同但不共享网络参数,在训练时两条分支相互独立。位置和运动两条分支的特征图沿维度拼接后送入后续的卷积层进一步提取特征。该网络采用 2 维卷积层,分别在姿态关键点的空间和时序上进行卷积操作,因此需要考虑到序列长度对时序卷积结果的影响。为了处理不同长度的序列,在卷积层后添加了时序池化步骤融合时序特征,得到最后的姿态特征。

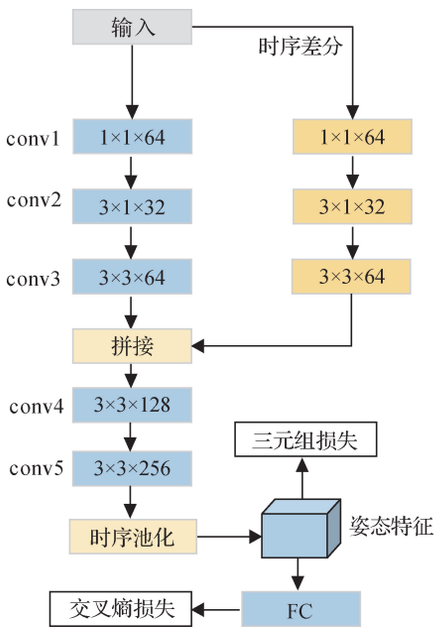


图 3　姿态分支网络结构

Fig. 3　Architecture of the pose branch network

1.4　特征融合模块

输入数据经过外观分支网络和姿态分支网络后,分别得到外观特征图 $\boldsymbol{f}_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}_{\mathrm{P}}$。其中,$\boldsymbol{f}_{\mathrm{A}}\in\mathbf{R}^{h_1\times w_1\times c_1}$,$\boldsymbol{f}_{\mathrm{P}}\in\mathbf{R}^{h_2\times w_2\times c_2}$。$h_1$、$w_1$、$c_1$ 和 h_2、w_2、c_2 分别表示外观特征图和姿态特征图的高度、宽度和通道数。外观特征图 $\boldsymbol{f}_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}_{\mathrm{P}}$ 经过池化融合后得到最终的外观特征 $\boldsymbol{f}_{\mathrm{a}}$ 和姿态特征 $\boldsymbol{f}_{\mathrm{p}}$。特征融合常见的方法有相加、加权平均、张量积和拼接等,在本文算法的双分支网络中,由于外观分支网络与姿态分支网络结构不同,特征维度有很大差距。在实施中,外观特征 $\boldsymbol{f}_{\mathrm{a}}\in\mathbf{R}^{15\,872\times1}$,而姿态特征 $\boldsymbol{f}_{\mathrm{p}}\in\mathbf{R}^{1\,024\times1}$,因此相加、加权平均并不适用于该算法的网络,同时外观特征维度较高导致张量积的计算量很大。相比之下,特征拼接的融合方法简单且适用于该算法的双分支网络,但仅使用特征拼接不能充分利用两种特征的互补性。

受 MMTM(multi-modal transfer module)(Vaezi Joze 等,2020)利用 SE(squeeze excitation)模块处理多模态模型中不同尺寸特征图的启发,本文算法设计了一种基于 SE 模块(Hu 等,2018)的特征融合模块(squeeze-and-excitation feature fusion module,SEFM),结构图如图 4(a)所示。SEFM 模块使用 SE 模块的结构以融合外观和姿态两种特征。具体融合方法为:

1)对外观特征图 $\boldsymbol{f}_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}_{\mathrm{P}}$ 分别池化下采样,得到全局外观特征 $\boldsymbol{f}'_{\mathrm{a}}\in\mathbf{R}^{1\times1\times c_1}$ 和全局姿态特征 $\boldsymbol{f}'_{\mathrm{p}}\in\mathbf{R}^{1\times1\times c_2}$,将两种特征拼接后得到拼接特征 $\boldsymbol{f}'_{\mathrm{c}}\in\mathbf{R}^{1\times1\times(c_1+c_2)}$,这里采用了全局最大池化将特征图的空间信息压缩到通道维度中,以便于融合不同尺寸的特征图。

