

基于 WiFi 信号和 Transformer 深度网络的行为识别方法

许睿锟 孙承烨 赵宏扬 郭子宇

天津工业大学，天津市，300380；

摘要：在物联网的技术背景下，本文设计了一种基于 WiFi 信号和 Transformer 深度网络的行为识别系统。首先通过部署 WIFI 信号采集设备进而获取实时的室内环境中的信号数据。然后利用子载波选择算法对提取到的 CSI 信息进行预处理，进而提出能够反映人体行为特征的信息。接着将提取到的信息作为 Transformer 深度网络的输入，利用 Transformer 网络中的多层自注意力机制和前馈神经网络，实现对特征的融合和行为分类。测试结果表明基于 WIFI 信号和 Transformer 的行为识别方法可以被广泛应用于不同环境。

关键词：人体行为识别；CSI 信息；Transformer 网络；机器学习

DOI: 10.69979/3029-2735.25.07.099

引言

数字化时代下，智能感知技术快速发展，室内人体行为识别作为核心领域，在智能家居、健康监护、安防等领域展现巨大潜力。

目前人体行为识别主要依赖机器视觉^[1]、可穿戴设备^[2]及雷达毫米波技术^[3]，但均存在局限性：视觉方案涉及隐私风险，可穿戴设备佩戴不便且成本较高，雷达技术存在精度局限。

相比传统方法，基于 WiFi 信号的人体行为识别是新兴领域。因 WiFi 广泛部署，设备收发信号时会产生 CSI 信息。CSI 作为物理层细粒度信息，因其稳定性成为研究主流^[4]。

基于 CSI 的人体行为识别方法主要分为机器学习与深度学习。传统机器学习依赖人工设计特征，需专业领域知识且泛化能力受限。深度学习通过 CNN、Transformer 等架构实现特征自动提取，这些神经网络经过深度学习后适用于动态环境决策，并且可以通过一些相对简单的算法来实现调整，这使得该方法尤为适用于 CSI 信息的人体行为识别。^[4]

综上，基于 WIFI 的人体行为识别方法具有重要意义。因此，本文设计了一个基于 CSI 信号的 Transformer 深度网络行为识别方法。

1 行为识别方法概述

1.1 理论基础

1.1.1 CSI 信号物理特性与行为感知机理

信道状态信息作为正交频分复用系统的物理层参数，能够表征无线信号在传播路径上的复数信道响应。其数学形式可定义为^[5]：

$$H(f, t) = \sum_{k=1}^{N_p} \alpha_k(t) e^{-j2\pi f \tau_k(t)} + \epsilon(f, t) \quad (1)$$

其中， N_p 为多径数量， $\alpha_k(t)$ 和 $\tau_k(t)$ 分别表示第 k

条路径在 t 时刻的衰减系数与时延， $\epsilon(f, t)$ 为测量噪声项。CSI 可刻画多径传播并区分频域和时域路径，其多维度参数包含信号强度及时频域信道状态细节，其中相位信息呈现周期性调制特性^[6]。

1.1.2 Hampel 滤波

Hampel 滤波作为处理数据异常值的有效方法，现被广泛应用。其基于滑动窗口机制，核心原理是在每个窗口内利用数据的统计特征识别并处理异常值。

计算时，先确定窗口大小 n 。对于时间序列 x_t ，在以 t 为中心的窗口 $[t - \frac{n-1}{2}, t + \frac{n-1}{2}]$ (n 为奇数) 内操作。先求窗口内数据中位数 M ，接着算出各数据与 M 的绝对偏差 $|x_i - M|$ ，并求出这些绝对偏差的中位数 MAD。对窗口内每个 x_i ，若 $|x_i - M| > k \times \text{MAD}$ ，就判定 x_i 为异常值，将其替换成 M ，反之则保留原值。

1.1.3 子载波优化选择算法

针对 CSI 数据高维度带来的维度灾难问题，提出改进型相关性筛选算法：

首先定义互相关矩阵：

$$C_{ij} = \frac{1}{T} \sum_{t=1}^T \frac{(x_i(t) - \mu_i)(x_j(t) - \mu_j)}{\sigma_i \sigma_j}, \quad i, j \in [1, N_{sc}] \quad (2)$$

