

DOI:10.19651/j.cnki.emt.2314185

基于多尺度特征融合的跨视角步态识别*

邹雪^{1,2} 谭棉^{1,2} 严晓波^{1,2} 王飞^{2,3} 王林²

(1. 贵州民族大学数据科学与信息工程学院 贵阳 550025; 2. 贵州省模式识别与智能系统重点实验室, 贵州民族大学 贵阳 550025; 3. 贵州民族大学人文科技学院 贵阳 550025)

摘要: 在跨视角步态识别中, 针对衣着遮挡情况下难以提取具有可辨别性和多样性的步态特征, 导致识别准确率下降的问题, 提出了一种基于多尺度特征融合网络的跨视角步态识别方法。该方法能够有效利用步态特征间的互补性, 获得具有可辨别性和多样性的步态特征, 从而解决因衣着遮挡造成可辨别性差以及单一性的问题, 进而提升跨视角步态识别的准确性。为验证所提方法的有效性, 在公共数据集 CASIA-B 上进行了验证, 实验结果表明所提方法在处理具有遮挡条件下的跨视角步态识别问题的识别性能达到了 73.4%, 同时在正常和背包两种行走条件下的识别性能分别达到了 95.5% 和 88.0%。此外, 我们的方法在处理遮挡条件下的识别性能优于同类典型的步态识别方法。

关键词: 跨视角步态识别; 多尺度特征融合; 步态特征; 可辨别性

中图分类号: TP75 **文献标识码:** A **国家标准学科分类代码:** 520.2

Cross-view gait recognition based on multi-scale feature fusion

Zou Xue^{1,2} Tan Mian^{1,2} Yan Xiaobo^{1,2} Wang Fei^{2,3} Wang Lin²(1. School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;
2. Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang 550025, China;
3. College of Humanities & Sciences, Guizhou Minzu University, Guiyang 550025, China)

Abstract: In cross-view gait recognition, it is difficult to extract distinguishable and diverse gait features in the case of clothing occlusion, which leads to the decrease of recognition accuracy. A multi-scale feature fusion network based cross-view gait recognition method is proposed. This method can effectively utilize the complementarity among gait features to obtain gait features with discriminability and diversity, thereby solving the problem of poor discriminability and uniformity caused by clothing occlusion, and thus improving the accuracy of cross-viewing Angle gait recognition. In order to verify the effectiveness of the proposed method, the public data set CASIA-B was used to verify the proposed method. The experimental results show that the proposed method achieves 73.4% recognition performance for the cross-viewing Angle gait recognition problem with occlusion, and 95.5% and 88.0% recognition performance under normal and backpack walking conditions, respectively. In addition, the performance of our method is better than that of other typical gait recognition methods under occluded conditions.

Keywords: cross-view gait recognition; multi-scale feature fusion; gait feature; discriminability

0 引言

步态是一个人行走的姿态, 具有周期性, 每个人都有自己独特的行走步态, 通过分析个人的行为动作来研究个体的心里状态。在包含一个步态周期的轮廓序列中, 可以观

察到每个位置的轮廓都具有独特的姿态。不同于人脸、指纹、虹膜等其它生物特征, 步态是唯一可在远距离不受控制的状态下获得的生物特征^[1-2]。跨视角步态识别是一项涉及计算机视觉和模式识别的研究领域, 旨在通过分析人体行走的特征来识别个体。这一技术对于人体动作分析、安

收稿日期: 2023-07-23

* 基金项目: 国家自然科学基金(62241206)、国家自然科学基金(62162012)、贵州省科技计划项目(黔科合基础-ZK[2022]一般 195、黔科合基础-ZK[2023]一般 143、黔科合基础-ZK[2022]一般 550、黔科合平台人才-ZCKJ[2021]007)、贵州省高层次创新型人才项目(黔科合平台人才-GCC[2023]027)、贵州省教育厅自然科学研究项目(黔教技[2022]015 号)、贵州省教育厅青年科技人才成长项目(黔教技[2023]012 号, 黔教技[2022]015 号, 黔教技[2023]061 号, 黔教技[2023]062 号, 黔教合 KY 字[2021]115)、贵州省模式识别与智能系统重点实验室开放课题(GZMUKL[2022]KF01)资助