2)将拼接特征 $\boldsymbol{f}'_{\mathrm{c}}$ 映射到低维空间,得到融合特征 $\boldsymbol{f}'_{\mathrm{m}}$,该过程可表示为

$$\boldsymbol{f}'_{\mathrm{m}}=\boldsymbol{W}\boldsymbol{f}'_{\mathrm{c}}+b \tag{2}$$

式中,$\boldsymbol{W}$ 表示权重,b 表示偏差;$\boldsymbol{f}'_{\mathrm{m}}=\mathbf{R}^{1\times1\times c_{\mathrm{e}}}$,$c_{\mathrm{e}}$ 表示映射后的特征通道数,该过程称为压缩(squeeze),算法实施中 $c_{\mathrm{e}}=(c_1+c_2)/2$。

3)将融合特征分别映射到全局外观特征空间和全局姿态特征空间,该过程称为激励(excitation),得到外观特征图和姿态特征图的激励值 $\boldsymbol{e}_{\mathrm{a}}$ 和 $\boldsymbol{e}_{\mathrm{p}}$,可表示为

$$\boldsymbol{e}_{\mathrm{a}}=\sigma(\boldsymbol{W}\boldsymbol{f}'_{\mathrm{m}}+b),\quad \boldsymbol{e}_{\mathrm{p}}=\sigma(\boldsymbol{W}\boldsymbol{f}'_{\mathrm{m}}+b) \tag{3}$$

式中,$\sigma(\cdot)$ 表示 sigmoid 函数;$\boldsymbol{e}_{\mathrm{a}}\in\mathbf{R}^{1\times1\times c_1}$,$\boldsymbol{e}_{\mathrm{p}}\in\mathbf{R}^{1\times1\times c_2}$;两次映射的权重和偏差值不共享,训练网络时各自独立训练。

4)将激励值 $\boldsymbol{e}_{\mathrm{a}}$ 和 $\boldsymbol{e}_{\mathrm{p}}$ 分别看做外观特征图 $\boldsymbol{f}_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}_{\mathrm{P}}$ 通道上的权重,将其扩展到与对应特征图相同的尺寸后再通过哈达玛积计算加权后的特征图。该过程与 SE 模块相似,可看做是通道维度上的注意力机制,不同之处在于 SEFM 有两种激励输出。最后将加权后的特征图与原始特征图相加,

得到新的外观特征图 $\boldsymbol{f}'_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}'_{\mathrm{P}}$。

5)分别对外观特征图 $\boldsymbol{f}'_{\mathrm{A}}$ 和姿态特征图 $\boldsymbol{f}'_{\mathrm{P}}$ 做池化下采样,将所得特征展开为1维向量后,拼接得到最终的步态特征 $\boldsymbol{f}$。

从整体结构上看,SEFM以特征拼接为基础融合两种特征的信息,并通过压缩—激励过程学习外观特征和姿态特征的权重,以达到自适应融合的目的。SE模块在通道维度上进行加权,这种注意力机制使模型更多关注有效特征并抑制不重要特征。而SEFM特征融合模块可看做在姿态和外观两种特征上加权,首先将两种特征在空间上压缩并拼接得到融合特征,压缩单元接收融合特征并进一步生成一个全局的联合表示,激励单元则根据这个联合表示强调姿态和外观两种特征中更重要的特征。

为了研究外观特征和姿态特征对融合结果的影响,在原始模块的基础上对激励过程进行改动,得到另外两种融合模块,分别如图4(b)和图4(c)所示,称为SEFM-A和SEFM-P。与原始模块不同的是,SEFM-A模块在激励阶段只计算外观特征图的激励值,即将姿态特征融入外观特征,姿态特征保持不变;而SEFM-P模块在激励阶段只计算姿态特征图的激励值,即将外观特征融入姿态特征,外观特征保持不变。

