高技术通讯 2026 年 第36卷 第7期:677 - 687

doi:10.3772/j.issn.1002 - 0470.2026.07.002

基于反残差自注意力模块与多模态、多特征融合的步态识别算法研究①

赵昶辰②* 刘 凯* 黄 斌** 冯远静*

(*浙江工业大学信息工程学院 杭州 310023)

(**北京航空航天大学杭州创新研究院 杭州 310051)

摘 要 基于可穿戴设备的步态识别是实现下肢穿戴设备人机交互的基础。考虑到单一模态、单一特征不能提供完整的信息来识别步态,且当前步态识别领域缺少结合注意力机制的轻量化多模态算法研究,本文提出一种融合深度学习特征与传统时域特征的轻量化多模态步态识别算法,探索多模态信号融合中2种特征新的融合范式。该算法分别对姿态信号和肌电(electromyography,EMG)信号同时提取时域特征与深度学习特征,采取层级式融合方式将二者分步融合。基于自注意力机制和倒置残差结构设计了一个轻量化的特征提取模块,在保留卷积神经网络提取能力的同时也增加了远程信息交互能力。同时,设计了一个融合模块来融合不同模态的特征,增强特征中有用信息同时忽略次要信息。该模型在公开数据集ENABL3S和自建数据集上的准确率分别为98.16%和95.38%。此外,本文还分析了惯性测量单元(inertial measurement unit,IMU)和EMG这2种模态的2种特征分别对步态识别精度的贡献,为后续步态识别算法的研究者提供了参考。

关键词 步态识别;肌电;惯性测量单元;自注意力;时域特征

多模态步态识别通过综合分析多种传感信号来识别人体的基本步态,如平地行走(level walking,LW)、上斜坡(ramp ascent,RA)、下斜坡(ramp descent,RD)、上楼梯(stair ascent,SA)和下楼梯(stair descent,SD)等,在康复机器人、助力外骨骼和仿生假肢等诸多应用领域都有着广泛的应用[1]。步态识别也是各类助力设备实现人机交互的关键[2]。助力设备根据识别的结果来调整控制策略或算法参数[3]。识别精度的高低会直接影响设备的控制效果,从而影响使用体验,甚至危及使用者的生命安全[4]。因此,如何提高运动意图识别算法的精度对助力设备的使用安全、用户体验以及使用效果具有重要意义。

生物电信号和机械信号是2种常用的步态识别传感信号[5-7]。生物电信号一般包括脑电(electroencephalography,EEG)、眼电(electrooculography,EOG)和肌电(electromyography,EMG)信号[8]。其中,肌电信号是下肢运动意图识别中常用的生物电信号,它是由贴在皮肤表面的电极采集的运动时肌肉收缩和舒张时的生物电信号。不同运动状态下肌肉收缩的程度和速度不同,肌电信号也会不同。机械信号一般是指加速度、角速度、足底压力和关节角等运动学、动力学信号。这些信号能更直观地反映不同种类的运动。不同模态的传感器描述相同步态不同方面的信息,这些信息的融合有利于提高识别精度。早期步态识别算法大多基于单一传感器[9-10]。近期一些研究使用了多种传感信号[11-12],但是大多只是将多种信号进行简单拼接,缺乏对模态间信号相互依赖性的深层分析,无法充分挖掘模态间的互补信息。

① 国家自然科学基金面上项目(62573171)和国家自然科学基金青年基金(62203037)资助项目。

② 男,1988年生,博士,副教授;研究方向:计算机视觉,视觉生理信号测量与人工智能;联系人,E-mail:cczhao@zjut.edu.cn。

(收稿日期:2024 - 03 - 08)

步态识别所使用的模型主要包括传统机器学习分类器以及深度神经网络。基于传统机器学习分类器的方法[9,13]常使用的特征包括时域特征、频域特征以及时频域特征,这类特征具有明确的物理意义,能够直观反映运动状态,且易于实现,对计算资源依赖较少。但这类特征较为简单、无法反映模态更深层次的特性。基于深度神经网络的方法[4,14-15]能自动提取更加深层和复杂的特征,在处理大规模数据时效果常常优于传统算法。但需要消耗大量的计算资源,且模型缺乏可解释性。2 种方法各有优缺点,且均已经在步态识别领域中取得了不错的效果。

为了最大程度地利用不同模态之间的信息,并发挥传统特征和深度学习特征各自的优势,同时弥补它们的劣势,本文提出一种融合深度学习特征与传统时域特征的多模态步态识别算法。该算法以惯性测量单元(inertial measurement unit,IMU)的姿态信号和 EMG 作为输入数据,探索不同模态特征之间新的融合范式,以期进一步提高步态识别的准确性。首先,基于自注意力机制和反残差结构设计了一个名为反残差自注意力模块(inverted residual self-attention,IRSA)的轻量化特征提取模块,以取代传统卷积提取深度学习特征,同时采用了 2 个编码器分别同时提取 IMU 和 EMG 特征,以增强网络的建模能力。其次,提出了一种多模态特征融合模块来自适应学习融合权重,并通过按通道加权求和,实现不同模态之间互补特征的融合和无关特征的抑制。最后,基于注意力池化层设计了全局特征表示,并通过 3 层线性层生成最终输出。此外,为弥补公开数据集的不足,本文收集了用于评估步态识别算法的基准数据集。本文进行了大量的对比实验和消融实验,证明了该算法特征提取的高效性和特征融合方式的有效性,也证明了结合深度学习特征和传统时域特征可以充分利用不同模态的信息进一步提高步态识别算法精度。