其中 $x_i(t)$ 为第 i 个子载波在时间窗 T 内的幅度序列， μ_i 和

σ_i 分别为均值和标准差。

算法工作流程为：

步骤 1：计算初始互相关矩阵 $C^{(0)}$

步骤 2：计算子载波选择因子 $\rho_k^{(n)} = \sum_{j \neq k} |C_{kj}^{(n)}|$

步骤3: 剔除 $\rho_{\max}^{(n)}$ 对应子载波, 更新矩阵为 $C^{(n+1)}$

步骤4: 重复步骤2-3直至 $\max(C_{ij}^{(n)}) < \theta$ (阈值 $\theta = 0.3$)

1.2 数据处理

由于环境噪声等因素, 原始CSI数据往往包含大量噪声, 因此需要对收集到的CSI行预处理。

本文使用了基于Hampel滤波与子载波选择相结合的CSI数据预处理方法。Hampel滤波通过检测并替换异常值消除噪声突发干扰。子载波选择利用CSI冗余性筛选关键子载波, 在降维同时保留有效信息。

下面是本文对CSI数据的相关处理方法:

1.2.1 Hampel滤波

由于采集到的CSI数据中含有大量异常值, 这些异常值幅值高, 很大程度上影响了Transformer神经网络的输入。为了过滤这些值, 本文对采集到的原始CSI数据进行了Hampel滤波, 处理结果如图1所示。

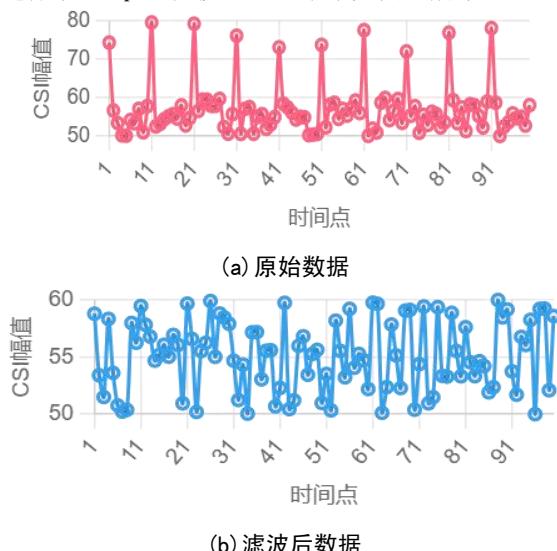


图1 经过Hampel滤波前后的数据对比

1.2.2 子载波选择

由于CSI数据通常由多个子载波的幅值和相位组成, 然而并非所有子载波都对目标重要。因此通过子载波选择方法筛选出最关键的子载波, 以减少冗余信息并提升信号质量。

本文通过构建协方差矩阵, 将多维数据投影到低维空间中, 从而提取出主要特征维度。在此基础上, 通过对子载波的主成分贡献进行量化分析, 可以选取对总信息量贡献较大的子载波, 从而构建优化的低维子载波组合。

通过皮尔逊相关系数计算比较子载波之间的相关性, 量化每个子载波与其他子载波的相关程度。根据相关性矩阵, 可以选取相关性最弱的一组子载波, 最大程

度避免冗余信息。

将Hampel滤波与子载波选择方法整合后, 对收集到的CSI数据进行整合优化, 显著降低了数据的噪声和冗余性, 同时保留了关键信息, 随后将经过预处理过的CSI数据转换成二维热图形式进行表示, 为后续分析提供输入。

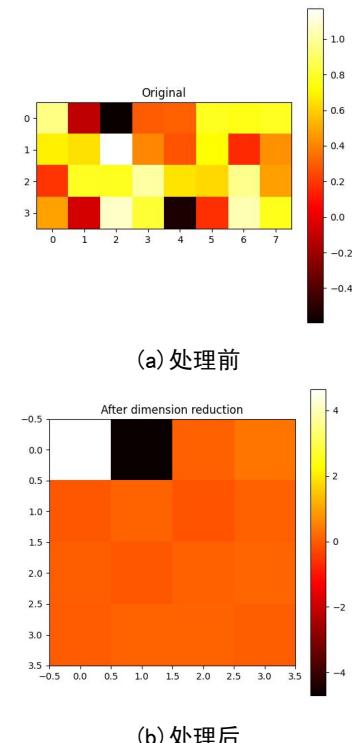


图2 处理前后的CSI热图

1.3 模型设定及训练优化

1.3.1 模型设定

Transformer是一种由多个编码器-解码器堆叠而成的模型, 于2017年由Google团队提出。由于其高度并行化并可以捕获长时间依赖关系, 使其能更好的解决长时依赖问题并在机器翻译上获得了很大成功。如图3所示, 每个编码器都包含两个主要的子层: 多头自注意力和基于位置的前馈神经网络。