在训练过程中,双分支网络使用了两种损失函数,分别为三元组损失 L_{triple} 和交叉熵损失 L_{CE}。其中三元组损失作用于融合后的步态特征,而交叉熵损失作用于融合前的姿态特征,总的损失为

$$L = L_{\mathrm{triple}} + \lambda L_{\mathrm{CE}} \tag{4}$$

式中,λ 为权重参数,算法实施中取 $\lambda = 2$。

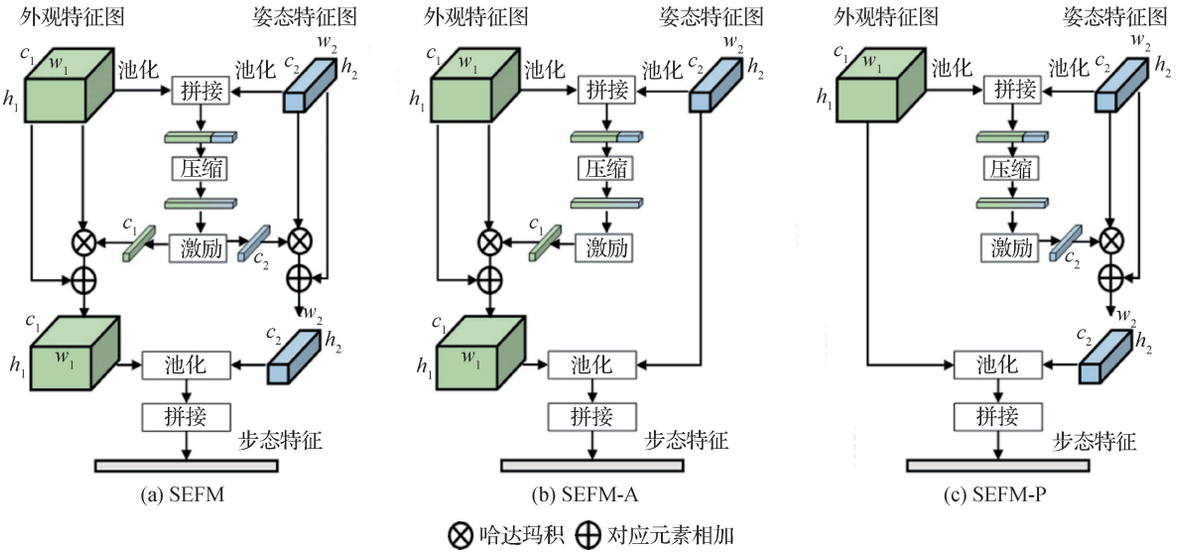


图4 3种不同的特征融合模块结构

Fig. 4 Architecture of three different feature fusion modules((a)SEFM;(b)SEFM-A;(c)SEFM-P)

2 实 验

为了评估双分支特征融合网络的步态识别算法的有效性,在CASIA-B数据集上进行实验,实验中的模型均使用Pytorch框架实现。

2.1 数据集

CAISA-B数据集(Yu等,2006)包含124人的步行视频,每个人含有3种不同行走状态下的10段视频,即正常(NM#1—6)、背包(BG#1—2)、换衣(CL#1—2),每段视频都包含11种角度(0°～180°)。实验中轮廓图像由数据集提供,姿态数据使用AlphaPose姿态估计算法从原始视频中提取,且每幅轮廓图像都有对应的姿态骨架。

实验采用跨视角和不同行走状态两种设置。跨视角设置参考GaitSet算法中的中训练集(medium-sample training,MT)和大训练集(large-sample training,LT)两种数据集划分方式。MT使用前62个行人的数据训练和后62个行人的数据测试;LT使用前74个行人的数据训练和后50人的数据测试。测试时,将前4个正常行走状态(NM#1—4)的数据作为检索库数据(gallery),剩下的数据(NM#5—6,BG#1—2,CL#1—2)按照状态划分为3个待检索数据(probe)分别测试,以评估模型不同行走状态条件下的识别能力,同时在测试时评估所有的视角,因此该实验设置更侧重模型的跨视角识别能力。不同行走