防监控、医疗辅助和智能交通等领域具有重要意义^[3]。

目前,现有的跨视角步态识别方法大致可分为两类:基于三维模型的方法,利用计算机图形学和深度传感器等技术,将人体行走的三维动作转化为数字化的三维模型数据进行分析 and 识别。这种方法可以提取更丰富的空间和时间信息,从而在不同视角下实现更准确的步态识别。基于二维视频序列的方法,主要依赖于计算机视觉和图像处理技术,在二维视频序列中提取并分析人体行走的特征,利用机器学习算法或深度学习模型对这些特征进行分类和识别。GaitSet^[4]使用二维 CNN 在帧级水平上提取全局特征,并将步态序列视为一个可以呈现时间信息的集合。然后利用设置池法将帧级特征聚合为独立的序列级特征。但由于信息量的限制,可能存在一定的识别准确性和视角鲁棒性方面的挑战。GaitNet^[5]提出了一种自动编码器框架,该框架从原始 RGB 图像中提取步态特征,然后利用长-短期记忆网络建立步态序列的时间变化模型。这仅利用全局特征来描述步态信息,而对步态序列的局部细节关注不够。为了获得不同个体更细微和可区分的步态特征,Zhang 等^[6]提出了局部特征提取器和时间注意模型来提取可区分的时空特征,同时制定了有效的损失函数来区分步态特征。GaitPart^[7]引入了一种新型的卷积层——聚焦卷积层,它将特征图水平分割得到多个部分步态序列,以探索更细粒度的局部信息。然而,局部信息表示丢失了步态帧的全局上下文信息,而忽略了局部区域之间的关系。李言等^[8]提出了一种针对步态识别中容易受环境干扰问题的特征提取方法,使模型在更为复杂的环境中具有良好的泛化能力。Wu 等^[9]研究了一种通过深度卷积神经网络的相似性学习进行步态识别的方法,它可以训练深度网络来识别步态模式的最具鉴别力的变化。张红颖等^[10]在步态特征提取阶段进行了水平多级化分块操作,逐级融合来减弱遮挡造成的局部特征的影响。

综上所述,虽然特征融合思想在跨视角步态识别方面取得了一定的研究,但相关工作大都是基于全局信息和局部特征信息融合。因此,在步态信息可辨别性和多样性上关于因遮挡情况造成的识别准确率低问题需要进一步研究。针对此问题,本文设计了一种多尺度特征融合网络(multi-scale feature fusion network, MSFFN),使用改进后的基线网络 GaitSet 提取不同尺度步态特征,然后利用特征融合模块进行融合,有效利用不同尺度步态特征间的互补性,获得具有可辨别性和多样性的步态特征,从而提升跨视角步态识别的准确性。本文的主要贡献在于:

- 1) 设计了一种步态特征提取网络,利用外观数据分别提取不同尺度的外观特征,并进一步融合得到更具有辨别性和多样性的步态特征,从而达到更高的准确率。
- 2) 设计了一种多尺度特征融合模块,能够有效利用不同尺度步态特征间的互补性,且适用于特征维数不匹配的情形。
- 3) 在 CASIA-B 数据集上与主流方法进行对比,验证

了方法的有效性。

1 跨视角步态识别

1.1 问题描述

受试者的完整周期轮廓如图 1 所示。

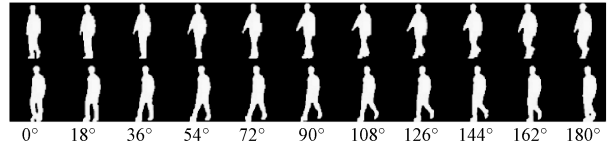


图 1 受试者的完整周期轮廓

针对遮挡情况下的跨视角步态识别中步态特征信息可辨别性差以及单一性的问题,导致识别准确率低的情形,本文设计了一种基于多尺度特征融合网络的跨视角步态识别方法。通过对基线网络 GaitSet 改进解决步态特征信息可辨别性差以及单一性的问题,提升步态信息的可辨别性和多样性;最后加入多尺度特征融合模块,利用多尺度步态信息间的互补性,提升网络在跨视角步态识别中的识别准确性,总体结构如图 2 所示。该方法主要是有两部分组成:步态特征提取和特征融合。

