Modeling and Simulation 建模与仿真, 2025, 14(4), 858-870
Published Online April 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.144336

Hans汉斯

TFIE-Gait：一种基于时频信息增强的步态识别模型

蔡　俊

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2025年3月24日；录用日期：2025年4月17日；发布日期：2025年4月27日

摘　要

在步态识别任务中，空间和时序信息对区分不同步态模式至关重要。然而，现有方法在开放环境数据集(如Grew)中主要依赖空间信息，未充分利用时序信息，且开放环境数据集中的噪声(如遮挡和运动暂停)会破坏步态序列时序信息，干扰时序特征提取，降低模型性能。为此，本文提出TFIE-Gait模型，引入时频信息增强模块(TFIE)和去噪采样模块(DAS)。TFIE模块结合时域和频域信息，通过多尺度卷积和自注意力机制提取关节时序特征及关节间依赖关系，并利用傅里叶变换在频域提取判别性特征。DAS模块利用频域去噪前后的序列数据差异联合分析，识别和去除异常数据帧，并利用交叉相关算法拼接子序列，恢复步态序列的周期性时序信息。实验表明，TFIE-Gait在开放环境数据集上显著优于基线模型。

关键词

步态识别，时频分析，时序建模，自注意力机制

TFIE-Gait: A Gait Recognition Model Based on Temporal-Frequency Information Enhancement

Jun Cai

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 24th, 2025; accepted: Apr. 17th, 2025; published: Apr. 27th, 2025

Abstract

In gait recognition tasks, spatial and temporal information are crucial for distinguishing different

文章引用: 蔡俊. TFIE-Gait: 一种基于时频信息增强的步态识别模型[J]. 建模与仿真, 2025, 14(4): 858-870.
DOI: 10.12677/mos.2025.144336

gait patterns. However, existing methods primarily rely on spatial information in open-environment datasets (e.g., GREW) and fail to fully utilize temporal information. Moreover, noise in open-environment datasets (e.g., occlusions and motion pauses) can disrupt the temporal information of gait sequences, interfere with temporal feature extraction, and degrade model performance. To address these issues, this paper proposes the TFIE-Gait model, which introduces a Time-Frequency Information Enhancement (TFIE) module and a Denoising and Sampling (DAS) module. The TFIE module integrates time-domain and frequency-domain information, leveraging multi-scale convolution and self-attention mechanisms to extract joint temporal features and inter-joint dependencies, while utilizing Fourier transform to extract discriminative features in the frequency domain. The DAS module jointly analyzes the differences between sequences before and after frequency-domain denoising to identify and remove abnormal data frames, and employs a cross-correlation algorithm to stitch subsequences, thereby restoring the periodic temporal information of gait sequences. Experimental results demonstrate that TFIE-Gait significantly outperforms baseline models on open-environment datasets.

Keywords

Gait Recognition, Time-Frequency Analysis, Temporal Modeling, Self-Attention Mechanism

1. 引言

步态识别是一种基于人类行走模式进行身份验证的生物特征识别技术。随着人工智能和计算机视觉的快速发展，步态识别因其远距离、无接触和难以伪装的优势，在安全监控、刑事侦查等领域展现出巨大潜力[1]。现有方法主要分为基于外观的方法和基于模型的方法。基于外观的方法依赖于人体轮廓图像序列，而基于模型的方法则利用 2D/3D 骨架[2] [3]或 SMPL 模型[4]等人体结构信息。尽管基于外观的方法在实验室数据集(如 CASIA-B [5]、OU-MVLP [6])上表现优异，但在开放环境数据集(如 Grew [7]、Gait3D [8])中，由于遮挡、停顿等复杂条件，其性能显著下降。这表明仅依赖空间特征难以取得性能的提升，亟需对步态时序信息进行建模作为空间信息的增强，以实现模型性能的提升。

现有方法中，一些研究通过基于集合的方法[9]隐式建模时序信息，认为步态的周期性特征可通过轮廓外观直接识别，从而忽略了相邻帧间的运动特征。然而，显式时序建模已被证明更具优势[10]。3D-CNN [11]被广泛用于从步态轮廓序列中显式地提取时序特征，但其计算复杂度高且由于步态轮廓序列中时序信息的稀疏性难以有效提取时序信息。相比之下，步态骨架序列作为关节点位置的时间序列，包含更丰富的时序信息。因此，本文选择从步态骨架序列中显式地提取时序信息。ST-GCN [12]通过结合图卷积和时间卷积，有效捕捉了骨架序列的时序特征，但其依赖预定义拓扑图，限制了关节点间时序关系的建模。因此本文提出利用多尺度卷积神经网络和自注意力机制分别构建单关节点的局部和全局时序特征，有效地提取时序特征。

此外，现有对步态时序特征提取的研究多集中于时域，忽略了频域信息的重要性。频域分析能有效揭示信号的周期性成分和频率特征，而这些特征在时域中可能并不明显。图 1 和图 2 展示了步态序列样本频域信息中幅度谱的可视化结果，可以看出步态序列的频域信息存在显著差异，表明频域特征具有判别性。受 Lei 等人[13]启发，可将频域特征提取视为图像分类问题，从而有效提取步态的频域特征。

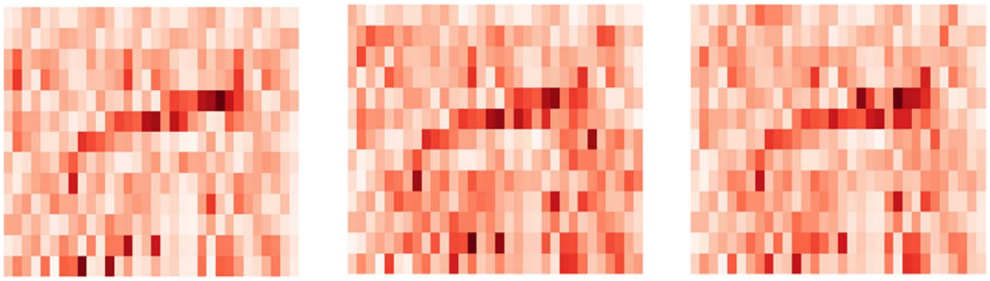

Figure 1. The visualization of the magnitude spectra from different sequences of the same individual
图 1. 同一个体不同序列的幅度谱可视化

