

基于SE-TCN-Transformer融合网络的下肢外骨骼机器人运动意图识别研究

高婧祯, 蓝艺亮, 张国成, 刘欣怡, 颜建军*

华东理工大学机械与动力工程学院, 上海

收稿日期: 2026年2月16日; 录用日期: 2026年3月13日; 发布日期: 2026年3月25日

摘要

针对下肢外骨骼机器人人机交互系统中对运动意图识别的高精度与实时性要求, 本文提出了一种基于SE-TCN-Transformer的新型融合网络架构。现有的卷积神经网络(CNN)和长短期记忆网络(LSTM)在处理长时序多模态传感器数据时, 往往面临局部特征提取不足、全局依赖捕捉困难以及难以有效抑制传感器噪声等挑战。为解决上述问题, 本研究设计了“局部特征增强-全局语义建模”的双阶段特征提取策略。首先, 引入嵌入Squeeze-and-Excitation (SE)通道注意力机制的时间卷积网络(Temporal Convolutional Network, TCN)作为前端, 通过因果空洞卷积提取高频局部运动特征, 并利用SE模块自适应重标定通道权重, 有效筛选关键传感器信息并抑制冗余噪声。其次, 采用Transformer编码器对特征序列进行全局建模, 利用多头自注意力机制精准捕捉步态相位间的长距离时序依赖。在公开数据集HuGaDB上的实验结果表明, 该模型对8种复杂日常步态的平均识别准确率达到96.73%, F1分数达到95.43%。此外, 基于Shapley Additive exPlanations (SHAP)的可解释性分析揭示, 足部垂直加速度与大腿角速度是步态识别中作用较大, 表明模型提取的关键特征分布与人体下肢生物力学的先验知识具有较高的一致性。特别是在坐下、起立等具有瞬态突变特征的转换动作识别中, 该模型表现出显著优势, 优于传统的SVM、LSTM及CNN-LSTM混合模型。该研究不仅验证了混合架构在多模态步态分析中的有效性, 也为外骨骼机器人的实时精准控制提供了有力的技术支撑。

关键词

下肢外骨骼, 步态识别, SE注意力机制, TCN, Transformer

Research on Motion Intention Recognition for Lower-Limb Exoskeleton Robots Based on SE-TCN-Transformer Fusion Network

*通讯作者。

文章引用: 高婧祯, 蓝艺亮, 张国成, 刘欣怡, 颜建军. 基于SE-TCN-Transformer融合网络的下肢外骨骼机器人运动意图识别研究[J]. 人工智能与机器人研究, 2026, 15(2): 660-671. DOI: 10.12677/airr.2026.152063

Jingzhen Gao, Yiliang Lan, Guocheng Zhang, Xinyi Liu, Jianjun Yan*

School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai

Received: February 16, 2026; accepted: March 13, 2026; published: March 25, 2026

Abstract

To achieve high-precision and real-time motion intention recognition for lower-limb exoskeletons, this paper proposes a novel hybrid architecture: the SE-TCN-Transformer. Addressing the limitations of conventional CNNs and LSTMs in processing long-sequence multimodal data, we introduce a “local enhancement-global modeling” strategy. First, a Temporal Convolutional Network (TCN) integrated with Squeeze-and-Excitation (SE) attention extracts high-frequency local features while adaptively suppressing sensor noise. Second, a Transformer encoder captures long-range temporal dependencies via multi-head self-attention. Experiments on the HuGaDB dataset demonstrate an average accuracy of 96.73% and an F1-score of 95.43% across eight gaits. SHAP-based interpretability analysis identifies foot vertical acceleration and thigh angular velocity as the most critical features, demonstrating a higher consistency with biomechanical priors. The proposed model significantly outperforms traditional SVM, LSTM, and CNN-LSTM baselines, particularly in recognizing transient actions like “Sit” and “Stand”, thereby providing robust technical support for real-time exoskeleton control.

Keywords

Lower-Limb Exoskeleton, Gait Recognition, SE Attention, TCN, Transformer

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着全球老龄化进程的加速及脑卒中等运动功能障碍患者数量的持续增长，下肢外骨骼机器人作为一种能够辅助人体运动及进行康复训练的智能装备，已成为学术界与工业界的关注焦点[1]。在外骨骼机器人的控制系统中，人机交互的自然性与顺应性至关重要，其核心挑战在于机器人能否从复杂的传感器数据中精准理解穿戴者的运动意图，从而实现“按需辅助”(Assist-as-Needed)的主动控制策略[2]。

早期的运动意图识别研究主要依赖支持向量机(Support Vector Machine, SVM) [3]、随机森林[4]等传统机器学习算法。这类方法虽然理论成熟，但依赖专家经验进行繁琐的人工特征工程，且对传感器噪声高度敏感[5]，难以适应不同受试者步态的非线性变异，泛化能力有限。近年来，随着深度学习技术的突破，端到端的特征学习范式逐渐取代了传统方法。其中，卷积神经网络(Convolutional Neural Network, CNN) [6]凭借其强大的空间特征提取能力被广泛应用，但受限于卷积核的局部感受野，CNN在处理长时序步态数据时往往难以捕捉全局依赖[7]。