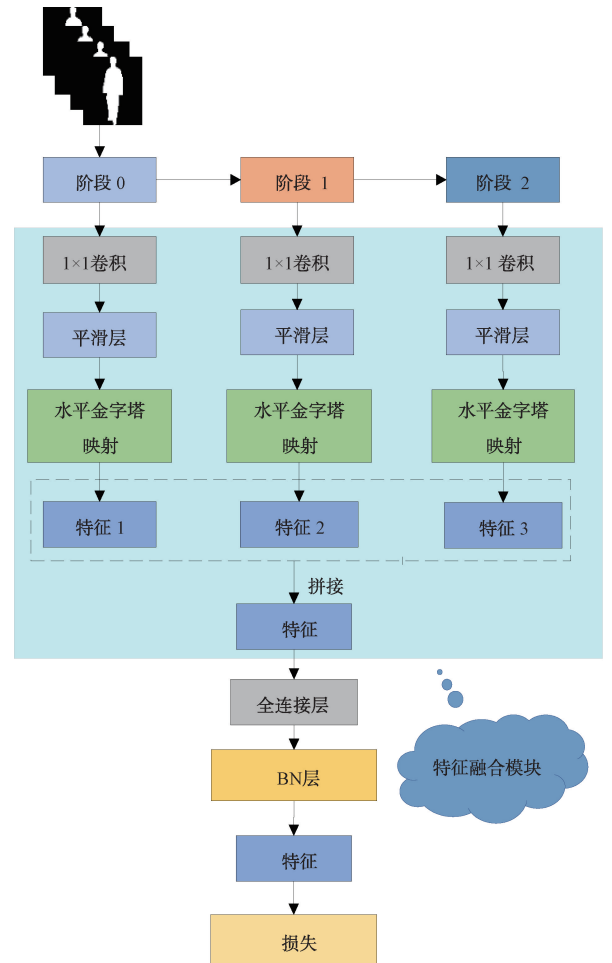


图 2 步态网络基本框架

本文的主要思想是:通过多尺度特征融合模块提取网络中不同层的步态特征,使步态特征的可辨别性增强,以及丰富其多样性。该网络的具体操作:

第 1 步是将基线网络 Gaitset^[4]分为三个阶段,然后每个阶段作为一个输出,阶段 0 (Stage 0)阶段输出的特征比较粗糙,但也是保留步态信息最多的阶段;阶段 1 (Stage 1)和阶段 2 (Stage 2)两阶段分别是经过卷积层后得到的步态信息,相对于 Stage 0 阶段来说,Stage 1 阶段得到的步态信息较少,Stage 2 阶段得到的步态信息更少,但 Stage 2 阶段得到的步态信息更精细;

第 2 步是分别将 3 个阶段的特征输入到多尺度特征融合模块(图 2 中浅蓝色部分)中,通过多尺度特征融合模块将每个阶段的步态特征(即:精细程度不同的步态特征)进行融合,得到最终的步态特征进行识别。

1.2 步态特征提取网络

针对步态特征可辨别性差和单一性的问题,本文采用 GaitSet^[4]作为外观步态特征提取基线网络,将网络分为 3 个阶段,在每个阶段的最后添加 Cat 操作,保证每阶段的输出,结构如图 3 所示。该网络的主干由 6 层卷积层和 2 层池化层组成,输入轮廓图像序列后得到特征图序列。除了提取轮廓图像的外观特征之外,该网络的分支还用于融合网络的浅层特征与深层特征。在图 3 中,池化操作表示对序列在时序上进行池化,而 \oplus 表示对应元素的相加操作。因此,外观步态特征提取网络还起到融合时序信息的作用。在每个阶段添加了拼接操作将两种特征图在通道维度上进行拼接,目的是为了便于后期的特征融合。