状态设置参考 GaitNet-pre 算法(Zhang 等,2019b)的实施方式,使用前 34 个行人的数据训练和后 90 个行人的数据测试。其中训练集中只包含从 54°～144°的数据,测试集中前 4 个正常行走状态(NM#1—4)的数据作为检索库数据,背包和换衣状态的数据作为两组待检索数据,分别测试相邻视角下的识别准确率。评价指标为 Rank-1 准确率,其中待检索样本和检索库样本角度相同时的测试结果不计入最终结果。

2.2　实验细节

训练时,每次迭代从训练样本里随机挑选 16 位行人,再从每个行人的数据中随机挑选 8 个序列,因此批尺寸为 128;之后从每个序列中随机选取连续的 30 帧作为输入数据,不足 30 帧的序列重复至 30 帧。所有网络均采用 Adam 优化器,姿态分支网络的初始学习率为 0.000 2,外观分支网络和特征融合模块的初始学习率为 0.000 1;在跨视角设置中,共迭代 60 000 次,并在第 45 000 次时将学习率衰减至各自的 0.1 倍;在不同行走状态设置中,共迭代20 000 次;三元组损失的阈值距离设为 0.2。

2.3　实验结果

表 1 是双分支网络与不同的特征融合方式的测试结果。前 3 行是双分支网络中单个分支网络的准确率,后 4 行是双分支网络结合不同的特征融合模块的结果。

表 1　不同特征融合方法的 Rank-1 准确率

Table 1　Rank-1 accuracies of different feature fusion methods

/%

算法	检索状态		
	NM	BG	CL
外观分支网络	91.43	83.29	66.70
姿态分支网络	59.73	32.93	20.46
双分支网络 + 特征拼接	92.59	84.64	68.12
双分支网络 + MMTM	79.51	63.57	51.30
双分支网络 + SEFM	83.52	66.91	54.55
双分支网络 + SEFM-A	81.90	64.65	52.18
双分支网络 + SEFM-P	**93.36**	**84.75**	**70.90**

注:加粗字体表示各列最优值,实验设置为 MT。

从表 1 可以看出,1)基于外观的方法在准确率上远高于基于模型的方法,因为一方面姿态数据本身包含的信息少于轮廓图像,另一方面受姿态提取算法的限制,获得的姿态关键点并不完全准确。但经过特征拼接后,得到的准确率优于外观分支网络,证明了外观特征和姿态特征存在互补性,二者融合后得到的步态特征更加准确。2)添加 MMTM、SEFM 和 SEFM-A 共 3 种特征融合模块后最终的准确率急剧降低,甚至低于外观分支网络;而添加 SEFM-P 特征融合模块后准确率提升。导致该结果的原因仍在于姿态数据,由于相机角度限制,行人在行走过程中存在自遮挡现象,此外原始 RGB 视频的分辨率较低,以及姿态估计算法的性能限制,这些因素共同导致了姿态数据中存在部分噪声。因此在特征融合的过程中,如果将姿态特征融入外观特征,反而会因引入噪声,导致外观特征也不准确。在几种特征融合方法中,MMTM 和 SEFM 将姿态特征和外观特征相互融合;SEFM-A 将姿态特征融入外观特征中,所以这 3 种融合模块都导致准确率降低。而 SEFM-P 将外观特征融入姿态特征,同时保持外观特征不变,避免了外观特征受噪声污染;特征拼接也同样保持了外观特征不变,因此这两种特征融合方法都提高了准确率。根据准确率对比,可以看出本文算法设计的几种 SEFM 模块都优于 MMTM 模块,而且 SEFM-P 达到了最高的准确率,证明了本文算法中特征融合模块的有效性。