针对时序建模问题，以长短期记忆网络(Long Short-Term Memory, LSTM)及其变体(如 Bi-LSTM)为代表的循环神经网络(Recurrent Neural Network, RNN)成为了当前步态识别的主流方法。Jeon [8]等人利用 Bi-LSTM 实现了对步态相位的连续估计，验证了其在时序特征提取上的有效性；Fang [9]等人提出的 CNN-LSTM 混合模型试图结合卷积层与循环层的优势。然而，LSTM 固有的串行计算特性不仅导致了较高的

推理延迟, 且在处理超长序列时仍存在梯度消失的风险, 难以满足外骨骼机器人对实时性的严苛要求。相比之下, Bai 等人[10]的研究证明, 时间卷积网络通过因果卷积和空洞卷积, 在序列建模任务中具备比 LSTM 更长的有效记忆长度和并行计算能力, 为解决上述问题提供了新的思路。

尽管 Transformer [11]架构通过自注意力机制打破了序列处理的递归限制, 在自然语言处理领域取得了巨大成功, 且 Song [12]等人提出的 GaitNet 也展示了其在步态特征学习中的潜力, 但直接将原始传感器数据输入 Transformer 仍存在缺陷。由于自注意力机制倾向于计算全局关联, 往往会导致局部高频特征(如脚后跟触地瞬间的冲击信号)被平滑, 从而丢失关键的运动细节。此外, 在多传感器融合场景下, 不同传感器通道对特定动作的贡献度存在显著差异, 缺乏有效的通道筛选机制将引入冗余噪声。

鉴于上述局限性, 本文提出了一种基于 SE-TCN-Transformer 的新型混合网络架构, 旨在兼顾步态识别的高精度与实时性。首先, 引入 TCN 作为特征提取的前端, 并在其残差块中嵌入 SE 注意力机制[13], 通过显式建模通道间的依赖关系重标定特征响应, 使网络能够自适应地学习不同传感器通道的重要性, 有效抑制无关噪声并弥补 Transformer 对局部细节敏感度的不足; 其次, 利用 Transformer 编码器对 TCN 输出的特征序列进行全局建模, 利用其多头自注意力机制精准捕捉步态相位间长距离的时序依赖关系。在公开数据集 HuGaDB [14]上的大量实验结果表明, 本文所提模型在处理复杂步态转换识别任务时, 不仅在准确率上优于现有的 CNN 及 LSTM 变体, 更在推理速度上展现出显著优势, 为下肢外骨骼机器人的实时精准控制提供了新的解决方案。

2. 方法与数据

2.1. 数据预处理

本研究基于 HuGaDB v2 数据集构建步态识别系统, 该数据集涵盖了 18 名受试者在不同地形下的运动数据。在数据预处理阶段, 选取分布于大腿、小腿及足部的 6 个惯性测量单元(Inertial Measurement Unit, IMU)和 2 个表面肌电(surface Electromyography, sEMG)传感器数据, 构建总维度为 $D = 38$ 的多模态特征向量[14]。采集数据的人体传感器位置分布如图 1 所示。为了保证外骨骼控制的实时性并增强训练数据的丰富度, 采用高重叠率的滑动窗口技术对数据进行切片。根据实验设置, 设定窗口大小 $W = 100$ (对应采样点数), 步长 $S = 10$, 即窗口重叠率(Overlap)高达 90%。针对原始数据集中的 12 类活动标签, 本文筛选并整合了其中与下肢外骨骼控制最相关的 8 类主要动作, 包括“行走”、“跑步”、“坐下”、“起立”、“坐姿”、“站姿”、“上楼梯”、“下楼梯”, 并对其进行整数标签映射处理, 以适配模型训练需求。

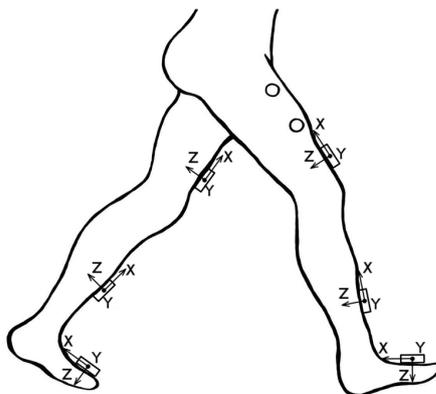


Figure 1. Human body sensor locations for HuGaDB data collection [14]

图 1. HuGaDB 采集数据的人体传感器位置分布[14]

2.2. SE-TCN-Transformer 算法

下肢外骨骼机器人的意图识别要求严格的实时性，为了进一步提升模型对多模态传感器数据的特征筛选能力，我们在 TCN 的残差块中嵌入了 SE 通道注意力机制，构建了 SE-TCN 融合模块。