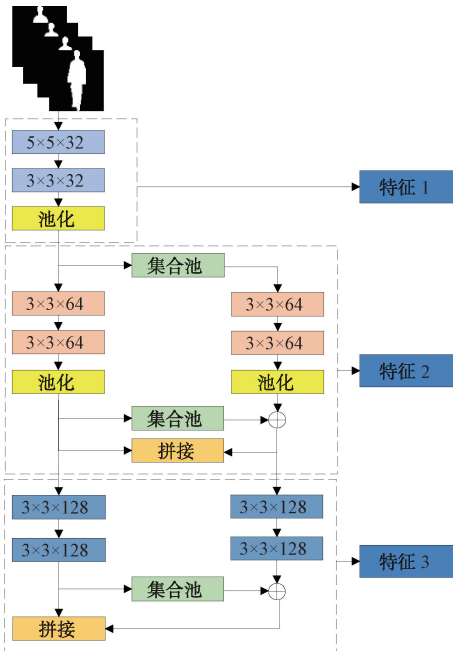


图 3 外观特征提取网络

1.3 多尺度步态特征融合模块

考虑到跨视角步态识别准确性问题,本文提出了多尺

度特征融合模块。该模块是以外观步态特征提取网络为基础,将网络分解成 3 个阶段:Stage 0、Stage 1、Stage 2,每个阶段的输出均是主干网络和分支网络进行拼接后的特征,将得到的特征作为每个阶段的 1×1 Conv 输入, 1×1 Conv 的作用是对通道数进行调整,使得每层的特征通道数都一样,固定为 256;再经过 3×3 卷积的平滑层(smoothing layer, SMO),缓解不同阶段之间的语义差距,最后经过水平金字塔映射^[4](horizontal pyramid mapping, HPM)得到每个阶段的特征。最后将 3 个阶段所得到的特征进行融合,得到最终的步态特征^[11],如图 4 所示。

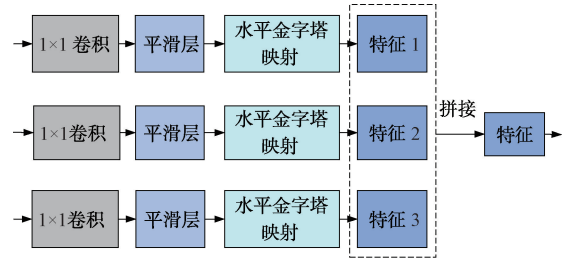


图 4 特征融合模块

假设主干网络输出的特征为 X , 分支网络输出的特征为 Y , 每个阶段的具体计算如下:

Stage 0:

$$F_1 = Feature_1(HPM(SMO(1 \times 1Conv(X_1)))) \quad (1)$$

Stage 1:

$$F_2 = Feature_2(HPM(SMO(1 \times 1Conv(Cat(X_2, Y_1)))))) \quad (2)$$

Stage 2:

$$F_3 = Feature_3(HPM(SMO(1 \times 1Conv(Cat(X_3, Y_2)))))) \quad (3)$$

其中, $1 \times 1Conv(\cdot)$ 是对通道维数进行调整, $SMO(\cdot)$ 是平滑层, $HPM(\cdot)$ 是水平金字塔映射操作, $F_i (i = 1, 2, 3)$ 为输出特征。

对每个阶段的特征进行融合得到:

$$F_{total} = Cat(F_1, F_2, F_3) \quad (4)$$

其中, F_{total} 为融合后的特征, $Cat(\cdot)$ 是通道维数上的串联运算。

本节重点描述了网络中的多尺度特征融合模块,将在第 2 节的实验部分对其进行验证。

2 实 验

为了验证多尺度特征融合模块在跨视角步态识别中解决遮挡情况下的识别性能问题,本节在公开数据集 CASIA-B 上进行了验证。

2.1 实验设置与分析

1)数据集

CASIA-B 数据集^[12]是一个流行的步态数据集,包含 124 名受试者,标记为 001 到 124。每个受试者有 3 种行走

条件,即正常(normal,NM)(每个受试者6个视频序列),带包(bag,BG)(每个受试者2个视频序列),穿外套或夹克(clothes,CL)(每个受试者2个视频序列)。每个序列同时被设置为11个视图($0^\circ, 18^\circ, \dots, 180^\circ$)。因此,该数据集总共包含 $124 \times (6+2+2) \times 11 = 113\ 640$ 个视频序列。根据目前的数据集情况,在进行实验时采用了3种常见的划分方式。根据训练集的规模不同,将这3种方法分别命名为划分小样本训练集(small sample training set,ST)、中样本训练集(medium sample training set,MT)和大样本训练集(large sample training set,LT)。在ST设置中,将前24个受试者(001-024)用于训练,剩下的100个受试者用于测试,两者之间没有重叠。在MT设置中,将前62个受试者(001-062)用于训练,剩下的62个受试者用于测试。在LT设置中,将前74个受试者(001-074)用于训练,剩下的50个受试者用于测试。在训练集中,将NM条件的前4个序列(即NM#1-4)作为验证集,其余的6个序列作为训练