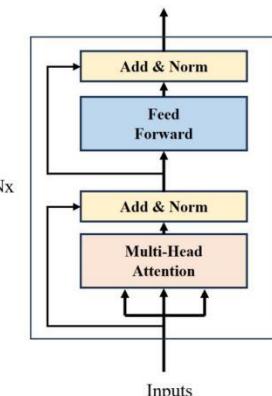


图3 Transformer网络基本框架

多头自注意力机制可以并行地处理并关注时序数据的不同特征，每个“头”都能学习输入序列的不同表示子空间，捕捉到不同类型的依赖关系。除了这两个子层外，每个编码器还包含残差链接和层归一化两个组成部分，它们共同作用于模型的各个层中，提高模型的训练效率和性能。^[7]

本文在输入层将 4×4 热度图展开成 16 维向量，作为网络输入。

模型中多头注意力机制的公式如下：

$$\text{MultiHead(Q, K, V)} = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^o \quad (4)$$

其中，每个注意力头的计算为：

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

本文此次采用 4 个注意力头 ($h=4$)，隐藏层维度为 64。

在做完多头注意力机制的设定后，转向前馈网络(FFN)，前馈网络的公式如下：

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (6)$$

本文的 Transformer 网络此次包含了两个线性层 ($W_1 \in \mathbb{R}^{16 \times 64}, W_2 \in \mathbb{R}^{64 \times 16}$)，并且还使用了 ReLU 函数使得网络具备拟合非线性数据的能力。

对于每层输出，通过残差连接和归一化方式稳定训练过程。

在分类层中，本文通过全连接层将先前展开的 CS I 热力图形成的 16 维向量映射到 5 个动作类别，完成模型的训练，使得模型具有识别对应行为的功能：

$$y_{\text{pred}} = \text{Softmax}(hW_c + b_c), W_c \in \mathbb{R}^{16 \times 5} \quad (7)$$

1.3.2 训练优化

首先定义得分 z ：

$$z = hW_c + b_c \quad (8)$$

其中 W_c 为权重矩阵， b_c 为偏置项。

之后将得分 z 转换为概率分布：

$$\hat{y}_c = \frac{e^{z_c}}{\sum_{j=1}^5 e^{z_j}} \quad (9)$$

通过此方式，便可以更方便地使用交叉熵函数计算模型预测与真实结果之间的差异，具体计算方式如下：

$$\mathcal{L} = - \sum_{c=1}^5 y_c \log(\hat{y}_c) \quad (10)$$

在损失计算完成后，通过链式法则计算损失函数对模型参数的梯度：

$$\frac{\partial \mathcal{L}}{\partial \theta} \quad (11)$$

这里的梯度包括两点：Transformer 编码器的参数

梯度以及全连接层的参数梯度。

在梯度计算完成后，使用 Adam 优化器更新模型参数，其中 Adam 优化器的相关公式如下：

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (12)$$

本文对每个小批量数据进行了如上操作，以保证数据集被完整遍历。并且采用多轮训练的方式，直至模型收敛，大大提升了模型性能。

2 实验结果与分析

2.1 实验设置

2.1.1 数据采集

实验招募 10 名受试者完成 5 类动作，每个动作重复十次，采用 CSI 工具采集 2.4 GHz 频段、无线信道 6、20 MHz 带宽、索引 8 调制编码方案下的 WiFi 信号。系统配置 2×3 MIMO，AP 配备双发射天线，NIC 配备三接收天线，形成 6 个发射 - 接收天线组合。每流通过 OFDM 调制捕获 30 个子载波，在 20 MHz 信道带宽内均匀分布，共采集 $6 \times 30 = 180$ 个子载波数据

2.1.2 评价指标

本文分类结果的评判指标将通过准确率、精确率、召回率、F1 分数四个指标进行评判。四者的关系如下式：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

上式中，TP 为实际正例且被正确预测的样本数，FP 是实际负例但被误判为正例的样本数，TN 为实际负例且被正确预测的样本数，FN 是实际正例但被误判为负例的样本数。同时，本文还采用归一化混淆矩阵来评估模型的性能。

2.2 实验结果分析

2.2.1 训练结果

在完成 1.4.2 节的相关参数设定后，本文对模型进行了训练。

从图 4 中可以看出随着训练轮次的增加，模型训练的损失在逐渐降低，这表明了模型在不断的学习中收敛，并且在最终趋于稳定。说明本次训练的模型训练较好，没有明显的过拟合现象。