模型的整体构建流程如图 2 所示。网络输入为包含加速度、角速度及肌电信号等多模态数据的时间序列窗口。

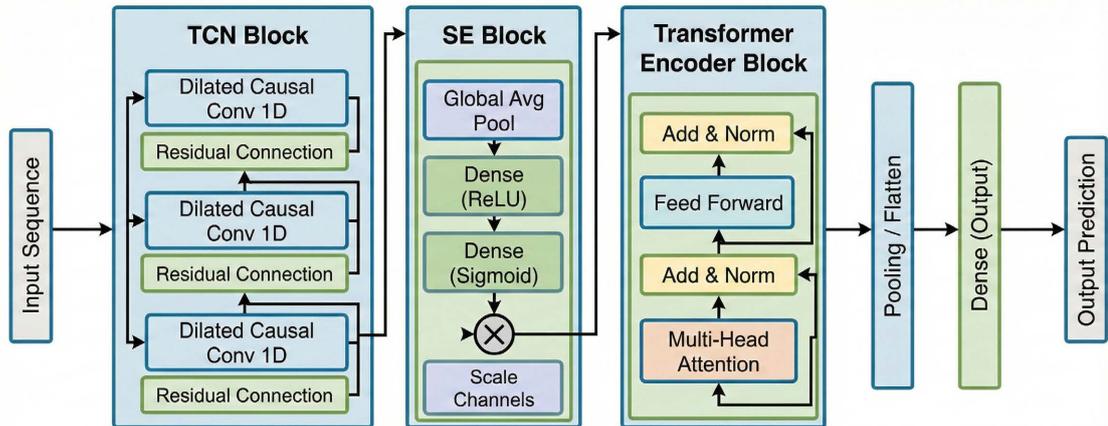


Figure 2. Gait phase prediction based on SE-TCN-Transformer
图 2. SE-TCN-Transformer 步态相位预测

2.2.1. TCN 特征提取层

我们在 TCN 的残差块中嵌入了 SE 通道注意力机制，构建了 SE-TCN 融合模块。本模型采用时间卷积网络作为核心特征提取模块。为了严格遵守时间序列数据的时序约束，模块首先引入了因果卷积(Causal Convolution)结构，确保 t 时刻的卷积输出仅依赖于 t 时刻及其之前的历史输入，从而有效避免了未来信息的泄露。此外，为了在控制计算复杂度的同时捕捉长时序依赖(Long-term Dependency)，模型进一步集成了空洞卷积(Dilated Convolution)机制。

输入数据张量 $X \in \mathbb{R}^{B \times W \times D}$ 首先通过 TCN 模块。该模块包含三层因果空洞卷积层(Causal Dilated Convolution)，膨胀系数 d 依次设定为 1、2 和 4。对于序列中的元素。对于序列中的元素 x_t ，其卷积输出 $F(s)$ 定义为：

$$F(s) = (x *_d k)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (1)$$

其中， K 代表卷积核大小(Kernel Size)， d 为膨胀系数， f 为滤波器权重， $*_d$ 表示空洞卷积运算。该设计使得网络的有效感受野(Receptive Field)随层数呈指数级增长，从而能够高效地捕捉局部高频运动特征与全局长时依赖。

同时，为了增强梯度的传播效率并防止网络退化，模块内部引入了残差连接(Residual Connection)；为提升模型的泛化能力并抑制过拟合，层间还应用了 Dropout 正则化技术。

2.2.2. SE 注意力机制

为了使网络具备特征筛选能力，我们在 TCN 的每个残差块(Residual Block)的两个卷积层之后、残差连接之前，嵌入了 SE 模块。该模块通过显式建模通道间的相互依赖关系，自适应地重标定通道特征响应。具体实现包含压缩、激励与重标定三个步骤。

首先是压缩(Squeeze)操作。由于卷积操作仅在局部感受野内聚合信息，无法获得通道的全局概貌。

我们通过全局平均池化(Global Average Pooling, GAP)将维度为 $C \times T$ (通道数 \times 时间步长)的特征图 U 在时间维度上进行压缩,生成一个包含全局感受野信息的通道描述符 $z \in \mathbb{R}^C$ 。对于第 c 个通道,其计算公式为:

$$z_c = F_{sq}(u_c) = \frac{1}{T} \sum_{t=1}^T u_c(t) \quad (2)$$

这一步将每个通道的时序信号压缩为一个实数值,表征该传感器通道在当前时间窗口内的整体活跃程度。

其次是激励(Excitation)操作。为了捕获通道间的非线性依赖关系,我们采用了一个由两个全连接层(FC)构成的瓶颈结构。首先将特征向量 z 通过第一个 FC 层降维至 C/r ,其中 r 为缩减比(Reduction Ratio),本实验设为 16;经过 ReLU 激活后,再通过第二个 FC 层恢复至 C 维。最后,利用 Sigmoid 函数将输出映射到(0, 1)区间,获得归一化的通道权重向量 s :

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

其中, δ 为 ReLU 函数, σ 为 Sigmoid 函数, $W_1 \in \mathbb{R}^{r \times C}$ 和 $W_2 \in \mathbb{R}^{C \times r}$ 为可学习的权重矩阵。这一过程使得网络能够学习到不同传感器通道之间的协同或抑制关系。