1 相关工作

早期的步态识别算法大多基于传统机器学习的方法。支持向量机(support vector machine,SVM)[16]、K-近邻(K-nearest neighbor,KNN)[9]、隐含狄利克雷分配(latent Dirichlet allocation,LDA)[17]和隐马尔科夫模型(hidden Markov model,HMM)[18]等方法都已经在步态识别领域取得了不错的效果,但上述模型在物理变化和人类生理变化等环境下意图识别算法可靠性和精度会下降。Liu 等人[19]用基于熵的自适应算法、从测试数据学习算法和转换支持向量机 3 种自适应分类器来增强算法的可靠性。上述方法都需要手动设计特征,而且特征不全面。基于深度学习的方法实现了特征的自动提取,且可以提取到更加丰富的特征,在步态识别领域的应用也越来越广泛。Su 等人[14]利用卷积神经网络(convolutional neural network,CNN)来识别 IMU 信号,通过使用多种尺度的卷积层来提取特征,从而提高了算法的鲁棒性。Su 等人[14]的方法对稳态步态识别效果不错,但对连续步态识别效果不佳。Lu 等人[15]为了克服连续意图识别分类器在分类概率不确定时,预测类别在不同类别之间切换的问题,提出了一种置信指数的度量标准来衡量分类器在输出中的确定程度,只在置信指数高时输出类别,他们的算法对连续步态的识别有较高的精度。

当前步态识别主流算法仍以传统机器学习和深度学习为主导,如何提高信息的利用效率成为提高识别精度的关键元素。已有研究从修改输入入手,通过对输入的时序信号进行线性插值、距离矩阵、图像变化和转化成频谱图等操作来提高最终的识别精度[20-21]。由于步态信号是时序信号,而上述深度学习方法都没有利用到时序依赖关系,已有研究者开始使用循环神经网络来进行步态识别[23-24]。Tran 等人[25]设计了一个新型的长短记忆网络(long short term memory,LSTM)用来进行步态识别,并且实验验证了所提出的 LSTM 结构优于其他 LSTM 结构。考虑到不同模态的传感器能够提供互补信息,有些研究者同时使用 IMU、EMG 或 EEG 等多种信息的组合来识别步态[26-27],但是它们融合方式较简单,只是简单对原始输入或中间特征的简单拼接,无法充分挖掘多模态的互补信息。

Transformer[28]是近些年最成功的深度学习模型之一,也推动了许多领域的发展。Transformer 中的

自注意力机制能够捕捉序列中的长距离依赖关系,有助于提高模型对时间序列数据的处理能力,这对于时序信号的分析是非常重要的。Zhang 等人[29]基于 Transformer 设计了 ViT-LLMR(vision Transformer based lower limb motion recognition)架构用于融合多通道 EEG 信号和 IMU 数据,以识别下肢运动。虽然他们的模型取得了较高的精度,但是模型参数量过于庞大,而步态识别算法需要部署到嵌入式设备或智能穿戴设备,这种大模型显然是不合适的。步态识别算法的使用场景要求在提升精度的同时保持模型的轻量化。近期在其他领域,模型的轻量化研究取得了较大进步,MobileNet[30]提出了深度可分离卷积(depth-wise convolution,Dw-Conv)来减少计算量和参数。MobileNetv2[31]提出了一种基于 Dw-Conv 的倒置残差块(inverted residual block,IRB)。Zhang 等人[32]利用 MobileNetv2 中的倒置残差块和 Transformer 中的多头注意力机制,只使用基本运算符基于注意力的模型构建一个轻量级的类似 IRB 的倒置残余移动块。这种结合注意力机制的轻量化的研究恰好满足运动意图识别的需要,促使注意力机制应用到步态意图识别。

增加模态种类和提高信息的利用效率是提升步态识别精度的主要手段。目前算法仍然存在一些不足。首先,缺少探究不同模态信息的融合方式的研究,存在信息冗余和权重分配的问题。其次,缺少结合注意力机制的轻量化识别模型,目前方法大多只关注了数据的局部依赖,或者采用了过于庞大的模型。为充分探索不同模态特征的提取方式以及充分挖掘不同模态特征的互补性,有效利用不同模态中的有用信息,同时减少无效信息的干扰,本文提出一种基于反残差自注意力与多模态、多特征融合的步态识别算法。基于注意力机制和反残差结构设计了一个轻量化的特征提取模块来代替普通卷积提取深度学习特征。本文算法首先利用设计的特征提取模块充分提取 IMU 和 EMG 的深度学习特征,通过多模态特征融合模块对 2 种模态的深度学习特征进行加权融合,保存有用信息同时忽略次要信息,融合的深度学习特征与手工设计与提取的时域特征进行融合,最后通过反残差自注意力模块、全局特征表示和线性层输出最终的识别结果。