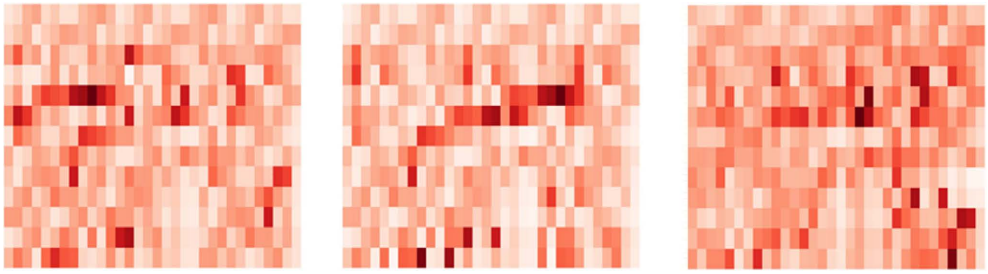

Figure 2. The visualization of the magnitude spectra across different individuals
图 2. 不同个体的步态序列可视化

基于上述观察，本文提出了一种基于时频域信息增强的步态识别模型——TFIE-Gait。该模型包含三个核心模块：空间信息模块、时频域信息增强模块(Time-Frequency Information Enhancement Module, TFIE)和去噪采样模块(Denoising and Sampling Module, DAS)。TFIE 模块通过多尺度卷积和自注意力机制提取时域特征，并通过傅里叶变换提取频域特征。DAS 模块利用频域去噪和交叉相关算法去除异常数据帧，恢复步态序列的周期性时序信息。

综上所述，本文的主要贡献包括以下三个方面：1) 提出了时频信息增强模块(TFIE)，该模块分别从时域和频域的角度构建步态的时序特征。2) 设计了去噪采样模块(DAS)，通过频域去噪和交叉相关算法恢复步态序列的周期性时序特征。3) 通过集成以上部分，提出一种基于时频域信息增强的步态识别模型(TFIE-Gait)。在 Grew 和 Gait3D 两个开放环境步态识别数据集上验证了提出模型的有效性。

2. 方法

2.1. 问题定义

步态识别是一种利用人类行走模式来进行身份验证和识别的生物特征识别技术。定义步态识别问题如下：给定一个包含行走个体的视频序列V，目标是从视频序列中提取特征G并进行匹配，以识别个体的身份。

为了提高去噪效果、特征提取质量和识别的鲁棒性，通常将原始步态视频序列V处理成步态轮廓序列$S \in R^{T*H*W}$和步态骨架序列$J \in R^{T*N*C}$。其中T表示帧数、H表示图像高度、W表示宽度、N表示人体关节数，C表示关节点坐标维度(本文中$C=2$)。

进一步的，在模型训练阶段，给定S和J，那么步态识别问题可以被定义为一个函数$f(\cdot)$。这个函数可以预测S和J对应的可能个体的身份，即$y'=f(S,J:\theta)$，其中θ是一组可学习参数，y'是个体身份标签。设$L(\cdot)$为损失函数，通过最小化$L(\cdot)$的平均值可以最优可学习参数θ，可以表示为：

$$\theta = \arg\min_{\theta} \frac{1}{|D|} \sum_{d_i \in D} L\left(y, f\left(S^{d_i}, J^{d_i} : \theta\right)\right) \tag{1}$$

其中 $|D|$ 表示训练集的大小， S^{d_i}, J^{d_i} 是 V^{d_i} 对应的轮廓图序列和骨架序列。

在推理阶段，给定步态轮廓序列 S 和步态骨架序列 J ，首先通过已训练好的特征提取部分 $f_{feat}(S,J)$ 提取步态特征 G 。然后，使用提取的步态特征 G 进行相似度计算和身份识别，这一过程可以表示为：

$$\hat{y} = \arg\max_{y'} \text{similarity}\left(G, G'_{y'}\right) \tag{2}$$

其中 $G'_{y'}$ 是数据库中存储的已知身份 y' 对应的步态特征序列。

2.2. 模型架构

TFIE-Gait 的架构如图 3 所示。模型使用轮廓和骨架作为输入，其中轮廓作为空间信息模块的输入，骨架作为时频信息增强模块的输入。在数据输入至对应模块之前，需先经 DAS 模块处理。DAS 模块利用噪声在频域中表现为高频的特点，在频域中对骨架序列进行去噪处理，然后将其恢复到时域并与原数据进行对比分析，以确定异常数据帧在序列中的索引集合。根据这个索引集合，原始的步态骨架序列被分割成多个骨架子序列集合。对于模块的输入，空间信息模块采用直接合并子序列并采样的策略而 TFIE 模块则基于序列间的交叉相关性来合并子序列并进行采样。