最后是重标定(Scale)操作。将学习到的通道权重 s_c 逐通道地乘到原始特征图 u_c 上,完成特征的重标定:

$$\tilde{u}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

通过这种方式,SE-TCN 能够自动增强那些包含关键步态信息的特征通道,同时抑制与当前任务无关的噪声通道,显著提升模型在复杂动态环境下的鲁棒性。

2.2.3. Transformer

经过 SE-TCN 提取的特征序列虽然包含局部时序细节,但缺乏对长周期步态模式的全局建模能力。因此,我们引入 Transformer 编码器作为时序聚合模块。

多头自注意力机制(Multi-Head Self-Attention)是在自注意力机制的基础上发展起来的,是自注意力机制的变体,可以增强模型的表达能力和泛化能力。在我们的步态分析任务中,不同的注意力头可以自适应地聚焦于不同的运动子空间(Sub-spaces),例如,某些头专注于大腿摆动幅度的周期性变化,而另一些头则捕获足底压力的瞬时突变。这种机制使得模型能够从复杂的传感器数据中,自动对齐具有生理意义的关键帧,从而提升特征的鲁棒性。

该机制允许网络同时关注步态序列中不同的子空间信息,例如同时聚焦大腿的摆动幅度与足底的压力变化。最终,Transformer 的输出经过全局平均池化压缩为特征向量,并通过全连接层映射到各活动类别。模型训练采用加权交叉熵损失函数:

$$L = -\sum_{i=1}^C w_i y_i \log(\hat{y}_i) \quad (5)$$

其中 w_i 为根据类别样本频率计算的权重,用于解决不同步态类别样本不平衡的问题。

3. 实验结果

3.1. 实验设置

为了客观评估模型在未见受试者上的泛化性能,本实验严格采用留一分组交叉验证(Leave-One-Group-Out, LOGO)策略。具体而言,即在 18 轮实验中,每轮依次选取 1 名受试者作为测试集,其余 17

名作为训练集，最终结果取 18 轮的平均值。

训练参数设置为：模型训练采用 AdamW 优化器，初始学习率设定为 0.001，权重衰减系数为 0.001，批量大小为 128，总迭代轮次为 40 次。为了增强模型的鲁棒性并防止过拟合，采取了多项正则化策略：设置丢弃率(Dropout)为 0.3；引入因子为 0.1 的标签平滑(Label Smoothing)机制；采用 Mixup 数据增强技术，混合系数 alpha 设定为 0.4。此外，为了帮助模型跳出局部最优解，应用了重启周期为 10 的余弦退火学习率调度器(Cosine Annealing Warm Restarts)，并将梯度裁剪阈值设定为 1.0 以保证训练过程的稳定性。

为了量化评估模型在步态分类任务上的综合性能，本实验选取了四个核心评价指标：准确率(Accuracy)、精确率(Precision)、召回率(Recall)以及 F1 分数(F1-Score)。其中准确率(Accuracy)直观地反映了模型对所有测试样本分类正确的比例，是衡量模型整体性能的基础指标。精确率(Precision)指的是在模型预测为正类的样本中，实际为正类的比例，它反映了模型的查准能力。召回率(Recall)则是指在所有实际为正类的样本中，被模型正确预测为正类的比例，体现了模型的查全能力。鉴于精确率和召回率往往存在相互制约的关系，本实验还引入了 F1 分数(F1-Score)，它是精确率和召回率的调和平均数，能够更综合、客观地评估模型的稳健性。各指标的计算公式如下：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (8)$$

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

其中，TP (True Positive)、FP (False Positive)、TN (True Negative)和 FN (False Negative)分别代表真阳性、假阳性、真阴性和假阴性的样本数量。在多分类任务中，我们将分别计算每一类步态的上述指标，并取其宏平均(Macro-average)作为最终的评估结果。

3.2. 模型对比实验

为了客观评估所提方法的性能，我们将 SE-TCN-Transformer 与传统的 SVM 模型、经典的 LSTM 以及 CNN-LSTM 模型进行了对比实验。SVM 模型作为浅层学习的基准，输入为窗口内的统计特征(均值、标准差、最大/最小值)。所有模型均在相同的硬件环境与数据集划分下进行测试，评价指标包括平均准确率、精确率、回调率和 F1 分数以及针对不同步态类别的识别鲁棒性。实验统计结果如表 1 所示。

Table 1. Performance comparison of different models

表 1. 模型对比实验结果

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	93.99	91.25	95.02	92.67
LSTM	95.03	92.80	92.46	92.10
CNN-LSTM	94.62	91.77	92.73	91.09
SE-TCN-Transformer	96.73	97.02	95.54	95.43

3.3. 消融实验

为了探究 SE-TCN-Transformer 内部各模块的贡献度，我们设计了逐步回归的消融实验，结果如表 2

所示。

Table 2. Results of ablation experiments
表 2. 消融实验结果

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Transformer	93.55	92.32	95.38	93.47
TCN-Transformer	95.23	94.37	95.32	94.28
SE-TCN-Transformer	96.73	97.02	95.54	95.43