2 本文步态识别算法

2.1 网络总体结构

本文步态识别网络的总体结构如图 1 所示,主要由反残差自注意力模块、时域特征提取模块、特征融合模块、全局特征表示层和解码器几部分组成。

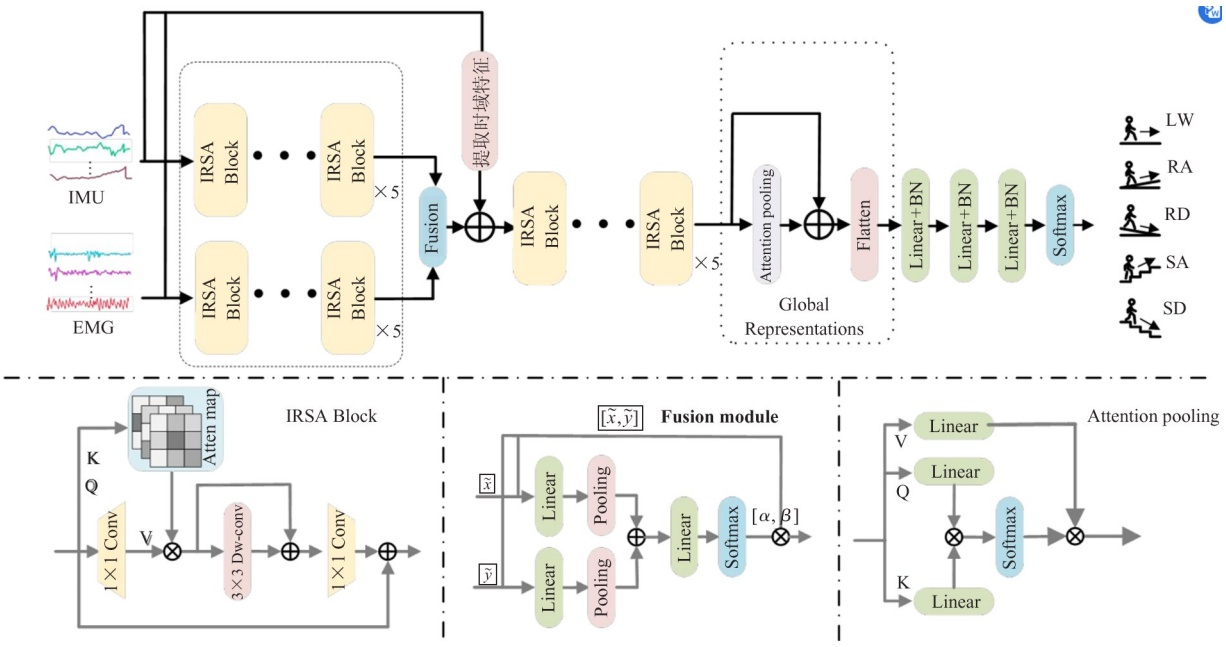


图 1 下肢运动模式识别网络的总体结构框图

该网络的输入是同一时刻滑动窗口中 IMU 信号和 EMG 信号,输出为 5 种常见步态类别,分别为平地行走,上、下斜坡以及上、下楼梯。该网络分别提取 IMU 和 EMG 信号的深度学习特征和传统时域特征,深度学习特征通过特征融合模块进行融合,再与时域特征结合,由反残差自注意力模块、全局特征表示和线性层构成的解码器输出最终的识别结果。

2.2　反残差自注意力模块

反残差自注意力模块被用来提取 IMU 和 EMG 数据的深度学习特征。这一模块的独特之处在于综合了注意力机制和倒置残差结构的优势,吸收了 CNN 建模局部特征的效率的同时也增加了自注意力机制的动态建模能力来学习远程交互,并使用了 1×1 卷积来对通道数进行扩张和压缩,以此优化计算效率,模块中扩展和压缩倍数为 6。IRSA 模块中的深度可分离卷积用于捕捉空间特征,同时模块使用了残差连接来提高性能和表征能力。IRSA 模块的通道扩展和压缩操作使得输入和输出的通道数相同。

自注意力机制包括 3 个输入:查询向量(query,Q)、键向量(key,K)和值向量(value,V),通过 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 捕捉元素之间的关联得到注意力矩阵,利用注意力权重对值向量加权融合得到最终的自注意力输出,计算公式为

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

式中:d_k 是查询向量 $\boldsymbol{Q}$ 和键向量 $\boldsymbol{K}$ 的维度,上标 T 表示矩阵转置。自注意力机制的灵活性使得模块能够自适应于不同时间步的运动特征,捕捉运动序列中关键部分的长距离依赖关系。这种动态关注的机制能够更精准地捕捉运动序列中的时空关联,为模型提供更强大的建模能力。通过在每个时间步动态计算注意力权重,模块能够更好地适应运动序列中变化的动作模式,从而提高了下肢运动意图的识别性能。整合了自注意力机制的模块,将更有可能在复杂的下肢运动模式中发现关键的动作特征,为识别运动意图提供更为准确和全面的信息。

反残差自注意力模块是基于反残差结构设计的。深度可分离卷积是该结构的关键,它通过将普通卷积分解为深度卷积和逐点卷积来减少模型的参数量和计算复杂度。深度卷积在每个输入通道单独进行卷积,每个通道都有自己的深度卷积核,有效地整合了信息。逐点卷积则对深度学习生成的特征图进行融合同时改变通道数。深度可分离卷积在减少计算量的同时也保持相对较好的性能,这对于移动设备、嵌入式系统或其他资源受限的环境中的轻量级模型设计非常有用。

2.3　特征融合模块

特征融合模块旨在融合 IMU 与 EMG 这 2 种模态的深度特征。该特征融合模块在 IMU 和 EMG 深度学习特征的融合过程中,通过动态权重学习和多层次信息处理,能够更好地捕捉 2 种模态之间的关联,强化有效信息,从而提高了对下肢运动意图识别任务的建模能力和性能。2 种模态的特征融合是通过加权相加的方式进行,见式(2)。

$$\boldsymbol{z} = [\alpha \ \ \beta]\begin{bmatrix}\tilde{\boldsymbol{x}}\\ \tilde{\boldsymbol{y}}\end{bmatrix} \tag{2}$$

式中:$\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \mathbb{R}^{T\times C}$ 为待融合特征,C 为通道数;α,$\beta \in \mathbb{R}^{C}$ 为权重。如图 1 中下所示,融合模块的权重 α、β 是网络训练过程中自动学习得到,它可被公式化,如式(3)所示。待融合的特征首先分别通过一层池化层将特征压缩成一维,然后通过一层共享权重的卷积层,再将 2 个模态的特征进行拼接,最后通过线性层与 SoftMax 层即可得到每个通道上融合权重。融合权重 α、β 与输入的特征图 $\tilde{\boldsymbol{x}}$、$\tilde{\boldsymbol{y}}$ 相乘得到融合的特征图 $\boldsymbol{z}$。

$$[\alpha \ \ \beta] = f(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \tag{3}$$

式中:$f(\cdot)$ 表示融合权重生成函数。

2.4　时域特征提取

时域特征可以提供一些深度学习特征可能忽略的局部运动信息,且包含更细粒度的运动信息,与深度学习结合有助于提供更全面的特征表示。深度学习模型对于噪声或突发性的运动变化较为敏感,时域特征的引入有助于对抗噪声干扰,提高模型的鲁棒性。本文使用平均绝对值、过零次数、波形长度、斜率变化次数和 6 阶自回归模型系数等共 10 个特征。这些特征的定义如下。

平均绝对值反映了滑动窗口内肌电信号的强

度,见式(4)。

$$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N} |\boldsymbol{x}_i| \tag{4}$$

式中:MAV 为平均绝对值,$\boldsymbol{x}_i$ 表示 i 时刻的肌电信号,N 表示信号窗口的长度,$|\cdot|$表示取绝对值。

过零次数表示信号波形穿过 x 轴的次数,见式(5)。

$$\mathrm{ZC} = \sum_{i=1}^{N-1}[\mathrm{sgn}(\boldsymbol{x}_i) \neq \mathrm{sgn}(\boldsymbol{x}_{i+1})] \tag{5}$$