集,分为3个子集,分别是包含BG#1-2的BG子集,包含CL#1-2的CL子集,以及NM子集条件#5-6。

2) 参数设置

实验时输入的均是一组大小为 64×44 的对齐轮廓,这些轮廓是由数据集直接提供,并根据文献[13]中描述的方法进行对齐。采用Adam优化器^[14]来训练本文的GaitSet-MSFFM网络。所有实验的代码都是用Python和Pytorch 1.13编写的。这些模型是在一台配备了2台NVIDIA A100gpu的计算机上进行的训练。除非另有说明,在训练阶段的设置基数被设置为30。三元组损失函数的边际被设置为0.2。HPM量表的数量S被设置为7。对于CASIA-B数据集,将C1和C2的通道数设置为32,C3和C4的通道数设置为64,C5和C6的通道数设置为128。在这些设置下,本文模型的平均计算复杂度为8.6gflop。如表1所示,列出了不同实验设置下的批量大小、学习速率和训练迭代次数。此外,在后续的评价中,采用rank-1准确性作为标准。

表1 CASIA-B不同实验设置下的批量大小、学习速率和训练迭代次数

		小样本训练集	中样本训练集	大样本训练集
交叉熵损失	批量大小	128	128	128
	学习速率	1×10^{-4}	1×10^{-4}	1×10^{-4}
	迭代次数	70k	80k	90k
三元组损失	批量大小	$P=8, k=16$	$P=8, k=16$	$P=8, k=16$
	学习速率	1×10^{-4}	1×10^{-4}	1×10^{-4}
	迭代次数	80k	80k	100k
微调三元组损失	批量大小	$P=8, k=16$	$P=8, k=16$	$P=8, k=16$
	学习速率	1×10^{-5}	1×10^{-5}	1×10^{-5}
	迭代次数	30k	40k	60k

3) 数据预处理

由于原始视频中的行人行走过程与相机距离不断变化,导致一个序列中轮廓图像的分辨率也在改变。为了避免这种数据的影响,需要对输入数据进行标准化处理。

假设原始RGB视频序列为 $\mathbf{I} = \{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_T\}$,其中 \mathbf{I} 表示单幅RGB图像,T表示序列中的图像数量。通过背景分割算法,得到了轮廓图像序列 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$,其中 \mathbf{x} 表示单幅轮廓图像。

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。然后,将轮廓图像保持宽高比例不变,但将其高度缩放到64个像素,并将图像的宽度左右平均填充到44个像素。最后,对所有轮廓图像进行标准化处理,以消除不同图像之间的亮度和对比度差异。

2.2 实验结果与分析

为了验证多尺度特征融合模块在跨视角步态识别中的有效性,本节实验主要和GaitSet^[4]、CNN-LB^[9]、CNN-3D^[9]、CNN-Ensemble^[9]、AE^[15]、MGAN^[16]、GaitNet^[17]实验作对比,结果如表2所示,比较小样本训练集(ST)、中样

本训练集(MT)和大样本训练集(LT)3种不同实验设置下的平均rank-1。Wu等在文献[9]中提出了几种基于CNN的模型,CNN-LB将两个步态序列的GEIs作为两个通道输入一个3层CNN,判断这两个GEIs是否属于同一个人;CNN-3D是在9个相邻的帧上运行3层3D-CNN,并对16个9帧的样本进行平均预测,得到最终的输出;CNN-Ensemble聚合了8个不同网络的输出。Yu等^[15]应用自动编码器(AE)来提取视角不变的特征。He等^[16]等人提出了一种多任务GAN(MGAN),将步态特征从一个角度投影到另一个角度,用于跨视角步态识别。除本文方法GaitSet-MSFFM外,其他结果均直接取自他们的最原始论文。所有的结果都在11个角度的不同视图上取平均值。

小样本训练集(ST)。本文的方法即使在训练集为24名受试者的情况下也取得了很高的性能,比文献[9]高出15%,文献[4]高出3%。原因主要是本文方法是基于文献[4]将网络分为了3个阶段,然后将每个阶段的输出单独作为 1×1 Conv的输入,再经过SMO,最后经过HPM获得不同阶段的特征,使步态特征信息具有可辨别性和多样性。