4. 讨论

4.1. 模型对比实验的结果分析

实验结果表明,本文提出的模型平均准确率(Accuracy)达到 96.73%,相比于传统的 SVM 和 LSTM 分别提升了 2.74%和 1.70%。更值得注意的是,在衡量模型综合性能的 F1 分数上,本模型达到了 95.43%,显著优于 CNN-LSTM 的 91.09%,提升幅度达 4.34%。CNN-LSTM 虽然结合了卷积与循环神经网络的优势,但在处理 HuGaDB 这种包含 38 维异构传感器数据及复杂动作转换的数据集时,表现出了一定的不稳定性,导致精确率和召回率的不平衡。相比之下,SE-TCN-Transformer 凭借 SE 模块对多通道特征的自适应筛选以及 Transformer 对全局时序依赖的建模能力,实现了查准率与查全率的均衡提升,证明了该混合架构在处理复杂多模态步态数据时的鲁棒性。

运动意图识别的混淆矩阵如图 3 所示,模型对“坐下”的识别准确率高达 97.34%,对“起立”的识别率亦达到 96.36%。这一表现显著优于通常难以捕捉短时突变的传统 LSTM 模型。这一优势主要归功于模型前端引入的多尺度 TCN 模块。不同于循环神经网络(RNN)存在的“记忆遗忘”问题,TCN 的因果膨胀卷积具备灵活的感受野,能够敏锐捕捉到动作起始与终止瞬间的高频加速度突变特征。同时,SE 注意力模块自动抑制了非相关传感器轴的噪声干扰,使得模型在极短的转换窗口内依然能提取出高置信度的特征向量。这种对意图转换的精准识别,对于保障外骨骼在起坐辅助过程中的安全性与流畅性至关重要。

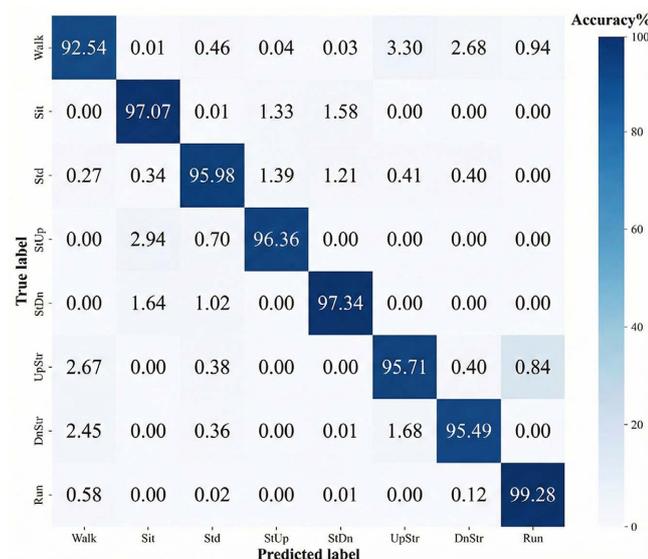


Figure 3. Confusion matrix of the SE-TCN-Transformer model
图 3. SE-TCN-Transformer 混淆矩阵

4.2. 消融实验的结果分析

消融实验以基础 Transformer 模型为基准, 依次引入 TCN 模块和 SE 模块, 通过对比各阶段模型在准确率、精确率、召回率及 F1 分数上的表现(如表 2 所示), 结合混淆矩阵(如图 3~5 所示)分析模型在不同步态类别上的具体的分类性能。

实验结果显示, 基础 Transformer 模型虽然具备处理长序列数据的能力, 但在测试集上的准确率仅为 93.55%, 且在处理局部特征时表现出一定的局限性。通过引入 TCN 模块构建 TCN-Transformer 结构后, 模型的准确率提升至 95.23%, 增长了 1.68 个百分点, 精确率与 F1 分数也分别提升了 2.05% 和 0.81%。这一提升证明了 TCN 模块在提取步态周期局部动态变化方面的优势, 有效弥补了单一 Transformer 在局部感受野上的不足。在此基础上, 进一步融合 SE 注意力机制后的 SE-TCN-Transformer 模型在各项指标上均达到最优, 准确率最终提升至 96.73%, 精确率达到了 97.02%。这表明 SE 模块通过自适应重标定特征通道权重, 成功抑制了噪声干扰并增强了关键特征的表达能力, 使得模型整体性能较基准模型实现了质的飞跃。

为了更细致地分析各模块对特定步态识别的改进作用, 对比图 4 (Transformer)、图 5 (TCN-Transformer)与图 3 (SE-TCN-Transformer)的混淆矩阵可以发现, 模型在易混淆步态和静态动作上的辨识度显著增强。在运动学特征高度相似的“上/下楼梯”与“行走”类别中, 基准 Transformer 模型存在明显的误判现象, 例如将 3.37% 的“下楼梯”样本误判为“行走”。而融合模型通过 TCN 对局部冲击特征的提取以及 SE 对关键通道的加权, 将这一误判率降低至 2.45%, 同时将“下楼梯”类别的整体识别准确率从 93.88% 提升至 95.49%。此外, 在“站立”这一静态动作的识别上, SE-TCN-Transformer 模型有效解决了基准模型因传感器微小漂移而导致的误判问题, 将识别准确率从 93.67% 大幅提升至 95.98%。