式中:ZC 表示过零次数;sgn 为符号函数,当输入为正时返回 1,输入为负返回 -1,输入为 0 则返回 0。

波形长度反映区间内相邻各数值之差的绝对值之和,见式(6)。

$$\mathrm{WL} = \sum_{i=1}^{N-1} |\boldsymbol{x}_{i+1} - \boldsymbol{x}_i| \tag{6}$$

式中:WL 表示波形长度。

斜率变化次数表示波形在区间内斜率变化的次数,见式(7)。

$$\mathrm{SC} = \sum_{i=2}^{N-1}[\mathrm{sgn}((\boldsymbol{x}_{i+1} - \boldsymbol{x}_i)(\boldsymbol{x}_i - \boldsymbol{x}_{i-1})) - 1] \tag{7}$$

式中:SC 表示斜率变化次数。

自回归特征基于自回归模型(autoregression, AR),描述时间序列信号的自相关性。自回归系数反映了当前值和历史值之间的线性关系,见式(8)。

$$\boldsymbol{x}_n = \sum_{i=1}^{P} a_i \boldsymbol{x}_{n-i} + \varepsilon \tag{8}$$

式中:P 为回归模型阶数,本文为 6;a_i 表示系数,是本文需要提取的特征;ε 是随机误差,$\varepsilon \in N(0,\sigma^2)$。

2.5 全局特征表示

受到 Transformer 启发,Alec 等人[33]提出了注意力池化层,用以替代传统的全局平均池化。在这个注意力池化层中,查询向量受到图像的全局平均池化所代表的条件的影响。由于简单的池化层可能使运动学信息丢失,本文参考了文献[34]中的全局表示方式,基于该注意力池化层设计了全局特征表示层以聚集关键特征信息,并将其与残差结构结合以获得全局表征。最终得到的全局特征表示如式(9)所示。

$$q^{\mathrm{Global}} = \mathrm{Flatten}(\mathrm{Concat}(F_{\mathrm{L}}, \mathrm{AttenPool}(F_{\mathrm{L}})) \tag{9}$$

式中:q^{Global} 是得到的全局特征表示;AttenPool(·)表示注意力池化操作;Concat(·)表示将 FQ 与其注意力池化结果拼接;Flatten(·)表示展平操作;F_{L} 表示待聚合特征。

2.6 损失函数

本文采用的损失函数为交叉熵损失(cross entropy loss),广泛应用于神经网络的训练,见式(10)。

$$H(p,q) = -\sum_{\boldsymbol{x}} p(\boldsymbol{a})\log(q(\boldsymbol{a})) \tag{10}$$

式中:$\boldsymbol{a}$ 为输入,p 为真实标签值,q 为预测值。

3 实验设置

3.1 ENABL3S

健全人双侧下肢运动信号数据集(encyclopedia of able-bodied bilateral lower limb locomotor signals, ENABL3S)是一个由运动学信号和肌电信号组成的公开步态数据集[35]。该数据集中包含平地行走、上斜坡、下斜坡、上楼梯和下楼梯 5 种日常步态。该数据集共使用 5 个 IMU 传感器分别采集两侧大腿、小腿和腰部的姿态数据,同时使用 14 个 EMG 传感器采集下肢 7 块肌肉的肌电信号,包括胫骨前肌、腓肠肌内侧肌、比目鱼肌、股外侧肌、股骨直肌、股二头肌和半腱肌。最终得到 30 个通道的 IMU 数据和 14 个通道的 EMG 数据。

3.2 自建数据集

为了验证本文方法在不同传感器设置和采集环境下的普适性,本文采集了含有 IMU 和 EMG 数据的步态数据集。采集受试者为 8 男、2 女,年龄 25 ± 3 岁,身高为 165 ± 20 cm,体重为 70 ± 20 kg。所有受试者均没有步行障碍,也无其他影响步行的相关疾病。

采集数据选用的 IMU 传感器为维特智能公司的六轴加速度计 BWT61CL,可以采集 3 个方向上的加速度(a_x, a_y, a_z)和角速度(g_x, g_y, g_z)。EMG 传感器为美国 Delsys 公司的 Trigno EMG 传感器。IMU 信号通过蓝牙回传至上位机,EMG 信号通过无线方式先传回基站,由基站传回上位机。上位机根据时间戳对信号进行同步。采集详细信息如表 1 所示。

IMU 和 EMG 传感器的放置位置如图 2 所示,共采集了步行时腿部 4 处 IMU 信号和 8 处 EMG 信号。4个IMU传感器使用医用胶布固定,分别检测双侧

表 1　IMU 和 EMG 传感器的规格参数

传感器	IMU	EMG
传感器型号	BWT61CL	Trigno EMG
净重/g	18.8	14.0
通信距离/m	<20	>40
采样频率/Hz	100	2 000
续航/h	4	6 ~ 8
通道数目	6(ax, ay, az, gx, gy, gz)	1
量程	加速度:±16 g 角速度:±2 000 °/s	12 mV
使用数量	4	8

图 2　IMU 和 EMG 传感器的放置位置

大小腿的加速度和角速度。8 个 EMG 传感器则使用双面粘合剂固定在对应肌肉表面,用来检测双侧股直肌,胫骨前肌、半腱肌和腓肠肌的肌电信号,这些肌肉在进行下肢运动时的肌电信号较为明显。数据采集场景包括平地,16 cm 阶梯和高宽比为 1∶12 的坡道 3 种环境。所有受试者被要求在上述环境中按照他们最舒适的速度重复平地行走、上/下斜坡、上/下楼梯 5 种步态各 30 次。在采集的过程中,为了防止肌肉疲劳对数据造成影响,每采集 10 组数据休息 5 min。