表 2 3 种不同实验设置的 CASIA-B 平均 rank-1

验证样本集 NM#1-4		0°~180°											mean	
查询样本集		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°		
小样本 训练集(24)	CNN-LB [9]	54.8	—	—	77.8	—	64.9	—	76.1	—	—	—	—	
	NM#5-6	GaitSet [4]	70.2	82.9	89.8	88.8	82.2	78.2	81.5	87.7	89.1	84.2	66.9	81.9
		Ours	69.5	87.7	93.8	91.3	86.9	81.8	86.2	91.7	93.6	86.9	67.9	85.2
	BG#1-2	GaitSet [4]	60.2	72.4	77.4	75.3	71.5	67.8	70.2	77.2	78.4	71.8	58.9	71.0
		Ours	64.4	76.2	82.7	83.1	77.1	72.5	77.0	81.7	83.6	75.2	60.0	75.8
	CL#1-2	GaitSet[4]	31.8	44.8	50.2	45.2	42.5	39.8	42.2	44.6	41.3	35.2	28.4	40.6
中样本 训练集(62)		Ours	37.0	50.3	56.0	49.6	45.8	44.8	46.9	47.7	45.5	38.6	30.0	44.8
		AE [15]	49.3	61.5	64.4	63.6	63.7	58.1	59.9	66.5	64.8	56.9	44.0	59.3
	NM#5-6	MGAN [16]	54.9	65.9	72.1	74.8	71.1	65.7	70.0	75.6	76.2	68.6	53.8	68.1
		GaitSet [4]	88.2	95.5	97.3	96.1	91.9	89.1	91.8	95.7	97.3	93.7	82.3	92.6
		Ours	87.9	95.1	97.8	96.4	91.4	89.7	91.5	96.0	97.4	93.9	84.8	92.9
		AE [15]	29.8	37.7	39.2	40.5	43.8	37.5	43.0	42.7	36.3	30.6	28.5	37.2
BG#1-2	MGAN [16]	48.5	58.5	59.7	58.0	53.7	49.8	54.0	61.3	59.5	55.9	43.1	54.7	
	GaitSet [4]	79.9	89.8	91.2	86.7	81.6	76.7	81.0	88.2	90.3	88.5	73.0	84.3	
	Ours	79.9	87.2	90.4	88.1	84.6	82.5	84.7	89.1	91.3	88.3	77.6	85.8	
	AE [15]	18.7	21.0	25.0	25.1	25.0	26.3	28.7	30.0	23.6	23.4	19.0	24.2	
CL#1-2	MGAN [16]	23.1	34.5	36.3	33.3	32.9	32.7	34.2	37.6	33.7	26.7	21.0	31.5	
	GaitSet [4]	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5	
	Ours	59.4	72.0	74.7	71.7	63.4	59.3	62.0	65.8	66.4	61.8	50.6	64.4	
大样本 训练集(74)	CNN-3D [9]	87.1	93.2	97.0	94.6	90.2	88.3	91.1	93.8	96.5	96.0	85.7	92.1	
	CNN-Ensemble [9]	88.7	95.1	98.2	96.4	94.1	91.5	93.9	97.5	98.4	95.8	85.6	94.1	
	NM#5-6	GaitNet [17]	93.1	92.6	90.8	92.4	87.6	95.1	94.2	95.8	92.6	90.4	90.2	92.3
		GaitSet [4]	91.7	97.1	98.5	97.0	93.6	91.9	94.7	96.8	97.7	97.0	88.7	94.9
		Ours	92.9	98.0	99.3	97.4	93.9	92.1	94.5	97.8	98.7	96.6	88.8	95.5
		CNN-LB [9]	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
BG#1-2	GaitNet [17]	88.8	88.7	88.7	94.3	85.4	92.7	91.1	92.6	84.9	84.4	86.7	88.9	
	GaitSet [4]	85.2	92.0	93.4	90.9	86.7	81.7	86.7	91.0	93.9	93.4	82.2	88.8	
	Ours	83.6	90.7	93.2	91.2	85.6	82.6	84.1	91.6	92.6	91.2	81.2	88.0	
	CNN-LB [9]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0	
CL#1-2	GaitNet [17]	50.1	60.7	72.4	72.7	74.6	78.4	70.3	68.2	53.5	44.1	40.8	62.3	
	GaitSet [4]	59.5	75.0	78.3	74.6	71.4	71.3	70.8	74.1	74.6	69.4	54.1	70.3	
	Ours	63.6	78.0	80.4	80.8	73.9	71.6	72.8	75.7	78.0	70.3	61.7	73.4	