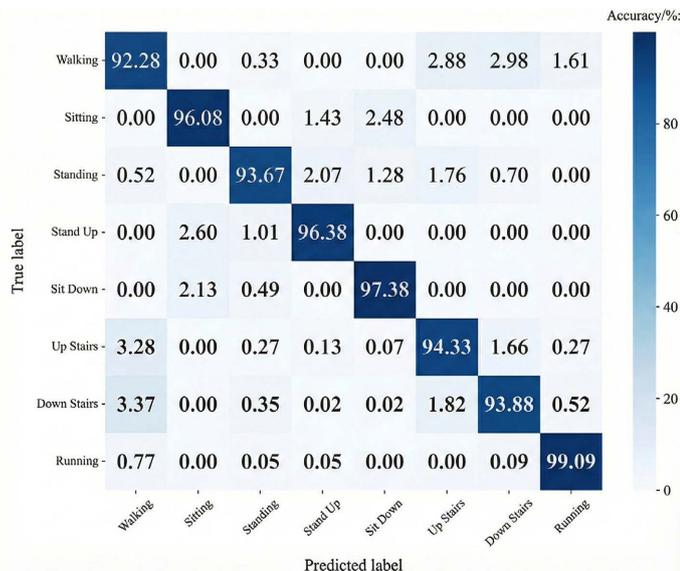


Figure 4. Confusion matrix of the Transformer model

图 4. Transformer 混淆矩阵

综合对比三组实验结果, SE-TCN-Transformer 模型相比于基准 Transformer 模型, 在准确率上累计提升了 3.18%, 在精确率上累计提升了 4.70%。消融实验充分证明, TCN 模块的局部特征提取能力与 SE 模块的特征筛选能力在运动意图识别任务中具有良好的互补性。两者的有机融合使得模型既能捕捉长短时序依赖, 又能聚焦关键特征, 有效解决了复杂步态模式下的精准识别问题。

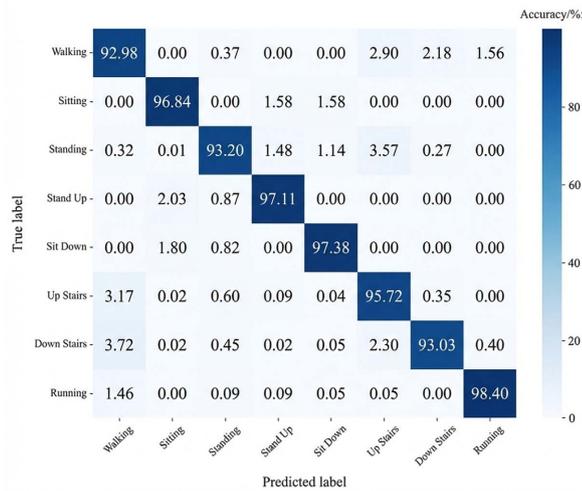


Figure 5. Confusion matrix of the TCN-Transformer model
图 5. TCN-Transformer 混淆矩阵

4.3. 模型可解释性分析

为了进一步验证 SE-TCN-Transformer 模型在步态特征提取上的物理可解释性，探究不同传感器通道对运动意图识别的贡献程度，本节引入 SHAP 方法进行归因分析。SHAP 基于合作博弈论，通过计算每个特征对模型预测结果的边际贡献，量化了各传感器通道的重要性。