3.3　数据预处理

为了消除传感器噪声以提高识别的精度,本文对采集的原始数据进行了滤波处理,滤波方式采用小波基为多贝西小波 8 的小波阈值滤波,选择 VisuShrink 为阈值,选用的阈值函数为软、硬阈值之间的折中。由于 IMU 和 EMG 传感器的采样频率存在差异,对采集到的 IMU 和 EMG 信号分别进行插值和降采样,使得采样频率统一为 500 Hz。使用长度为 300 ms,步长为 50 ms 的滑动窗口对连续的 IMU 和 EMG 信号进行同步采样。为了加速模型的收敛,在利用神经网络进行预测之前,还需要对数据进行归一化。利用 min-max 归一化对窗口内的数据进行归一化。min-max 归一化通过对原始数据进行线性变化,使其数值映射到[0,1],见式(11)。

$$\boldsymbol{X}_{\text{norm}} = \frac{\boldsymbol{X} - \boldsymbol{X}_{\min}}{\boldsymbol{X}_{\max} - \boldsymbol{X}_{\min}} \tag{11}$$

式中:$\boldsymbol{X}$ 表示输入数据,$\boldsymbol{X}_{\min}$ 表示输入数据中的最小值,$\boldsymbol{X}_{\max}$ 为最大值。

3.4　超参数设置

模型的训练和预测在 NVIDIA GeForce RTX 3080 Ti 显卡上进行。本文中的所有实验均进行 5 轮训练,实验结果为 5 轮实验结果的平均值。实验中模型训练时的各个超参数设置如表 2 所示。

表 2　训练模型时参数的设置

参数	数值
# of epochs	50
Batch size	32
优化器	SGD
学习率	0.001
Weight decay	0.000 1
Momentum	0.000 5

3.5　对比方法和评价指标

实验选用的对比算法包括 Su 等人[14]、Lu 等人[15]、Lee 等人[22] 和 Zhang 等人[32] 提出的基于深度神经网络的分类方法,以及 LDA、SVM 和 KNN 这 3 种基于传统机器学习分类器的方法。实验中,所有算法的输入是同一时间窗口内的 IMU 和 EMG 信号。在使用传统机器学习分类器时,对 IMU 信号提取了平均值、标准差、最大值、最小值、初始值和最终值等特征,对 EMG 信号提取了平均绝对值(mean absolute value,MAV)、过零点数(zero crossing,ZC)、斜率变化数(signal slope changes,SSC)和 6 阶自回归模型系数等特征[13,33]。

实验中采用的评价指标为准确度(accuracy,Acc)、召回率(recall,Rec)和 $F1$ 分数。

Acc 表示 5 种步态中所有预测样本中,预测正确的步态样本数目占总样本数的比例,见式(12)。

$$\mathrm{Acc} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{TP}_i + \mathrm{TN}_i}{\mathrm{TP}_i + \mathrm{TN}_i + \mathrm{FP}_i + \mathrm{FN}_i} \tag{12}$$

式中:$n=5$ 表示 5 种不同步态,TP_i 表示真实标签为 i 的样本被预测正确的数量,TN_i 表示真实标签不为 i 的且被预测正确的数量,FP_i 表示真实标签为 i 的样本被预测为其他类别的数量,FN_i 表示预测标签为 i 而真实标签为其他类别的样本数目。

Rec 是所有实际为 i 类中真正为 i 类的样本占实际为 i 类样本的比例,见式(13)。

$$\mathrm{Rec} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} \tag{13}$$

$F1$ 分数是评价分类性能的关键评价指标,用于综合考虑召回率和精准率(precision,Pre)的表现,见式(14)。

$$F1 = \frac{2\times\mathrm{Pre}\times\mathrm{Rec}}{\mathrm{Pre}+\mathrm{Rec}} \tag{14}$$

式中:Pre 定义如式(15)所示。

$$\mathrm{Pre} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i} \tag{15}$$

4 实验结果

4.1 分类实验结果

表 3 为本文算法和对比算法在 ENABL3S 数据集和自建数据集上的实验结果。从 5 种步态识别的平均结果来看,本文算法的平均 Acc、Rec 和 $F1$ 分数分别为 98.16%、98.19% 和 98.17%,均优于其他的所有对比算法。Su、Lu、Lee 和 Zhang 等人提出的深度学习算法的 Acc、Rec 和 $F1$ 分数也都超过了 95.00%,其中 Zhang 的识别 Acc、Rec 和 $F1$ 分数分别为 97.15%、97.16% 和 97.13%,在所有的对比算法中排名第 2,这凸显了深度学习方法和注意力机制在意图识别中的优越性。在传统方法中 LDA 和 SVM 的结果表现效果不佳,Acc、Rec 和 $F1$ 分数均低于 90.00%,而 KNN 的识别效果良好,仅次于本文算法、Lu 和 Zhang 的算法。在自建数据集上,本文算法在 5 种步态上的平均 Acc、Rec 和 $F1$ 分数分别为 95.38%、95.35% 和 95.37%,仍然优于其他对比算法。在对比的 4 种深度学习方法中,Zhang 的算法的识别效果在所有的对比方法中排第 2,仅次于本文算法,Su 的识别效果不佳,Acc、Rec 和 $F1$ 分数均低于 90.00%,而 Lu 和 Lee 的算法效果接近,二者的 Acc、Rec 和 $F1$ 分数均在 92.00% 左右。传统方法中,LDA 和 SVM 的 Acc、Rec 和 $F1$ 分数均低于 85.00%,效果不如深度学习的方法,而 KNN 的识别精度较高,仅次于本文算法和 Zhang 的算法。