中样本训练集(MT)和大样本训练集(LT)。一般来说,深度学习方法的性能在很大程度上取决于训练集的规模。因此,按照先前文献的建议,对训练集和测试集使用两个不同的划分,即 MT 和 LT,来评估本文的 GaitSet-MSFFN。表 2 显示,在 LT 上,GaitNet 模型只在 0°、90°、180°这几个特殊的角度上的识别精度高,而本文的模型在 0°、90°、180°相对于 GaitNet 模型,精度差距小,同时还满足在其他视角下的高精度结果,在 NM 数据子集上,其中除 180°外,所有视图的结果都超过 90%。这一结果表明,当该数据子集的训练数据越多时,精度就会有明显的提高。同时,本文模型在 BG 数据子集上都取得了令人满意的性能。在 CL 数据子

集上,GaitNet 模型除了在 72°、90°上的性能较好外,其他视角下的识别性能均本文模型的识别性能差,尽管本文的模型仍然超过到目前为止文献[9]报告的最佳性能将近 20%、文献[4]的最佳性能 3%,但相对于另外两种条件下的性能还是有很大的差距。出现这种结果的主要原因有:1)外套可以完全改变一个人的外表,例如,一个被试者穿外套比穿 T 恤看起来更大。2)外套可以隐藏四肢和身体的运动。3)在训练集中,CL 数据子集的比例明显低于 NM 数据子集的比例,要求模型具有更强的辨别能力。

2.3 消融实验

为了平衡代表性和计算成本,本文在设置池中考虑了

3个统计函数:最大值函数 $\max(\cdot)$ 、均值函数 $\text{mean}(\cdot)$ 和中值函数 $\text{median}(\cdot)$,在本节将作出全面的比较,如表3所示。

表3 消融实验

序号	能量图	设置集	设置集池化			MGP	MSFFN	NM	BG	CL
			最大值	均值	中值					
1	✓						89.0	76.3	50.7	
2		✓	✓				95.4	88.7	69.9	
3		✓		✓			95.0	86.3	66.3	
4		✓			✓		94.8	84.9	63.7	
5		✓	✓			✓	94.9	88.8	70.3	
6		✓	✓			✓	✓	95.5	88.0	73.4

设置集(Set)与能量图(GEI)相比。表3的前两行显示了将步态作为一个集合的有效性。在完全相同的网络中,使用Set的结果比使用GEI的结果好,在NM子集上超过6%,在BG子集上超过12.4%,在CL子集上超过19%。

MGP的影响。表3的第2行和第5行显示了MGP改进了所有3个测试子集,即从主通道的不同层中提取的特征包含不同的判别信息。

MSFFN的影响。在表3中,比较第5行和第6行可以得到,加入多尺度特征融合模块之后的NM和CL两个数据集上的结果均优于不加多尺度特征融合模块的结果,尽管在BG数据集上的结果略低,但综合所有子集的情况来看,影响不大,可以忽略。

HPM量表和HPM权重独立性的影响。如表4所示,HPM的规模越大,性能就越好。此外,表4的最后两行比较了HPM中全连接层的权重独立性的影响。可以看出,使用独立的权值使得每个数据子集的精度均有提高。在CL数据子集上效果最明显。在实验过程中,也注意到引入独立的权值会使网络收敛得更快。

表4 不同HPM尺度和权重独立性实验的影响

HPM 尺度	HPM 权重		NM	BG	CL
	共享	独立			
1(no HPM)		✓	91.8	82.4	59.4
2		✓	91.8	82.9	60.1
3		✓	91.7	83.0	64.2
4		✓	93.9	86.9	64.5
5		✓	94.9	88.8	70.3
6	✓		91.1	82.0	60.9
		✓	95.5	88.0	73.4

3 结 论

本文针对遮挡情况下的跨视角步态识别问题,提出了一种基于多尺度特征融合网络(MSFFN)的跨视角步态识

别方法,提升网络因可辨别性和多样性在遮挡情况下的识别准确率。实验表明,与现有的步态识别方法相比,多尺度特征融合网络(MSFFN)能更有效地利用不同尺度步态特征间的互补性,提高步态特征的可辨别性和多样性。同时,在公共数据集CASIA-B上的广泛实验结果验证了我们的方法在处理具有遮挡条件下的跨视角步态识别性能上达到了73.4%,相较于基线网络GaitSet提升了3.1%。同时也保持了正常行走和背包两种条件下的识别性能基本不变。因此,我们的方法在处理遮挡情况下的识别性能均优于同类典型的步态识别方法。