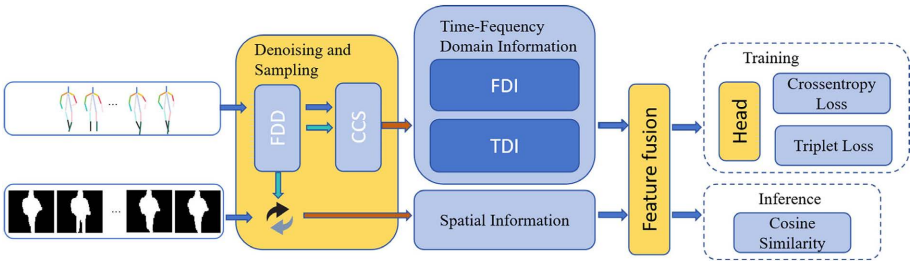

Figure 3. TFIE-Gait architecture
图 3. TFIE-Gait 架构

空间信息模块使用经典基于空间的方法来提取空间特征。TFIE 模块则分为时域信息模块(Time-Domain Information submodule, TDI)和频域信息模块(Frequency-Domain Information submodule, FDI)。在 TDI 模块中，本文提出了局部－全局时序块(Local-Global Temproal Block, LGT Block)来构建单关节点时序特征和关节点之间的时序依赖关系，即步态骨架序列的局部和全局时序特征。在 FDI 模块中，首先将输入转换到频域，得到一个可以看作图像的二维幅度谱矩阵。使用在图像分类任务中表现出色的 ResNet101 [14]模型来提取幅度谱的判别性特征。得到这三种特征后，以空间特征为主导，以时频特征为辅助进行融合，最终输出判别性特征。

2.3. 频域信息模块

之前的工作[12] [15]-[18]都在时域视角对步态时序特征进行建模。然而，值得注意的是，人类的行走模式是周期性的，包含丰富的周期性时序信息。周期性时序信息非常适合转换到频域进行分析。这是因为频域分析能够有效地揭示信号中的周期性成分和频率特征，这些特征在时域中可能并不明显。因此，本文提出了频域模块来提取步态序列中的周期性时序特征。

首先，使用离散傅里叶变换(DFT) [19]将时间序列转换为其频域表示，揭示数据中存在的频率分量。具体来说，对于长度为 T 的骨架序列，其对应的 DFT 频谱可以按照公式(3)计算，如下：

$$F_n^c[k]=\sum_{t=0}^{T-1}J_n^c[t]\mathrm{e}^{-2\pi i\frac{kt}{T}},\ k=0,1,\cdots,T-1 \tag{3}$$

其中 $F_n^c[k]$ 是关节 n 在位置维度 c 上的 DFT 频谱在频率索引 k 处值。$J_n^c[t]$ 是关节 n 在时间步 t 处，位置维度上的原始数据点。T 是帧长度，也就是 DFT 的大小。k 是频率索引，它的范围是从 0 到 $T-1$。i 是虚数单位。经过公式(3)计算后得到步态序列的频域信号 $F\in R^{T*N*C*2}$，其中，最后一个维度的第一个元素表示实部，第二个元素表示虚部。

频率谱中的幅度谱可以揭示数据中的主要振荡频率和能量分布，这对于识别特定模式或状态变化非常重要。此外，幅度谱通常比相位谱更加稳健，更不易受到噪声和其他不确定因素的影响。因此，仅将振幅分量作为输入进行后续处理。因此，对应的频率谱 F 对应的幅度谱为 $M=\sqrt{\left(F[\cdots,0]\right)^2+\left(F[\cdots,1]\right)^2}$，$M\in R^{T*N*C}$。

受 Lei 等人[13]思想的启发，通过对相同个体的不同步态骨架序列和不同个体的步态骨架序列的可视化频谱信息进行观察，可以发现相同个体的步态序列的频谱信息可视化结果相似，而不同个体的频谱信息可视化结果存在显著差异。这一现象表明，Lei 等人提出的将频域信息任务转化为图像分类任务的方法在步态骨架序列数据的频谱信息上仍然有效。因此，本文使用 ResNet101 模型构建频域模块。不同于传统图像分类任务中 ResNet 模型的输入图像通常包含三个通道对应于图像的 RGB 分量，在本文提出的模块中，输入由两个通道组成，分别对应于步态骨架序列数据中关节点的水平和垂直直坐标序列的频谱信息振幅分量。这个过程可以表示为：

$$F^F=\mathrm{ResNet101}(M) \tag{4}$$

其中 F^F 表示频域视角下的步态时序特征。

2.4. 时域信息模块

对于在时域中步态时序信息的提取，GaitGraph [15]和 Gait-D [20]都采用了基于改进的空间-时间图卷积网络(ST-GCN)的方法。然而，ST-GCN 中的图卷积依赖于事先给定的拓扑图来表示关节点之间的关系，这种方式限制了关节点之间时序关系的建模。因此，本文基于卷积神经网络(CNN)和自注意力机制构建了局部全局时序模块(LGT Block)来建模时序特征，如图 4 所示。LGT Block 由一个 CNN 层、一个自注意力层和一个批量归一化层(Batch Normalization, BN)组成。其中，CNN 层用于提取单个关节点的时序特征，通过控制 CNN 层的卷积核长度，可以对关节点在时间维度上的不同序列长度进行时序信息的建模。CNN 层能够构建关节的多尺度时序信息，但由于关节点之间的时序信息也存在依赖关系，因此使用自注意力层来捕获关节之间时序信息的依赖关系。此外，加入了 BN 层对神经网络中间层的输出进行标准化处理，以确保每一层的输入具有相似的分布，从而提高模型的收敛效率。本文通过堆叠 LGT Block 构建了时域模块，用于提取步态序列的时序特征。

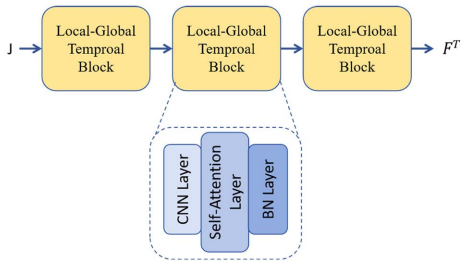

Figure 4. Time domain information module
图 4. 时域信息模块

2.4.1. CNN 层

对于步态骨架序列数据中的每个关节点，本文采用卷积神经网络(CNN)来进行时序建模。与通常的CNN 卷积核为大小为$n\times n$不同，为了适应序列数据，卷积核的大小被设置为$1\times r_n$，其中r_n是卷积核n的长度。通过控制 CNN 的层数和卷积核的长度，可以得到不同的感受野，即对时序数据的不同时间长度的时序特征进行特征提取。不同大小的感受野捕获了不同时间长度的局部特征，有助于在时间序列数据中表示不同的时间模式。