表 2 和表 3 分别列举了跨视角设置下基于特征融合的算法和近年主流的步态识别算法在 CASIA-B 数据集上采用 MT 和 LT 两种划分方式的准确率。其中 PoseGait(Liao 等,2020)和 JointsGait(Li 等,2020)是基于姿态模型的方法;GaitSet(Chao 等,2019)、CNN-LB(CNN-matching local features at the bottom layer)(Wu 等,2017)和 GaitNet-pre(Zhang 等,2019b)是基于外观的方法,使用轮廓图像作为输入;EV-Gait(event-based gait recognition)(Wang 等,2019)也是基于外观的方法,但其输入数据为事件图像(event image)。结合表 1,本文算法的外观分支网络采用与 GaitSet 算法相同的网络结构。在 MT 实验设置下,3 种步行状态的平均准确率分别为 91.4%、83.3% 和 66.7%,而 GaitSet(MT)的平均准确率分别为 92.0%、84.3% 和 62.5%。这是因为本文算法在实验实施中调整了序列数据的采样方式,

表 2　跨视角设置下不同算法在 CASIA-B 数据集上采用 MT 方式时的平均 Rank-1 准确率

Table 2　Rank-1 accuracies of different methods on the CASIA-B dataset with MT under different views

/%

检索数据	方法	视角										均值
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM#5—6	PoseGait	49.7	61.6	67.0	66.7	60.8	59.0	62.5	61.4	67.3	62.0	47.5	60.5
	GaitSet(MT)	86.8	95.2	**98.0**	94.5	**91.5**	89.1	91.1	95.0	**97.4**	93.7	80.2	92.0
	SEFM-P(本文)	**89.6**	**95.4**	97.7	**97.0**	91.2	**89.3**	**92.2**	**97.2**	97.1	**94.8**	**85.6**	**93.4**
BG#1—2	PoseGait	32.6	42.2	45.3	44.6	41.9	41.5	39.7	41.0	42.5	37.3	27.6	39.6
	GaitSet(MT)	79.9	**89.8**	**91.2**	86.7	81.6	76.7	81.0	**88.2**	90.3	**88.5**	73.0	84.3
	SEFM-P(本文)	**80.2**	87.3	90.8	**87.6**	**83.1**	**79.8**	**82.4**	86.5	**90.4**	87.9	**76.2**	**84.8**
CL#1—2	PoseGait	20.6	24.0	33.5	33.5	32.7	30.5	36.0	36.1	33.8	27.6	19.0	29.8
	GaitSet(MT)	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
	SEFM-P(本文)	**62.0**	**74.0**	**78.1**	**76.2**	**70.6**	**65.4**	**70.6**	**74.5**	**77.4**	**70.6**	**60.3**	**70.9**

注:加粗字体表示每组各列最优值。

表 3　跨视角设置下不同算法在 CASIA-B 数据集上采用 LT 方式时的平均 Rank-1 准确率

Table 3　Rank-1 accuracies of different methods on the CASIA-B dataset with LT under different views

/%

检索数据	方法	视角										均值
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM#5—6	JointsGait	68.1	73.6	77.9	76.4	77.5	79.1	78.4	76.0	69.5	71.9	70.1	74.4
	EV-Gait	77.3	89.3	94.0	91.8	92.3	96.2	91.8	91.8	91.4	87.8	85.7	89.9
	GaitNet-pre	91.2	92.0	90.5	95.6	86.9	92.6	93.5	96.0	90.9	88.8	89.0	91.6
	GaitSet(LT)	90.8	**97.9**	**99.4**	96.9	93.6	91.7	**95.0**	97.8	**98.9**	96.8	85.8	95.0
	SEFM-P(本文)	**94.0**	97.7	98.6	**97.4**	**94.3**	**92.4**	94.4	**98.3**	98.4	**98.3**	**88.9**	**95.7**
BG#1—2	JointsGait	54.3	59.1	60.6	59.7	63.0	65.7	62.4	59.0	58.1	58.6	50.1	59.1
	EV-Gait	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
	GaitNet-pre	83.0	87.8	88.3	93.3	82.6	74.8	89.5	91.0	86.1	81.2	85.6	85.7
	GaitSet(LT)	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	**94.4**	79.0	87.2
	SEFM-P(本文)	**85.8**	**91.9**	**92.9**	**89.1**	**85.5**	**82.2**	**84.1**	**90.9**	**92.9**	91.5	**79.0**	**87.8**
CL#1—2	JointsGait	48.1	46.9	49.6	50.5	51.0	52.3	49.0	46.0	48.7	53.6	52.0	49.8
	EV-Gait	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
	GaitNet-pre	42.1	58.2	65.1	70.7	68.0	70.6	65.3	69.4	51.5	50.1	36.6	58.9
	GaitSet(LT)	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	SEFM-P(本文)	**72.6**	**83.4**	**85.4**	**80.9**	**74.1**	**71.3**	**76.7**	**76.1**	**80.3**	**80.1**	**66.5**	**77.0**