表 3 本文算法与对比方法在 2 个数据集上分类准确度、召回率和 $F1$ 分数的比较 %

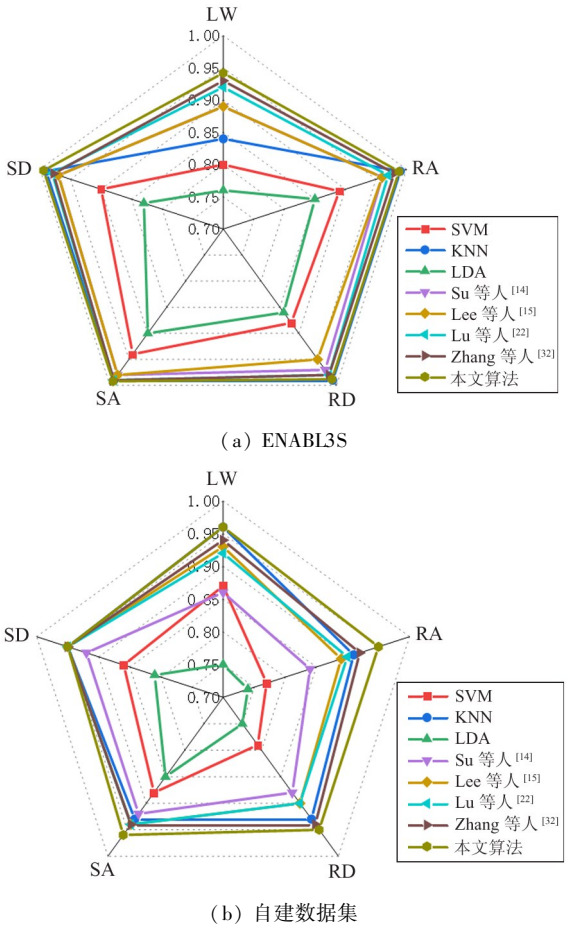
数据集	评价指标	LDA	SVM	KNN	Su 等人[14]	Lu 等人[15]	Lee 等人[22]	Zhang 等人[32]	本文算法
ENABL3S	Acc	84.16	88.88	96.20	95.46	96.38	95.60	97.15	98.16
	Rec	84.17	88.89	96.24	95.49	96.40	95.57	97.16	98.19
	$F1$	84.19	88.87	96.23	95.48	96.39	95.59	97.13	98.17
自建数据集	Acc	78.22	83.27	93.37	88.57	92.02	92.36	93.62	95.38
	Rec	78.19	83.25	93.39	88.57	91.98	92.34	93.61	95.35
	$F1$	78.18	83.27	93.38	88.57	92.01	92.35	93.64	95.37

图 3 为对比算法和本文算法在 5 种步态上识别精度的雷达图,5 个轴分别代表 5 种步态,轴上离原点距离越远表示识别准确度越高。在 2 个数据集上,本文算法 5 种步态的识别精度都是最优或并列最优。在 ENABL3S 数据集上,本文算法在平地行走步态相较于其他算法精度有较大提升。在自建数据集上,平地行走、上斜坡和上楼梯 3 种步态上都有较大提升。在 2 个数据集上的实验结果表明,本文算法对于 5 种步态的识别能力都较好。本文算法在不同实验环境与采集设备下的 5 种步态数据均取得了不错的识别效果,说明该算法具有较强的鲁棒性。

4.2 消融实验

为验证不同传感器的不同特征对运动意图识别精度的贡献,本文对 IMU 和 EMG 提取了不同特征进行了实验。如图 4 所示,Abn1 和 Abn2 分别单独对 IMU 和 EMG 信号提取时域特征,Abn3 组合了

IMU 和 EMG 的时域特征，Abn4 和 Abn5 则分别提取 IMU 和 EMG 的深度学习特征，Abn6 将 IMU 和 EMG 拼接后使用 1 个编码器来共同提取深度学习特征，Abn7 则是使用 2 个编码器分别提取 IMU 和 EMG 的深度学习特征再融合。实验结果如表 4 所示。在 2 个数据集上，不同的网络结构识别精度趋势变化一致。Abn1、Abn2、Abn4 和 Abn5 的实验结果表明，无论是时域特征还是深度学习特征，利用 IMU 进行识别的精度都优于 EMG 的识别精度。其中，使用 CNN 编码器提取 IMU 特征在所有单模态识别方法中效果最好。Abn3 和 Abn6 的结果显示，多模态相较于单模态的特征能提高识别精度，这也说明不同模态之间能相互提供互补信息。Abn7 利用双通道提取 IMU 和 EMG 的网络结构能够充分提取 IMU 和 EMG 的特有特征，同时利用融合模块加强有用信息的权重来融合二者的特征。最终的网络结构在充分利用不同模态的深度学习特征同时，加入时域特征，识别精度也进一步提高。