LGT Block 中的每个 CNN 层由三层 CNN 构成。前两层 CNN 卷积核长度的集合都设置为$\{1,2,3,5,7,\cdots,r_n\}$，其中r_n表示第n个内核大小，集合中的数是从 1 到r_n的质数。根据哥德巴赫猜想，任何偶数正整数都可以表示为两个素数的和。因此，随着两个 CNN 层的叠加，模型的感受野可以覆盖从 1 到$2r_n$的所有尺寸。第三层卷积核长度的集合为{1, 2}，作用是对前面的特征图进行特征聚合，以便于输入接下来的自注意力层。

考虑到人体的不同关节的运动模式明显不同，并且步态识别任务属于细粒度的分类任务，权值共享的卷积核难以有效地提取出针对特定关节的样本级特征。因此，在本文中，不同关节点的序列数据使用具有不同权重的 CNN 来提取针对特定关节的样本级特征。总体过程可以表示为：

$$j_{n,l+1}=\text{concat}\left(\text{Conv}^1\left(j_{n,l}\right),\text{Conv}^2\left(j_{n,l}\right),\cdots,\text{Conv}^{r_n}\left(j_{n,l}\right)\right) \tag{5}$$

$$J_{l+1}=\text{concat}\left(j_{1,l+1},j_{2,l+1},\cdots,j_{n,l+1}\right) \tag{6}$$

其中等式(5)表示将关节n的序列使用不同卷积核长度的卷积核卷积，并连接起来。等式(6)表示将所有关节点卷积后的结果连接起来，以形成完整的关节时序特征表示。

2.4.2. 自注意力层

在方法[15] [20]中，使用了基于 ST-GCN 的方法来构建关节点以及关节点之间的时序关系。然而，这种方法依赖于事先给定的拓扑图来确定关节点之间的关系，这种方式一定程度上限制了关节点之间依赖关系建模的自由度。因此，本文提出使用基于自注意力机制的方法来构建关节点以及关节点之间的时序特征依赖关系，如图 5。

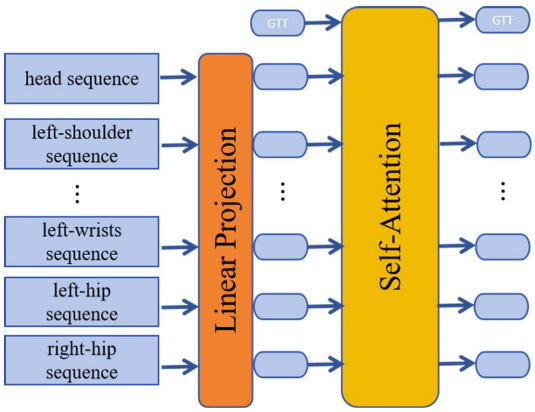

Figure 5. Self-attention layer
图 5. 自注意力层

本文使用多头自注意力机制作为自注意力层的核心。计算过程如公式(7) (8) (9) (10)所示，首先，自注意力层接收 cnn 层的输出$X\in R^{N\times T'\times C'}$作为自注意力模块的输入，并将其映射到指定输入维度$d_{input}$即

$X \in R^{N\times d_{\text{input}}}$。然后，使用 h 个独立的注意力头来计算不同子空间的注意力。对于每个注意力头，分别计算查询、键和值：$Q_i = XW_i^Q, K_i = XW_i^K, V_i = XW_i^V$，其中 $W_i^Q \in R^{d_{\text{input}}\times d_q}$，$W_i^K \in R^{d_{\text{input}}\times d_k}$，$W_i^V \in R^{d_{\text{input}}\times d_k}$ 分别是查询、键、值的权重矩阵。接下来，计算注意力得分矩阵和加权值矩阵并拼接 h 个注意力头的输出，形成一个新的矩阵 $Z \in R^{N\times(h\times d_v)}$。最后，$Z$ 经过全连接层处理后得到输出 O。这里使用了缩放点积注意力机制，$\frac{1}{\sqrt{d_k}}$ 是一个缩放因子，有助于避免在计算点积时数值过大。

$$A_i = \text{softmax}\left(\frac{Q_i K_i^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{7}$$

$$Z_i = A_i V_i \tag{8}$$

$$\mathrm{Z} = \text{ConCat}\left(Z_1, Z_2, \cdots, Z_h\right) \tag{9}$$

$$O = ZW^O \tag{10}$$

为了获取整体的时序特征，在自注意力层增加一个类似于 Vit [21]中 Classification Token 的特殊标记叫做 Global-Temporal Token (GTT)，来捕获步态骨架序列的全局时序信息。并且，基于分层的思想，在时域模块中的每个自注意力层都加入 GTT 以捕获不同层次的全局时序特征。将不同层次的 GTT 输出和最后的自注意力层连接起来，作为最终的具有判别性的时序特征，如图 6。

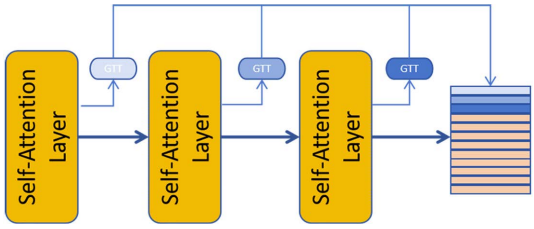

Figure 6. Illustration of GTT
图 6. GTT 功能示意图

2.5. 去噪采样模块

有效去除步态序列中的噪声数据并采样出包含有效时序信息的输入序列是时序模块有效建模步态时序信息的前提。本文提出了 DAS 模块去除噪声并采样。DAS 模块包含一个基于频域去噪的去噪模块(Frequency Domain Denoising Module, FDD)和一个基于交叉相关性的采样模块(Cross-Correlation Sampling Module, CCS)。接下来将详细介绍这两个模块。