注:加粗字体表示每组各列最优值。

在略微降低 NM、BG 状态准确率的条件下,大幅提高了 CL 状态的准确率。与另一种基于姿态的方法 PoseGait 相比,本文算法的姿态分支网络的准确率较低。这是因为考虑到提取姿态的计算复杂度,本

文算法采用了 2D 姿态数据,而 PoseGait 采用的 3D 姿态数据在跨视角时更有优势,因此具有更高的准确率。在结合姿态分支网络、外观分支网络和特征融合模块 SEFM-P 后,本文算法的双分支网络准确率进一步提升,分别为 93.4% 、84.8% 和 70.9% 。相比于 GaitSet(MT),在 3 种行走状态下的准确率分别提升了 1.4% 、0.5% 和 8.4% 。在 LT 实验设置中,本文算法同样优于 CNN-LB、GaitSet(LT)、GaitNet-pre 和 EV-Gait 等方法,达到了最高的准确率。需要说明的是,与外观分支网络 GaitSet 相比,本文算法在 CL 行走状态下提升较大,说明本文算法对衣物变化等外观干扰因素具有更好的鲁棒性;但在 NM 和 BG 行走状态下提升较小,这是因为姿态分支网络的准确率远低于外观分支网络,限制了特征融合的效果。根据表 2 的结果,基于姿态的算法如PoseGait、JointsGait 等,其准确率远低于其他几种基于外观的算法,这也是目前基于姿态的算法面临的主要问题。

表 4 为不同行走状态设置下的实验结果。可以看出,与 JUCNet(joint unique-gait and cross-gait network)(Zhang 等,2019a)和 CNN-LB 相比,本文算法在 CL 行走状态下的平均准确率为 90.0% ,远高于 CNN-LB 算法的 62.5% 。在 BG 行走状态下的平均准确率为 94.9% ,优于 JUCNet 算法的 93.2% 。GaitNet-pre 算法通过特征解耦提取与行走状态无关的步态特征,在 CL 行走状态下达到了 89.8% 的平均准确率,与本文算法的 90.0% 相差不大,但在 BG 行走条件下的准确率低于本文算法。与外观分支网络 GaitSet 算法相比,本文算法在 BG 和 CL 行走状态下的准确率分别提高了 1.6% 和 6.7% ,说明 SEFM-P 特征融合模块能够有效融合外观和姿态两种特征,且融合姿态特征后能够提升算法对背包、衣物等外观变化的鲁棒性。

表 4　不同行走状态设置下不同算法在 CASIA-B 数据集上的平均 Rank-1 准确率

Table 4　Rank-1 accuracies of different methods on the CASIA-B dataset under different walk conditions

/%

检索数据	方法	视角(待检测样本/已知参考样本)						均值
		54°/ 36°	54°/ 72°	90°/ 72°	80°/ 108°	126°/ 108°	126°/ 144°	
CL	GaitSet*	71.5	81.0	92.8	92.2	82.1	80.5	83.3
	GaitNet-pre	**87.0**	90.0	94.2	86.5	**89.8**	**91.2**	89.8
	JUCNet	–	–	–	–	–	–	–
	CNN-LB	49.7	62.0	78.3	75.6	58.1	51.4	62.5
	SEFM-P(本文)	82.7	**91.1**	**95.0**	**93.3**	89.4	88.8	**90.0**
BG	GaitSet*	91.0	91.0	**96.7**	93.3	93.3	94.4	93.3
	GaitNet-pre	91.6	90.0	95.6	87.4	90.1	93.8	91.4
	JUCNet	91.8	93.9	95.9	95.9	93.9	87.8	93.2
	CNN-LB	**92.7**	90.4	93.3	88.9	93.3	86.0	90.8
	SEFM-P(本文)	91.6	**95.5**	96.1	**95.0**	**96.1**	**95.0**	**94.9**