4.3.1. 全局特征重要性分析

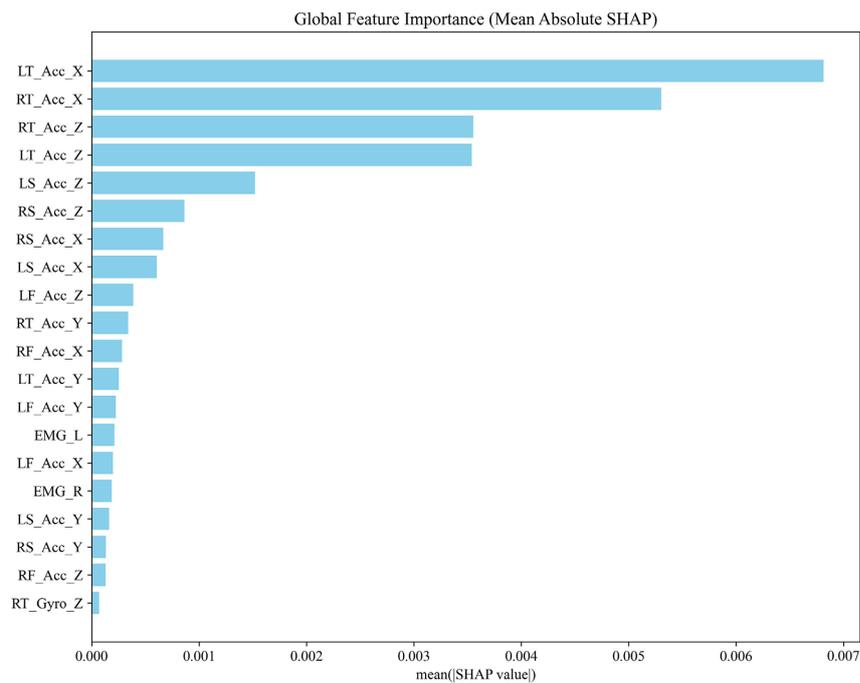


Figure 6. SHAP global feature importance summary
图 6. SHAP 全局特征重要性排序

模型在测试集上的全局特征重要性排序如图 6 所示。从图中可以看出，排名前 10 的关键特征几乎全

部由大腿(Thigh)和足部(Foot)的加速度(Acc)数据占据。其中,左大腿 X 轴加速度与右大腿 X 轴加速度的贡献度最为显著,均值远超其他特征。这种纯数据驱动下得出的特征权重分布,与下肢生物力学规律表现出了极高的吻合度,即大腿作为下肢运动链的驱动源,其在矢状面上的加速度变化确实包含了解析步态周期最核心的动力学信息。

值得注意的是,相比于运动学数据,表面肌电信号的全局重要性排名较低。这一现象侧面印证了本文引入 SE 注意力机制的有效性:由于 EMG 信号在实际采集过程中易受汗液和电极位移噪声干扰,SE 模块通过自适应重标定,自动降低了这些低信噪比通道的权重,使模型更聚焦于鲁棒性更强的惯性传感器特征,从而保证了识别的高精度。

4.3.2. 特定步态类别的决策归因

为了深入探究模型是如何利用特征值的变化来精准区分步态类别的,图 7 显示了针对“行走(Walking)”类别的 SHAP 蜂群图(Beeswarm Plot)。图中横坐标代表 SHAP 值(对预测概率的影响),点的颜色代表特征数值的大小(红色为高值,蓝色为低值)。

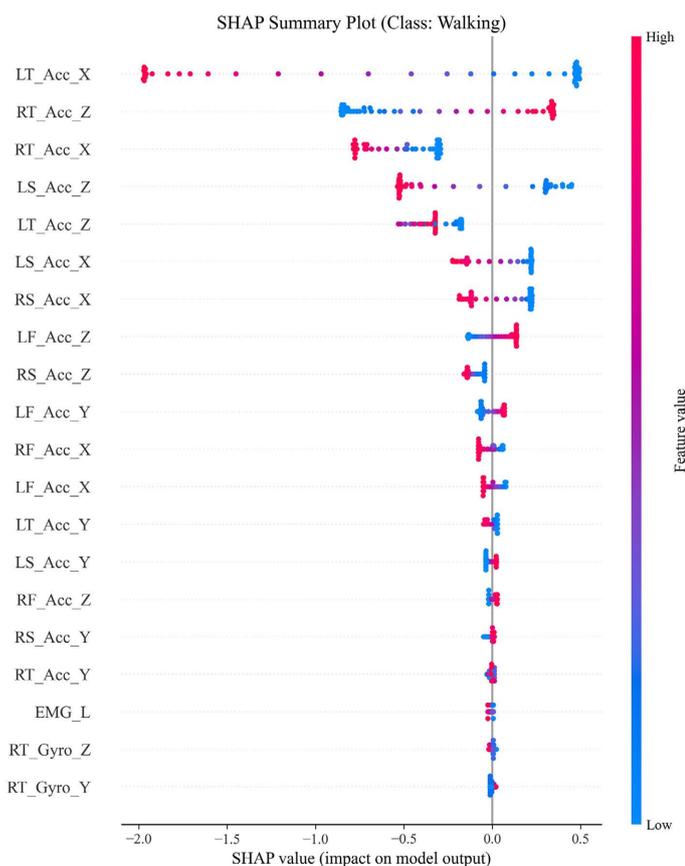


Figure 7. SHAP feature contribution distribution

图 7. SHAP 特征贡献分布图

观察对模型决策影响最大的特征 LT_Acc_X, 其数据点呈现出极具规律的“红-左, 蓝-右”分布。这意味着, 当该特征取高值时, SHAP 值为负, 强烈抑制模型将当前动作判断为“行走”; 反之, 当取低值时, SHAP 值为正, 显著提升“行走”的预测概率。这揭示了模型成功捕捉到了行走与其他高动态动作的本质区别, 行走过程中大腿在 X 轴方向的加速度幅值相对较低且平稳, 而过高的加速度通常对应着更

剧烈的运动模式。与此同时，排名第二的 RT_Acc_Z 呈现出截然相反的“红 - 右，蓝 - 左”分布，即该轴向的高值反而对判断为“行走”起正向贡献。这种对不同肢体、不同轴向特征采取完全不同的加权策略，证明了模型并非简单地依赖“运动幅度大小”进行分类，而是通过 Transformer 结构深刻理解了双下肢交替运动时复杂的时空依赖关系和相位互补特征。