（a）ENABL3S

（b）自建数据集

图 3　雷达图

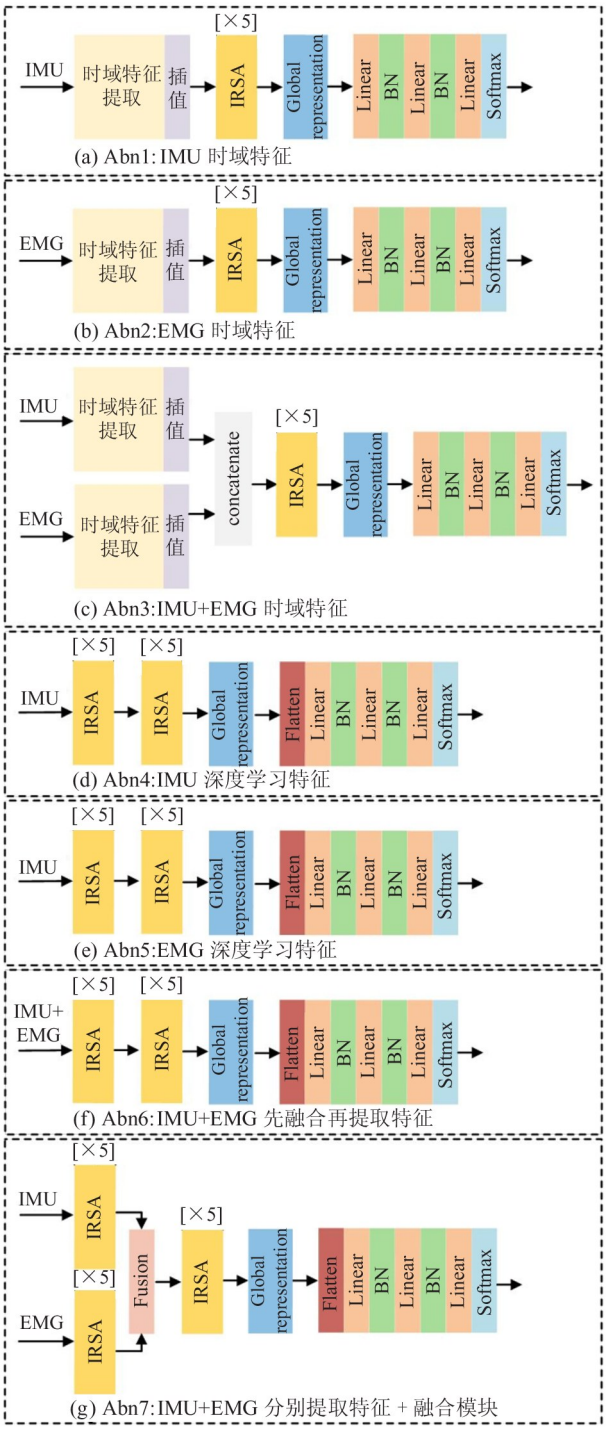


图 4　消融实验中的网络结构配置

为验证 IRSA 模块和全局表示层的效果，对这 2 个模块进行了消融实验，对于 IRSA 模块的消融，将其替代为卷积层、批量归一化层和 ReLU 激活函数结合的常规卷积模块。实验结果如表 5 所示，当把特征提取模块换成常规卷积模块时，ENABL3S 上准确度降低了0.59%，自建数据集上降低了1.10%，

表 4　消融实验中不同网络结构的准确度比较　%

	Abn1	Abn2	Abn3	Abn4	Abn5	Abn6	Abn7	本文算法
ENABL3S	94.24	89.85	95.77	97.15	78.72	97.26	97.93	98.16
自建数据集	88.80	80.54	94.01	93.62	66.42	94.26	94.66	95.38

表 5　特征提取模块和全局特征表示层消融实验中准确度比较结果　%

	普通卷积	无全局表示	所提网络
ENABL3S	97.57	97.87	98.16
自建数据集	94.28	94.78	95.38

说明所提的特征提取模块可提取到比普通卷积更加丰富的特征。当去掉全局表示层后,ENABL3S 上准确度降低了 0.29%,自建数据集上降低了 0.60%,这也证明了基于注意力池化层设计的全局表示层可以增强特征表示。

为测试所提网络对 IRSA 模块数量的敏感性,本文将所提算法的 IRSA 模块数量分别改为 3、4、5 和 6,并在 2 个数据集上进行了实验,结果如表 6 所示。实验结果表明,当模块的数目小于 5 时,随着数量的增加,识别精度也随之增加;当模块数为 5 时,2 个数据集上的识别精度达到最高;当模块数增加到 6 层时,识别精度下降。出现这种现象的原因在于,当模块数较少时,网络难以捕捉到深层特征,类别区分度不高,从而导致识别精度受到限制。随着 IRSA 模块数量的增加,网络可以更好地捕捉到高层特征,类别区分度上升,从而提高网络的识别精度。当数目继续增加时,参数量变多,模型复杂度增加,导致过拟合从而精度下降。实验结果证明,所提网络的使用 5 层 IRSA 模块是最优的,在提高精度的同时,也避免了过拟合的问题。

表 6　改变所提网络中编码器特征提取模块数量的准确度比较结果　%

	3 层	4 层	5 层	6 层
ENABL3S	97.57	97.87	98.16	98.06
自建数据集	94.28	94.78	95.38	95.06

5　结 论

本文针对人体行走步态识别算法中多模态信号的特征提取方式和融合方式如何提高识别精度这一问题,提出一种融合深度学习特征与传统时域特征的步态识别算法,该算法通过利用轻量化的反残差自注意力特征提取模块来分别提取 IMU 和 EMG 信号的深度学习特征,同时对其提取时域特征。提出特征融合模块对深度学习特征进行融合,并在融合之后与时域特征拼接,经过解码器得到最终的行走步态输出。在公开数据集 ENABL3S 和自建数据集上进行了性能验证,相比于现有方法取得了较高的识别精度。本研究表明,将深度学习特征和传统时域特征进行融合的方式有助于提升步态识别的精度,并且优于仅使用深度学习的方法和仅使用传统方法提取特征的方法,这也说明不同模态和不同特征之间都存在互补信息。未来工作将进一步探索这种融合方式下的融合和解码,以期获得更加高效、准确的步态识别性能。

参考文献

[1] 丁其川,熊安斌,赵新刚,等. 基于表面肌电的运动意图识别方法研究及应用综述[J]. 自动化学报, 2016, 42(1):13-25.

[2] 彭亮,侯增广,王晨,等. 康复辅助机器人及其物理人机交互方法[J]. 自动化学报, 2018,44(11):2000-2010.

[3] 王启宁,郑恩昊,陈保君,等. 面向人机融合的智能动力下肢假肢研究现状与挑战[J]. 自动化学报, 2016, 42(12):1780-1793.

[4] 苏本跃,倪钰,盛敏,等. 基于改进卷积神经网络的动力下肢假肢运动意图识别[J]. 控制与决策, 2021, 36(12):3031-3038.

[5] Lu Z, Zhuo W, Zhang N, et al. A novel motion intention recognition approach for soft exoskeleton via IMU[J]. Electronics, 2020,9(12):2176.

[6] Zhang P F, Gao X S, Miao M D, et al. Design and control of a lower limb rehabilitation robot based on human motion intention recognition with multi-source sensor information[J]. Machines, 2022,10(12):1125.

[7] Maike S, Dominik S, Axel G, et al. From the laboratory to the field: IMU-based shot and pass detection in football training and game scenarios using deep learning[J]. Sensors, 2021, 21(9): 3071.

[8] Nam Y, Koo B, Clchocki A, et al. GOM-Face: GKP, EOG, and EMG-based multimodal interface with application to humanoid robot control[J]. IEEE Transactions on Biomedical Engineering, 2014, 61(2): 453 – 462.

[9] Huang H, Todd A K, Robert D L. A strategy for identifying locomotion modes using surface electromyography [J]. IEEE Transactions on Biomedical Engineering, 2009, 56(1): 65 – 73.

[10] Jin D W, Yang J K, Zhang R H, et al. Terrain identification for prosthetic knees based on electromyographic signal features [J]. Tsinghua Science and Technology, 2006, 11: 74 – 79.