注:加粗字体表示每组各列最优值;GaitSet* 为复现的结果;“–”表示论文未提供实验结果。

3　结　论

步态识别算法包括基于外观的方法和基于模型的方法两类,结合两类方法的优点,设计了一种双分支特征融合网络的步态识别算法。该算法通过两条分支网络分别提取外观特征和姿态特征,再利用特征融合模块融合两种特征以得到更准确的步态表示,适用于能够同时获得轮廓图像和人体姿态的场景。CASIA-B 数据集上的实验结果表明,该算法与目前主流的算法相比具有优势,达到了更高的准确率。与其他步态识别算法相比,该算法能够利用更丰富的步态信息,对于衣物外观变化这类干扰因素具有更好的鲁棒性;与其他特征融合方法相比,该算

法在融合两种特征的同时,能够避免外观特征受姿态特征中的噪声影响,得到更准确的特征。尽管该算法在准确率上有所提升,但对姿态数据的处理上还存在不足。首先,该算法没有对姿态数据中的噪声进行处理,导致姿态特征的准确性较低;其次,该算法使用两条分支网络分别提取特征,效率有所降低。在未来的工作中,将围绕以上问题进一步改进。一方面提高网络对姿态数据中噪声的鲁棒性,以得到更准确的步态表示;另一方面优化网络结构,提高算法效率使其满足实际应用的需求。

参考文献(References)

Ben X, Gong C, Zhang P, Jia X T, Wu Q and Meng W X. 2019a. Coupled patch alignment for matching cross-view gaits. IEEE Transactions on Image Processing, 28(6): 3142-3157 [DOI: 10.1109/TIP.2019.2894362]

Ben X, Gong C, Zhang P, Yan R, Wu Q and Meng W X. 2020. Coupled bilinear discriminant projection for cross-view gait recognition. IEEE Transactions on Circuits and Systems for Video Technology, 30(3): 734-747 [DOI: 10.1109/TCSVT.2019.2893736]

Ben X, Zhang P, Lai Z H, Yan R, Zhai X L and Meng W X. 2019b. A general tensor representation framework for cross-view gait recognition. Pattern Recognition, 90: 87-98 [DOI: 10.1016/j.patcog.2019.01.017]

Ben X Y, Xu S and Wang K J. 2012. Review on pedestrian gait feature expression and recognition. Pattern Recognition and Artificial Intelligence, 25(1): 71-81 (贲晛烨, 徐森, 王科俊. 2012. 行人步态的特征表达及识别综述. 模式识别与人工智能, 25(1): 71-81) [DOI: 10.16451/j.cnki.issn1003-6059.2012.01.010]

Chao H Q, He Y W, Zhang J P and Feng J F. 2019. GaitSet: regarding gait as a set for cross-view gait recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 8126-8133 [DOI: 10.1609/aaai.v33i01.33018126]

Chao H Q, Wang K, He Y W, Zhang J P and Feng J F. 2021. GaitSet: cross-view gait recognition through utilizing gait as a deep set. IEEE Transactions on Pattern Analysis and Machine Intelligence: #3057879 [DOI: 10.1109/TPAMI.2021.3057879]

Fang H S, Xie S Q, Tai Y W and Lu C W. 2017. RMPE: regional multi-person pose estimation//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2353-2362 [DOI: 10.1109/ICCV.2017.256]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Kastaniotis D, Theodorakopoulos I, Theoharatos C, Economou G and Fotopoulos S. 2015. A framework for gait-based recognition using Kinect. Pattern Recognition Letters, 68: 327-335 [DOI: 10.1016/j.patrec.2015.06.020]