2.5.1. 频域去噪模块

人类的步态行走模式存在很强的周期性特征，通常一个步态序列样本中包含多个周期。然而，样本序列中经常存在异常数据帧，如行走停顿、物体遮挡等，这些异常帧会导致步态序列中出现不符合周期性的异常帧。异常帧会破坏步态序列的周期性特征，进而影响模型的建模效果。

本文提出利用一个去噪模块 FDD，利用在频域上噪声通常表现为高频的特点，去除步态序列中的异常数据帧。使用阈值为 f 的低通滤波器滤除掉噪声部分，得到干净的频域特征。再将去噪后的频域信号经过傅里叶逆变换回到时域。这一过程可以表示为：

$$\hat{I}_n^c[t] = \frac{1}{T}\sum_{k=0}^{T-1}\left(\sum_{t=0}^{T-1} I_n^c[t]\mathrm{e}^{-2\pi i\frac{kt}{T}}\right)\cdot H(f)\cdot \mathrm{e}^{2\pi i\frac{kt}{T}}, t = 0,1,\cdots,T-1 \tag{11}$$

其中 $H(f)$ 表示在指定的截止频率 f 之内的频率分量保留，高于 f 的频率分量衰减。

然后再通过欧氏距离来衡量重建序列与原序列数据点之间的差异程度。当同一时间步 t 有超过 k_n 个关节点被判定为异常(数据变化幅度超过 f)，则判定该时间步 t 对应的数据帧为异常数据帧。这一过程可以表示为：

$$d(n,t)=\sqrt{\sum_{c=1}^{C}\left(\hat{I}_n^c[t]-I_n^c[t]\right)^2} \tag{12}$$

$$n_t=\sum_{n=1}^{N}1\left(d(n,t)>j\right) \tag{13}$$

$$T_{\text{anomalous}}=\left\{t\middle|n_k\right\rangle k_n\},\ t=0,1,\cdots,T-1 \tag{14}$$

$$S=\left\{S_k\middle|S_k=\left\{X(:,t,:)\middle|t\in T_k\right\},T_k\cap T_{\text{anomalous}}=\varnothing\right\} \tag{15}$$

在获得异常点集合后，通过移除这些异常数据帧，将原始步态序列分割成一组子序列 S，这一过程见公式(15)。

2.5.2. 交叉相关采样模块

不同个体步态骨架序列样本存在序列长度不一致的问题，因此需要从样本中截取出一段能够有效代表整个序列时序信息的子序列。特定到步态骨架序列数据上，由于人类的行走模式具有明显的周期性，因此需要在样本序列中截取长度为 k 的具有完整的周期性时序信息子序列(k 为超参数)。通过已知的异常数据帧索引集合 $T_{\text{anomalous}}$，可以合理的规避这些异常数据点，截取干净的序列数据。通过设置 k 为合适的值，能够有效的获取到包含有效步态时序信息的定长序列。当 t 的值小于子序列集合中的最长子序列时，直接从子序列中截取。当不存在长度大于 k 的子序列时，最直接的方法是直接拼接最长子序列的相邻子序列得到新的子序列，重复这一操作，直到存在子序列长度大于 k。然而，这种方法虽然简单高效，却必然会存在周期性被破坏的问题，因为拼接的两个序列的结束点和起点在一个步态周期中由于异常点的存在是不对齐的。

为了得到有效的序列周期性时序信息，本文提出了一种基于序列交叉相关的序列拼接方法即 CCS 模块。给定序列 $A=[a_1,a_2,\cdots,a_m]$ 和 $B=[b_1,b_2,\cdots,b_n]$，通过公式(16)计算 A 和 B 在不同位移量 k 下的相关性。其中，b_{i+k} 超出范围时取值为 0，k 表示序列 B 相对于序列 A 的位移量，取值范围是 $[-(n-1),m-1]$。

$$R_{AB}(k)=\sum_{i=1}^{m}a_i b_{i+k} \tag{16}$$

$$k_{max}=\arg\max_k R_{AB}(k) \tag{17}$$

$$\delta=k_{\max}-(n-1) \tag{18}$$

如公式(17)在位移量 $k_{\max}$ 处，序列 A 和序列 B 交叉相似性最大[22]。通过公式(18)将交叉相关性索引 $k_{\max}$ 转换为序列 B 相对于 A 的实际位移量 δ。根据偏移量 δ 来拼接序列 A 和 B，如式(19)。在偏移量 δ 处拼接序列能够最大程度的还原序列的周期性特征，并且重叠部分最少。

$$\mathrm{C}=\begin{cases}\left[a_1,a_2,\cdots,a_m,b_{\delta+1},b_{\delta+2\cdots},b_n\right], & \delta>=0\\ \left[b_1,b_2,\cdots,b_{n+\delta},a_1,a_2,\cdots,a_m\right], & \delta<0\end{cases} \tag{19}$$

综上所述，截取具有完整周期性时序特征的子序列的过程可以表示为：

$$\mathrm{I}(t)=\begin{cases}\mathrm{cutout}_k\left(\max\limits_{s_n}\left(\mathrm{length}\left(s_n\right)\right)\right), \exists s\in S, \mathrm{length}(s)>=k \\ \mathrm{cutout}_k\left(\mathrm{CCS}(S)\right), \nexists s\in S, \mathrm{length}(s)>=k\end{cases} \tag{20}$$

其中，cutout_k 表示从一个序列中截取一段长为 k 的连续子序列。$\mathrm{CCS}(S)$ 表示将 S 中长度最长的子序列和其相邻的子序列拼接成一个新的子序列。当所有子序列都拼接完成但最后的一个子序列长度任然小于 k 时，复制一个相同的子序列继续进行拼接。