从图 7 的纵向分布来看，特征的影响力呈现出明显的断崖式差异。排名前列的惯性传感器特征分布范围极广，而排名靠后的特征，尤其是部分 Gyro 角速度和 EMG 肌电信号，其数据点紧密聚集在 SHAP = 0 的中心轴线上。这直观地反映了 SE 注意力机制的核心贡献，即模型能够自动识别并抑制那些对“行走”识别贡献度低或含噪量大的传感器通道，而将决策权重高度集中在最具区分力的关键运动学特征上。这种特征筛选机制正是本模型在多模态数据冗余环境下仍能保持高鲁棒性和高准确率的关键所在。

5. 讨论

本文针对下肢外骨骼机器人运动意图识别高精度与实时性要求，提出了一种 SE-TCN-Transformer 融合网络架构。通过集成 TCN 的高频特征提取能力、SE 模块的通道注意力筛选机制以及 Transformer 的全局时序建模优势，该模型在 HuGaDB 数据集上的留一交叉验证中实现了 96.73% 的准确率与 95.43% 的 F1 分数，显著优于 SVM 及 CNN-LSTM 基准。实验结果不仅验证了模型在处理复杂步态转换时的卓越鲁棒性，更通过消融实验揭示了各模块协同工作的物理机理。此外，基于 SHAP 的可解释性分析进一步证实，模型能够自适应聚焦于足部垂直加速度与大腿角速度等高价值运动学特征，并有效抑制低信噪比的肌电干扰。这种基于数据驱动学到的特征重要性分布，与人体下肢生物力学规律表现出较高的一致性，为黑盒模型在医疗辅助设备中的可靠应用提供了良好的物理可解释性基础。针对传统滑动窗口易平滑短时动作的痛点，模型凭借 TCN 对高频突变的捕捉与 SE 的去噪能力，在极具挑战的“坐下”与“起立”瞬态动作上分别实现了 97.34% 和 96.36% 的高识别率，有效解决了动作切换延迟与漏检问题。此外，全局建模能力有效区分了高度相似的行走与楼梯步态。未来工作将聚焦于模型轻量化，以实现嵌入式端侧的实时部署。

参考文献

- [1] Belal, M., Alsheikh, N., Aljarah, A. and Hussain, I. (2024) Deep Learning Approaches for Enhanced Lower-Limb Exoskeleton Control: A Review. *IEEE Access*, **12**, 143883-143907. <https://doi.org/10.1109/access.2024.3414175>
- [2] 李畅, 郑义, 王优强, 等. 下肢康复外骨骼机器人研究现状及应用进展[J]. *机械传动*, 2025, 49(11): 1-16.
- [3] Lau, H., Tong, K. and Zhu, H. (2008) Support Vector Machine for Classification of Walking Conditions Using Miniature Kinematic Sensors. *Medical & Biological Engineering & Computing*, **46**, 563-573. <https://doi.org/10.1007/s11517-008-0327-x>
- [4] Casale, P., Pujol, O. and Radeva, P. (2011) Human Activity Recognition from Accelerometer Data Using a Wearable Device. In: Vitrià, J., Sanches, J.M. and Hernández, M., Eds., *Pattern Recognition and Image Analysis*, Springer, 289-296. https://doi.org/10.1007/978-3-642-21257-4_36
- [5] Figo, D., Diniz, P.C., Ferreira, D.R. and Cardoso, J.M.P. (2010) Preprocessing Techniques for Context Recognition from Accelerometer Data. *Personal and Ubiquitous Computing*, **14**, 645-662. <https://doi.org/10.1007/s00779-010-0293-9>
- [6] Jiang, W. and Yin, Z. (2015) Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. *Proceedings of the 23rd ACM International Conference on Multimedia*, Brisbane, 26-30 October 2015, 1307-1310. <https://doi.org/10.1145/2733373.2806333>
- [7] Ordóñez, F. and Roggen, D. (2016) Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, **16**, Article 115. <https://doi.org/10.3390/s16010115>
- [8] Jeon, S., Choi, J. and Cheong, J. (2020) Gait Phase Estimation from Lower Limb Orientation Using Deep Learning. *IEEE Access*, **8**, 217207-217216.
- [9] Fang, H. and Liang, Y. (2019) Human Activity Recognition Based on Feature Selection and CNN-LSTM. 2019 *IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, 6-9 December 2019, 1997-2001.

-
- [10] Bai, S., Kolter, J.Z. and Koltun, V. (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv: 1803.01271
- [11] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS) 2017*, Long Beach, 4-9 December 2017, 5998-6008.
- [12] Song, C., Huang, Y., Huang, Y., Jia, N. and Wang, L. (2019) GaitNet: An End-To-End Network for Gait Based Human Identification. *Pattern Recognition*, **96**, Article ID: 106988. <https://doi.org/10.1016/j.patcog.2019.106988>
- [13] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-And-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [14] Chereshnev, R. and Kertész-Farkas, A. (2017) HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks. In: van der Aalst, W., *et al.*, Eds., *Analysis of Images, Social Networks and Texts*, Springer, 131-141. https://doi.org/10.1007/978-3-319-73013-4_12