Li C, Zhong Q Y, Xie D and Pu S L. 2018. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization: 786-792 [DOI: 10.24963/ijcai.2018/109]

Li N, Zhao X B and Ma C. 2020. JointsGait: a model-based gait recognition method based on gait graph convolutional networks and joints relationship pyramid mapping [EB/OL]. [2020-12-09]. https://arxiv.org/pdf/2005.08625.pdf

Liao R J, Cao C S, Garcia E B, Yu S Q and Huang Y Z. 2017. Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations//Proceedings of the 12th Chinese Conference on Biometric Recognition. Shenzhen, China: Springer: 474-483 [DOI: 10.1007/978-3-319-69923-3_51]

Liao R J, Yu S Q, An W Z and Huang Y Z. 2020. A model-based gait recognition method with body pose and human prior knowledge. Pattern Recognition, 98: #107069 [DOI: 10.1016/j.patcog.2019.107069]

Sadeghzadehyazdi N, Batabyal T, Glandon A, Dhar N K, Familoni B O, Iftekharuddin K M and Acton S T. 2019. Glidar3DJ: a view-invariant gait identification via flash lidar data correction//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE: 2606-2610 [DOI: 10.1109/ICIP.2019.8803237]

Vaezi Joze H R, Shaban A, Iuzzolino M L and Koishida K. 2020. MMTM: multimodal transfer module for CNN fusion//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 13286-13296 [DOI: 10.1109/CVPR42600.2020.01330]

Wang Y X, Du B W, Shen Y R, Wu K, Zhao G R, Sun J G and Wen H K. 2019. EV-Gait: event-based robust gait recognition using dynamic vision sensors//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 6351-6360 [DOI: 10.1109/CVPR.2019.00652]

Wolf T, Babaee M and Rigoll G. 2016. Multi-view gait recognition using 3D convolutional neural networks//Proceedings of 2016 IEEE International Conference on Image Processing (ICIP). Phoenix, USA: IEEE: 4165-4169 [DOI: 10.1109/ICIP.2016.7533144]

Wu Z F, Huang Y Z, Wang L, Wang X G and Tan T N. 2017. A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(2): 209-226 [DOI: 10.1109/TPAMI.2016.2545669]

Yu S Q, Liao R J, An W Z, Chen H F, García E B, Huang Y Z and

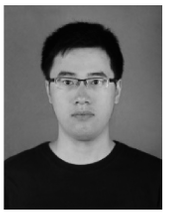
Poh N. 2019. GaitGANv2: invariant gait feature extraction using generative adversarial networks. Pattern Recognition, 87: 179-189 [DOI: 10.1016/j.patcog.2018.10.019]

Yu S Q, Tan D L and Tan T N. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition//Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong, China: IEEE: 441-444 [DOI: 10.1109/ICPR.2006.67]

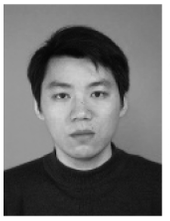
Zhang K H, Luo W H, Ma L, Liu W and Li H D. 2019a. Learning joint gait representation via quintuplet loss minimization//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4695-4704 [DOI: 10.1109/CVPR.2019.00483]

Zhang Z Y, Tran L, Yin X, Atoum Y, Liu X M, Wan J and Wang N X. 2019b. Gait recognition via disentangled representation learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4705-4714 [DOI: 10.1109/CVPR.2019.00484]

作者简介



徐硕，1997 年生，男，硕士研究生，主要研究方向为步态识别与深度学习。
E-mail：xareus@hotmail.com



唐俊，通信作者，男，教授，主要研究方向为计算机视觉与模式识别。
E-mail：tangjunahu@163.com

郑锋，男，助理教授，主要研究方向为计算机视觉、机器学习与人机交互。E-mail：zhengf@sustc.edu.cn
鲍文霞，女，副教授，主要研究方向为机器学习、图像视频处理和模式识别。E-mail：bwxia@ahu.edu.cn