3. 实验

为了评估所提出的模型的性能，本文在流行的开放环境步态识别数据集上进行了广泛的验证实验。在本节中，首先介绍了使用的数据集和实现细节。接着，对实验结果进行全面的讨论和分析。最后，进行了广泛的消融研究，以验证本文所提出的方法的有效性。

3.1. 数据集

提出的 TFIE-Gait 在两个开放环境大规模数据集上进行了充分实验，分别是 Grew [7]和 Gait3D [8]。Grew 是步态识别领域中首个大规模的户外步态识别数据集，包含了 26,345 个不同的个体，128,671 个步态序列，提供了自动处理的轮廓、步态能量图(GEIs)、光流以及 2D 和 3D 姿态估计数据。Gait3D 是一个大规模的 3D 表示基础的数据集，它包含 4000 个个体和超过 25,000 个步态序列，提供了从视频帧中恢复的 3D SMPL 模型。同时，也提供了步态序列的轮廓图和 2D 姿态估计数据。与传统的实验室或受控环境下收集的数据集不同，以上两个数据集专注于开放环境，这使得步态识别研究更贴近实际应用场景。

3.2. 数据预处理

考虑到 Grew 和 Gait3D 数据集中 2D 姿态估计数据是相对于原始图像的绝对坐标，包含了行走轨迹、拍摄距离等，原始图片尺寸等与步态无关的信息。本文引入中心归一化处理[17]，即选择两个髋关节中心作为身体的中心点(即重心)，然后计算其他所有关节点相对于这个参考点的偏移量。这样，即使整个人体在图像中的位置改变，关节间的相对位置仍保持不变。引入尺度归一化处理，可以有效避免拍摄距离带来的影响。

3.3. 与基线方法的对比

表 1 展示了本文提出的方法与基线方法在两个户外数据集上的性能对比。所涉及的基线方法包括 GaitSet [9]、GaitPart [16]、GaitGL [11]、GaitBase [23]、DeepGaitV2-2D [10]以及 SwimGait-2D [10]。本文以 DeepGaitV2-2D 和 SwimGait-2D 作为空间信息提取器，结合 TFIE 模块与 DAS 模块进行实验。在两个数据集(Grew 和 Gait3D)上，本文提出的方法相较于基线方法均取得了一定程度的性能提升。

Table 1. The comparison of effectiveness with baseline methods. Ours(*) denotes the results of the baseline model *combined with the TFIE and DAS modules
表 1. 与基线方法的效果对比。Ours(*)表示基线模型*结合了 TFIE 和 DAS 模块的结果

Method	Grew			Gait3D		
	R-1	R-5	R-10	R-1	R-5	mAP
GaitSet	47.4	64.9	71.1	36.9	58.7	30.4
GaitPart	44.1	60.4	68.8	27.8	46.9	21.1
GaitGL	50.8	66.4	72.7	29.5	48.8	22.4

续表

GaitBase	59.7	75.7	80.4	63.8	77.3	52.9
DeepGaitV2-2D	67.7	81.3	86.0	67.5	84.2	60.6
swimGait-2D	69.9	82.5	86.1	68.8	84.1	60.9
Ours (DeepGaitV2-2D)	72.9	84.7	89.2	70.1	85.9	62.1
Ours (SwimGait-2D)	**74.1**	**86.4**	**89.0**	**72.4**	**86.1**	**62.7**

具体而言，首先，在 Grew 数据集上，Ours (SwimGait-2D)的 Rank-1 准确率为 74.1%，Rank-5 准确率为 86.4%，而 Ours (DeepGaitV2-2D)的 Rank-10 准确率达到了 89.2%。在 Gait3D 数据集上，Ours (SwimGait-2D)的 Rank-1 准确率为 72.4%，平均精度(mAP)为 62.7%。与基线方法相比，这些结果均体现了一定程度的性能提升。其次，通过对比 DeepGaitV2-2D 与 Ours (DeepGaitV2-2D)以及 SwimGait-2D 与 Ours (SwimGait-2D)的实验结果可以发现，从骨架中提取时序信息作为空间信息的增强是一种有效提升步态特征判别能力的方法。在两个数据集上，本文提出的方法在不改变空间信息的情况下，均使模型性能得到了一定程度的提升。

3.4. 消融实验

在本小节中，在 Grew 和 Gait3D 数据集上进行了一系列定量和定性消融研究，以分析 TFIEGait 及其组成部分的有效性。

3.4.1. 各个模块的影响

表 2 展示了 TFIE 模块与 DAS 模块在应用于几种经典基线方法时的效果。表中每一列分别列出了基线方法本身、基线方法结合 TFIE 模块、基线方法结合 DAS 模块以及基线方法结合所有模块的性能表现。从表中可以看出，总体而言，本文提出的两个模块的组合有效提升了各基线方法上模型的性能。具体而言，在 Grew 数据集上，GaitGL 的 Rank-1 准确率最高提升了 5.5%；在 Gait3D 数据集上，GaitPart 的 Rank-1 准确率最高提升了 7.5%。

Table 2. The impact of the TFIE module and the DAS module on various baseline models. ALL represents TFIE + DAS
表 2. TFIE 模块和 DAS 模块对各基线模型的影响。ALL 表示 TFIE 与 DAS 的结合