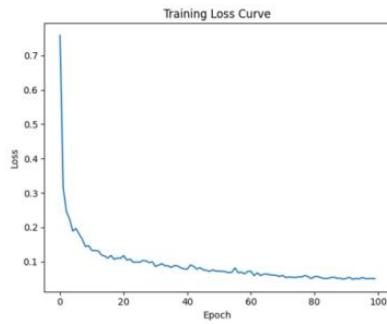


图 4 模型训练损失曲线

从图 5 可以看出随着训练轮次增加，准确率逐渐上升并趋于稳定，最终达到 96% 以上，说明模型的泛化能力较强。

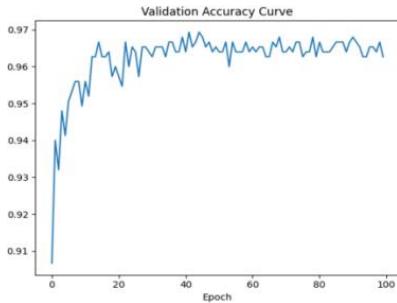


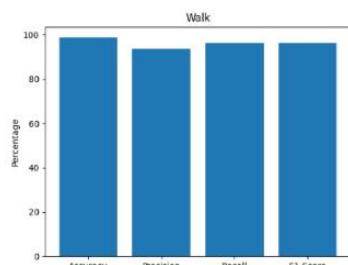
图 5 模型预测准确度曲线

综上，本文所训练的模型的训练结果较好，经过 100 轮次的训练后准确率可以稳定在 96% 以上，表明模型的泛化能力较强，并且无明显过拟合。

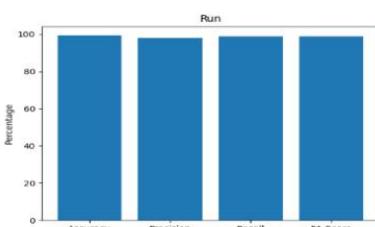
2.2.2 行为识别结果

(1) 嘈杂环境下的识别结果

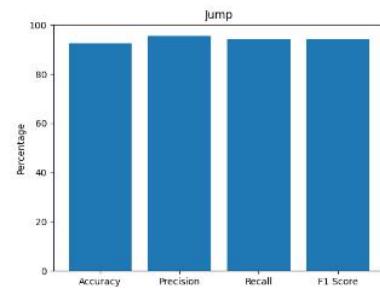
本次实验在嘈杂环境下设置了五个行为，分别是：走、跑、跳、蹲、跌倒。这几个行为的准确率、精确率、召回率、F1 分数值如图 11-图 15 所示。