虽然MSFFN能够解决一定程度上的遮挡情况下的跨视角步态识别问题,但在遮挡情况下的识别准确率仍有待提高。因此,后续的工作将集中在:探索基于注意力机制的跨视角步态识别问题,以及混合注意力机制的GaitSet-MSFFN识别性能问题。

参考文献

- [1] ZHAI X L, BEN X Y, LIU C, et al. Decomposing identity and view for cross-view gait recognition[C]. 2022 IEEE International Conference on Multimedia and Expo(ICME), Taipei, Taiwan, 2022, 1-6.
- [2] 曾维,何刚强,罗伟洋,等. 基于ICNet模型的人体步态识别研究[J]. 电子测量技术,2022,45(4):120-125.
- [3] XIA C S, MUNAKATA A, SUGIURA Y C. Privacy-aware gait identification with ultralow-dimensional data using a distance sensor [J]. IEEE Sensors Journal, 2023, 23(9): 10109-10117.
- [4] CHAO H Q, WANG K, HE Y W, et al. GaitSet: Cross-view gait recognition through utilizing gait as a deep set[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2022, 44 (7): 3467-3478.
- [5] ZHANG Z, TRAN L, YIN X, et al. Gait recognition via disentangled representation learning [C]. CVPR, 2019.
- [6] ZHANG Y Z, HUANG Y Z, YU S Q, et al. Cross-view

- gait recognition by discriminative feature learning[J]. IEEE Transactions on Image Processing (TIP), 2020, 29: 1001-1015.
- [7] FAN C, PENG Y, CAO C, et al. Gaitpart: Temporal part-based model for gait recognition[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020, 14225-14233.
- [8] 李言, 曾维, 蒋毅, 等. 基于改进型对抗网络的步态特征提取方研究[J]. 电子测量技术, 2022, 45(9): 121-126.
- [9] WU Z F, HUANG Y Z, WANG L, et al. A comprehensive study on cross-view gait based human identification with deep CNNs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017, 39(2): 209-226.
- [10] 张红颖, 田鹏华. 结合残差网络与多级分块结构的步态识别方法[J]. 电子测量与仪器学报, 2022, 36(6): 66-72.
- [11] 张智, 常超伟, 王雷, 等. 结合整体和局部特征的步态识别方法[J]. 火力与指挥控制, 2023, 48(4): 141-146.
- [12] TAKEMURA N, MAKIHARA Y, MURAMATSU D, et al. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition[J]. Ijps Transactions on Computer Vision & Applications, 2018, 10(1): 1-14.
- [13] BEN X Y, GONG C, ZHANG P, et al. Coupled patch alignment for matching cross-view gaits [J]. IEEE Transactions on Image Processing (TIP), 2019, 28(6): 3142-3157.
- [14] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. Computer Science, 2014, 1-15, DOI: 10.48550/arXiv.1412.6980.
- [15] YU S Q, CHEN H F, REYES E B, et al. GaitGAN: Invariant gait feature extraction using generative adversarial networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, 532-539.
- [16] HE Y W, ZHANG J P, SHAN H G, et al. Multi-task GANs for view-specific feature learning in gait recognition [J]. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 102-113.
- [17] ZHANG Z Y, TRAN L, LIU F, et al. On learning disentangled representations for gait recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2022, 44(1): 345-360.

作者简介

邹雪, 硕士研究生, 主要研究方向为图像处理与步态识别。

E-mail: 2217060877@qq.com

谭棉, 副教授, 主要研究方向为工业缺陷检测与智能算法。

E-mail: tanmian@gzmu.edu.cn

严晓波, 硕士, 主要研究方向为图像处理、模式识别、统计建模与应用及大数据分析。

E-mail: xbYan@gzmu.edu.cn

王飞, 硕士, 主要研究方向为计算机视觉、深度学习。

E-mail: wangfei10248@163.com

王林(通信作者), 教授, 主要研究方向为图像处理与机器视觉。

E-mail: wanglin@gzmu.edu.cn