[11] Camargo J, Flanagan W, Csomay S, et al. A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors[J]. IEEE Transactions on Biomedical Engineering, 2021, 68(5): 1569 – 1578.

[12] Li H T, Han S L, Pan M C. Lower-limb motion classification for hemiparetic patients through IMU and EMG signal processing[C]// Proceedings of the International Conference on Biomedical Engineering. Hanoi, Vietnam: IEEE, 2016: 113 – 118.

[13] Hu B, Elliott R, Levi H. Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities[J]. Frontiers in Robotics and AI, 2018, 5: 78.

[14] Su B Y, Wang J, Liu S Q, et al. A CNN-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(5): 1032 – 1042.

[15] Lu Z, Narayan A, Yu H. A deep learning based end-to-end locomotion mode detection method for lower limb wearable robot control[C]// Proceedings of the International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE, 2020: 4091 – 4097.

[16] Zhu C Y, Luo L, Mai J G, et al. Recognizing continuous multiple degrees of freedom foot movements with inertial sensors[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 431 – 440.

[17] Chen T, Lin T Y, Hong Y W P. Gait phase segmentation using weighted dynamic time warping and k-nearest neighbors graph embedding[C]// Proceedings of the International Conference on Acoustics, Speech and Signal Processing. Barcelona, Spain: IEEE, 2020: 1180 – 1184.

[18] Liu Z J, Lin W, Geng Y L, et al. Intent pattern recognition of lower-limb motion based on mechanical sensors [J]. IEEE/CAA Journal of Automatica Sinica, 2017, 7(7): 651 – 660.

[19] Liu M, Zhang F, Huang H. An adaptive classification strategy for reliable locomotion mode recognition [J]. Sensors, 2017, 17(9): 1 – 18.

[20] Hur T, Bang J, Huynh T, et al. Iss2Image: a novel signal-encoding technique for CNN-based human activity recognition[J]. Sensors, 2018, 18(11): 3910.

[21] Zhang R F, Li C P. Motion sequence recognition with multi-sensors using deep convolutional neural network[C]// Proceedings of the Intelligent Data Analysis and Applications. Berlin, Germany: Springer, 2015: 13 – 23.

[22] Lee U H, Bi J, Patel R, et al. Image transformation and CNNs: a strategy for encoding human locomotor intent for autonomous wearable robots [J]. IEEE Robotics and Automation Letters, 2020, 5(4): 5440 – 5447.

[23] Zou Q, Wang Y L, Wang Q, et al. Deep learning-based gait recognition using smartphones in the wild[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3197 – 3212.

[24] Fernandez-lopez P, Liu-Jimenez J, Kiyokawa K, et al. Recurrent neural network for inertial gait user recognition in smartphones[J]. Sensors, 2019, 19(18): 4054.

[25] Tran L, Hoang T, Nguyen T, et al. Multi-model long short-term memory network for gait recognition using window-based data segment [J]. IEEE Access, 2021, 9: 23826 – 23839.

[26] Cheng X, Zhang D G. Lower limb motor intent recognition based on multi-source information[J], Machine Design and Research, 2020, 36(6): 54 – 58.

[27] Liu Y X, Wang R, Gutierrez-farewik E M. A muscle synergy-inspired method of detecting human movement intentions based on wearable sensor fusion [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 1089 – 1098.

[28] Ashish V, Noam S, Niki P, et al. Attention is all you

need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc, 2017:6000-6010.

[29] Zhang H Y, Ke Y, Cao G S, et al. ViT-LLMR: vision transformer-based lower limb motion recognition from fusion signals of MMG and IMU[J]. Biomedical Signal Processing and Control, 2023,82:104508.

[30] Andrew G H, Zhu M L, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2024-04-09]. https://arxiv.org/pdf/1704.04861.

[31] Sandler M, Howard A, Zhu M L, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018:4510-4520.

[32] Zhang J N, Li X T, Liu L, et al. Rethinking mobile block for efficient attention-based models[C]//Proceedings of the International Conference on Computer Vision. Paris, France: IEEE, 2023:1389-1400.

[33] Alec R, Jong W K, Chris H, et al. Learning transferable visual models from natural language supervision[EB/OL]. (2021-02-26)[2024-04-09]. https://arxiv.org/pdf/2103.00020.

[34] Li H, Zhu J G, Jiang X H, et al. Uni-perceiver v2: a generalist model for large-scale vision and vision-language tasks[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023:2691-2700.

[35] Hu B, Elliott, Levi H. Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals[J]. Frontiers in Robotics and AI, 2018,5:14.

A novel gait recognition algorithm based on inverted residual self-attention module and multi-modality, multi-feature fusion

Zhao Changchen*, Liu Kai*, Huang Bin**, Feng Yuanjing*

(*College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

(**Hangzhou Innovation Institute, Beihang University, Hangzhou 310051)

Abstract

Gait recognition is fundamental to human-machine interaction for lower-limb wearable devices. Considering that a single modality and unitary feature cannot provide comprehensive information for gait recognition, and that there is limited number of researches incorporating attention mechanisms, this paper proposes a lightweight multi-modal gait recognition algorithm that integrates deep learning features with traditional time-domain features and explores a new fusion paradigm of the two types of features. The proposed algorithm extracts time-domain features and deep learning features from both posture signals and electromyography (EMG) signals simultaneously and then adopts a hierarchical fusion method to integrate them step by step. A lightweight feature extraction module based on the self-attention mechanism and inverted residual structure is designed to retain the extraction capability of convolutional neural networks and increase the capacity for remote information interaction. Additionally, a fusion module is designed to integrate features from different modalities, enhancing the useful information in the features while ignoring the unimportant information. The proposed algorithm achieves accuracies of 98.16% and 95.38% on a public dataset ENABL3S and a self-collected dataset, respectively. Moreover, the contribution of inertial measurement unit (IMU) and EMG features to the gait recognition accuracy is analyzed, providing references for researchers in the subsequent gait recognition algorithm research.

Key words: gait recognition, electromyography, inertial measurement unit, self-attention, time domain feature