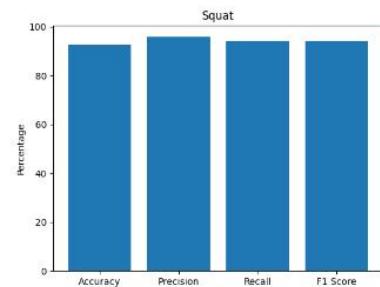
(a) 走



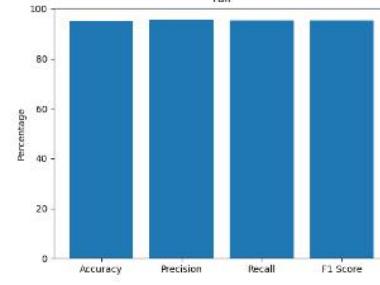
(b) 跑



(c) 跳



(d) 蹲



(e) 跌倒

图 6 嘈杂环境各行为识别结果

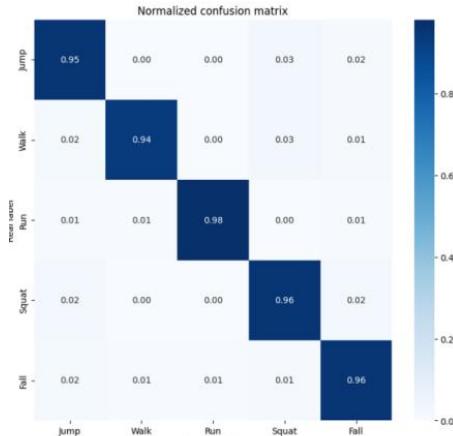


图 7 嘐杂环境归一化混淆矩阵

(2) 安静环境下的识别结果

在安静环境下，本文也对模型的行为识别进行了实验。模型的识别评价指标如下所示：

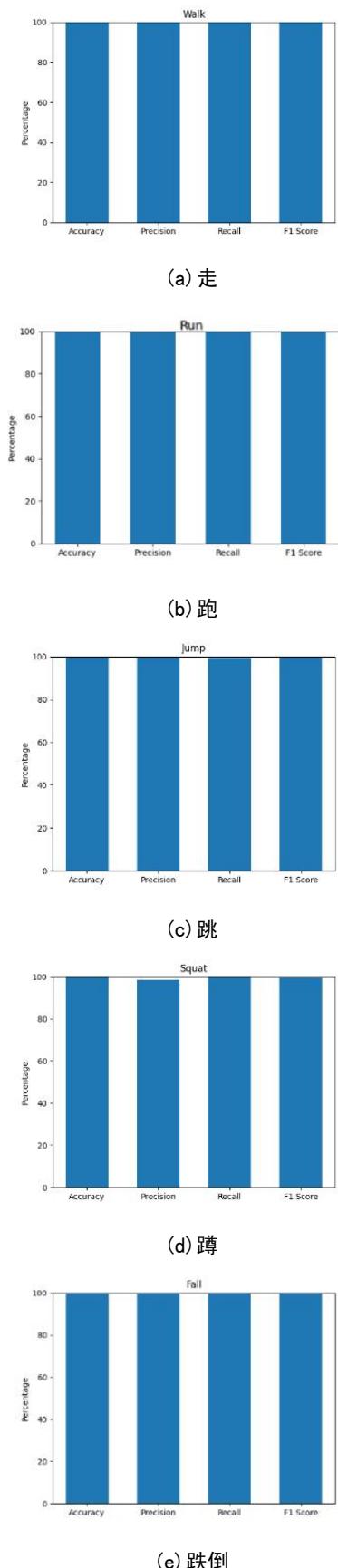


图 8 安静环境下各行为识别结果

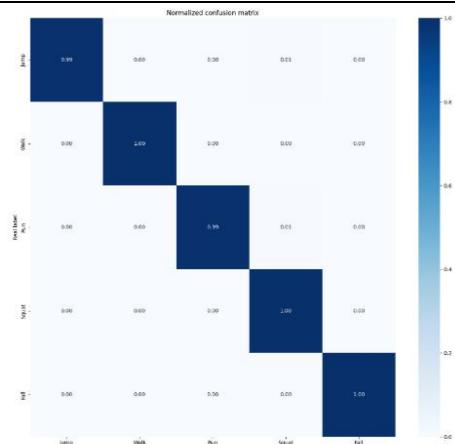


图 9 安静环境归一化混淆矩阵

从总体来看，在嘈杂环境下：该模型的各项指标都在 0.95 以上；在安静环境下：各项指标均在 0.98 以上，并且从归一化混淆矩阵来看，实验时安静环境下的识别准确度几乎为 100%。这说明该模型性能良好，可以有效用于行为识别。

3 结论

本文设计了一种基于 WiFi 信号与 Transformer 深度网络的行为识别方法。通过部署设备采集信号，利用子载波选择算法与 Hampel 滤波预处理 CSI 信息，并转为二维热度图输入 Transformer 网络，借其多层自注意力机制和前馈神经网络实现特征融合与分类。实验显示，模型训练效果佳，100 轮后准确率稳超 96%，无过拟合。训练后的模型在安静和嘈杂环境下的识别结果优秀，可以有效应用体行为识别。

参考文献

- [1] 牟少樊. 基于机器视觉的驾驶员不安全行为识别预警研究[D]. 武汉工程大学, 2025
- [2] 刘子寒. 基于可穿戴设备的人体行为识别方法研究[D]. 中国矿业大学
- [3] 管德蛟. 面向毫米波雷达图像识别人体行为的迁移学习方法研究[D]. 中南大学, 2025.
- [4] 史伟光, 姜皓元. 一种基于 CSI 数据预处理的人体行为识别方法[J]. 天津工业大学学报, 2024, 43 (06).
- [5] 王伟, 李强. 基于 CSI 的无线感知物理模型构建方法. 电子学报, 2018, 46 (3): 521–528.
- [6] 陈志, 等. WiFi 信号多普勒效应与人体运动关联分析. 计算机研究与发展, 2021, 58 (6): 1334–1345.
- [7] 陈栋, 黄国勇. 基于 Transformer 编码器和手工特征的航空发动机剩余寿命预测[J]. 机床与液压, 2025.