Method	Grew			Gait3D		
	R-1	R-5	R-10	R-1	R-5	mAP
GaitSet	47.4	64.9	71.1	36.9	58.7	30.4
GaitSet + TFIE	48.1↑	65.4↑	72.5↑	37.7↑	59.5↑	31.6↑
GaitSet + DAS	48.6↑	66.8↑	73.2↑	37.9↑	60.6↑	31.1↑
GaitSet + ALL	**51.2**	**68.9**	**74.8**	**40.1**	**62.5**	**32.2**
GaitPart	44.1	60.4	68.8	27.8	46.9	21.1
GaitPart + TFIE	43.6↓	60.1↓	67.4↓	29.2↑	45.5↓	23.4↑
GaitPart + DAS	45.3↑	61.4↑	68.9↑	31.7↑	49.0↑	24.9↑
GaitPart + ALL	46.6	65.6	71.2	35.3	52	30.4
GaitGL	50.8	66.4	72.7	29.5	48.8	22.4
GaitGL + TFIE	52.2↑	65.5↑	74.1↑	32.3↑	50.2↑	24.2↑

续表

GaitGL + DAS	53.7↑	67.2↑	75.6↑	32.9↑	52.9↑	26.5↑
GaitGL + ALL	**56.3**	**68.9**	**77.3**	**36.3**	**55.2**	**31.2**
GaitBase	59.7	75.7	80.4	63.8	77.3	52.9
GaitBase + TFIE	61.8↑	76.6↑	81.8↑	65.0↑	79.2↑	55.9↑
GaitBase + DAS	60.7↑	75.9↑	81.4↑	64.3↑	76.9↑	54.0↑
GaitBase + ALL	**64.7**	**78.2**	**84.5**	**67**	**81.2**	**57.3**
SwimGait-2D	69.9	82.5	86.1	68.8	84.1	60.9
SwimGait-2D + TFIE	71.4↑	84.2↑	88.2↑	70.1↑	85.9↑	61.9↑
SwimGait-2D + DAS	71.1↑	84.7↑	87.9↑	70.8↑	85.1↑	62.5↑
SwimGait-2D + ALL	**74.1**	**86.4**	**89.0**	**72.4**	**86.1**	**62.7**

从单个模块的效果来看，总体而言，TFIE 模块在基线模型中可以单独发挥作用，表明其能够有效提取时序信息作为空间信息的补充，从而提升模型性能。DAS 模块同样能够有效提升各基线模型的性能。然而，通过观察 DAS 模块在不同基线模型上带来的提升幅度可以发现，DAS 模块对 GaitSet、GaitPart 和 GaitGL 的模型性能提升显著，而对 GaitBase 和 SwimGait 的提升效果相对较弱。造成这一现象的可能原因是，后者的模型相较于前者具有更强的鲁棒性。DAS 模块去除异常数据帧的作用对于鲁棒性较强的模型可能并不关键，但对于鲁棒性较弱的模型则能够发挥重要作用。

此外，通过对比 Baseline + TFIE 相较于 Baseline 的提升幅度与 Baseline + ALL (即 Baseline + DAS + TFIE)相较于 Baseline + DAS 的提升幅度可以发现，TFIE 模块对模型的增强效果在使用 DAS 模块对骨架序列去噪后显著优于直接使用原始数据的情况。这表明 DAS 模块有效去除了数据中的噪声，并在一定程度上恢复了序列中时序信息的周期性特征。

3.4.2. TFIE 模块的进一步分析

针对本文提出的 TFIE 模块，进一步进行了消融实验，结果如表 3 所示。其中，第一行为基线模型的结果，以 SwimGait 作为基线模型。第四行为在基线模型的基础上加入 DAS 模块对序列进行去噪后的结果，前文已证明该操作能够有效消除步态序列中的噪声，并在一定程度上恢复序列的周期性时序特征。第二行与第三行分别为在基线模型的基础上加入频域分支与时域分支的结果，可以将其视为第一组实验。第五行与第六行分别为在第四行去噪采样的基础上加入频域分支与时域分支的结果，可以将其视为第二组实验。

Table 3. The impact of the TDI and the FDI submodules in TFIE module on model performance
表 3. TFIE 模块中 TDI 子模块和 FDI 子模块对模型性能的影响

	Baseline	DAS	FDI	TDI	Grew			Gait3D		
					R-1	R-5	R-10	R-1	R-5	mAP
1	√				69.9	82.5	86.1	68.8	84.1	60.9
2	√		√		68.2↓	80.6↓	85.8↓	69.6↑	83.2↓	61.5↑
3	√			√	70.4↑	82.7↑	87.4↑	70.2↑	84.9↑	61.7↑
4	√	√			71.1	84.7	87.9	69.8	85.1	62.1

续表

5	√	√	√		71.6↑	83.6↑	87.9↑	70.8↑	85.5↑	61.9↑
6	√	√		√	73.4↑	84.9↑	88.7↑	71.5↑	85.9↑	62.2↑
7	√	√	√	√	**74.1**	**86.4**	**89.0**	**72.4**	**86.1**	**62.7**

通过对比这两组实验的结果可以发现，频域分支在没有 DAS 模块去噪的情况下表现较差，在 Gait3D 数据集上仅使 Rank-1 (R1)准确率提升了 0.8%，在 Grew 数据集上甚至使准确率下降了 1.7%。而在加入 DAS 模块后，频域分支在两个数据集上分别使 Rank-1 准确率提升了 1.7%与 2.0%。这表明 FDI 子模块依赖于数据中良好的周期性时序特征。同样地，通过观察这两组数据可以看出，TDI 子模块在没有加入 DAS 模块清除噪声并恢复周期性时序特征的情况下，在两个数据集上分别使 R1 准确率提升了 0.5%与 1.4%。而在加入 DAS 模块后，提升幅度分别达到了 3.5%与 2.7%。这说明 TDI 子模块同样会受到噪声与周期性时序特征的影响。同时，这也证明了本文提出的使用自注意力机制对关节点时序特征的依赖关系进行建模是一种有效的时序特征建模方法。该方法不仅能够捕捉单个关节点的时序特征，还能够考虑关节点时序特征之间的依赖关系以及不同层级关节之间的时序依赖关系。

4. 结论

本文提出了一个时频信息增强模块(TFIE 模块)，该模块从时域和频域两个角度提取步态序列特征。频域角度侧重于利用卷积神经网络从步态序列的频域幅度谱中提取判别性特征；而时域角度则着重于对单个关节点的时序特征以及关节点之间的时序依赖关系进行建模。除此之外，本文还提出了基于频域去噪的方法和基于交叉相关性的采样方法。前者利用噪声在频域中表现为高频的特性，在频域进行去噪并还原到时域，随后通过分析多个关节点数据变化来定位和去除异常数据；后者利用子序列之间的交叉相关性对去噪后的子序列进行合并，从而在一定程度上恢复步态序列的周期性时序特征。实验结果表明，TFIE-Gait 相较于基线模型表现良好，提出的 TFIE 模块和 DAS 模块能够有效提升模型性能。在未来的工作中，我们将进一步从频域角度研究步态特征，以期更有效地提取步态的时序特征，增强模型特征的判别能力，并提升模型的准确率。

基金项目

本研究得到了国家自然科学基金(项目编号：61772342 和 62203304)的资助。

参考文献

[1] Connor, P. and Ross, A. (2018) Biometric Recognition by Gait: A Survey of Modalities and Features. *Computer Vision and Image Understanding*, **167**, 1-27. https://doi.org/10.1016/j.cviu.2018.01.007

[2] Cao, Z., Simon, T., Wei, S. and Sheikh, Y. (2017) Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1302-1310. https://doi.org/10.1109/cvpr.2017.143

[3] Martinez, J., Hossain, R., Romero, J. and Little, J.J. (2017) A Simple Yet Effective Baseline for 3d Human Pose Estimation. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2659-2668. https://doi.org/10.1109/iccv.2017.288

[4] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. and Black, M.J. (2023) SMPL: A Skinned Multi-Person Linear Model. In: Whitton, M.C., Ed., *Seminal Graphics Papers: Pushing the Boundaries*, *Volume* 2, ACM, 851-866. https://doi.org/10.1145/3596711.3596800

[5] Yu, S., Tan, D. and Tan, T. (2006) A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. 18*th International Conference on Pattern Recognition* (*ICPR*'06), Hong Kong SAR, 20-24 August 2006, 441-444.

[6] Takemura, N., Makihara, Y., Muramatsu, D., Echigo, T. and Yagi, Y. (2018) Multi-View Large Population Gait Dataset and Its Performance Evaluation for Cross-View Gait Recognition. *IPSJ Transactions on Computer Vision and Applications*, **10**, Article No. 4. https://doi.org/10.1186/s41074-018-0039-6

[7] Zhu, Z., Guo, X., Yang, T., *et al*. (2021) Gait Recognition in the Wild: A Benchmark. *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2021, Montreal, 11-17 October 2021, 14789-14799.

[8] Zheng, J., Liu, X., Liu, W., He, L., Yan, C. and Mei, T. (2022) Gait Recognition in the Wild with Dense 3D Representations and a Benchmark. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 20196-20205. https://doi.org/10.1109/cvpr52688.2022.01959

[9] Chao, H., He, Y., Zhang, J. and Feng, J. (2019) GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 8126-8133. https://doi.org/10.1609/aaai.v33i01.33018126

[10] Fan, C., Hou, S., Huang, Y., *et al*. (2023) Exploring Deep Models for Practical Gait Recognition. arXiv: 2303.03301.

[11] Lin, B., Zhang, S. and Yu, X. (2021) Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 14648-14656. https://doi.org/10.1109/iccv48922.2021.01438

[12] Wang, Q., Zhang, K. and Asghar, M.A. (2022) Skeleton-based ST-GCN for Human Action Recognition with Extended Skeleton Graph and Partitioning Strategy. *IEEE Access*, **10**, 41403-41410. https://doi.org/10.1109/access.2022.3164711

[13] Lei, T., Li, J. and Yang, K. (2024) Time and Frequency-Domain Feature Fusion Network for Multivariate Time Series Classification. *Expert Systems with Applications*, **252**, Article ID: 124155. https://doi.org/10.1016/j.eswa.2024.124155

[14] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[15] Teepe, T., Gilg, J., Herzog, F., Hormann, S. and Rigoll, G. (2022) Towards a Deeper Understanding of Skeleton-Based Gait Recognition. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022, 1568-1576. https://doi.org/10.1109/cvprw56347.2022.00163

[16] Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., *et al*. (2020) GaitPart: Temporal Part-Based Model for Gait Recognition. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 14213-14221. https://doi.org/10.1109/cvpr42600.2020.01423

[17] Huang, Z., Xue, D., Shen, X., Tian, X., Li, H., Huang, J., *et al*. (2021) 3D Local Convolutional Neural Networks for Gait Recognition. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 14900-14909. https://doi.org/10.1109/iccv48922.2021.01465

[18] Liao, R., Cao, C., Garcia, E.B., Yu, S. and Huang, Y. (2017) Pose-Based Temporal-Spatial Network (PTSN) for Gait Recognition with Carrying and Clothing Variations. In: Zhou, J., *et al*., Eds., *Biometric Recognition*, Springer, 474-483. https://doi.org/10.1007/978-3-319-69923-3_51

[19] Cooley, J.W. and Tukey, J.W. (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, **19**, 297-301. https://doi.org/10.1090/s0025-5718-1965-0178586-1

[20] Gao, S., Yun, J., Zhao, Y. and Liu, L. (2021) Gait-D: Skeleton-Based Gait Feature Decomposition for Gait Recognition. *IET Computer Vision*, **16**, 111-125. https://doi.org/10.1049/cvi2.12070

[21] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2020) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.

[22] Oppenheim, A.V. (1999) Discrete-Time Signal Processing. Pearson Education India.

[23] Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y. and Yu, S. (2023) OpenGait: Revisiting Gait Recognition toward Better Practicality. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 9707-9716. https://doi.org/10.1109/cvpr52729.2023.00936