

DOI: 10.13382/j.jemi.B2508472

融合分组多尺度注意力的 WiFi 人体活动提取与感知方法^{*}

李 振^{1,2} 赵志彪^{1,2} 石 岩^{1,2} 何育浪^{1,2} 周 琪^{1,2}

(1. 天津职业技术师范大学自动化与电气工程学院 天津 300222;

2. 天津市信息传感与智能控制重点实验室 天津 300222)

摘 要:基于 WiFi 信道状态信息的人体活动感知已成为当前的研究热点。针对现有方法难以提取有效活动数据和依赖人工特征设计的问题,提出了一种融合分组多尺度注意力的 WiFi 人体活动提取与感知方法。首先,采用天线选择和共轭相乘的方法提取对活动感知最敏感的幅值与相位特征,并通过小波降噪进行预处理。然后,通过联合幅值与相位的方差计算滑动窗口因子与活动相关因子,并结合经验阈值提取有效的活动数据序列。最后,构建基于分组多尺度注意力机制的端到端深度学习模型(D-GMAFi)来捕捉序列中的前后关联、多尺度空间信息,并动态关注多个维度上的重要特征信息。通过在自采集的数据集 WiACT 上进行大量仿真实验验证,D-GMAFi 在单人单场景和多人多场景下的平均准确率分别达到了 94.63% 和 91.06%,整体性能优于基准模型,并在不同用户与场景下保持了稳定的识别效果,体现出一定程度的鲁棒性与面向应用场景的适应能力。结果表明该方法在基于 WiFi 人体活动感知任务中具有良好的潜力,为 WiFi 无线感知技术在实际场景中的应用提供了参考。

关键词: 人体活动感知;信道状态信息;多尺度卷积;分组注意力机制;深度学习

中图分类号: TP183;TN83 **文献标识码:** A **国家标准学科分类代码:** 520.20.40

WiFi human activity extraction and sensing approach integrating grouped multi-scale attention

Li Zhen^{1,2} Zhao Zhibiao^{1,2} Shi Yan^{1,2} He Yulang^{1,2} Zhou Qi^{1,2}

(1. School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China;

2. Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin 300222, China)

Abstract: Human activity sensing based on WiFi Channel State Information has emerged as a prominent research focus in recent years. To address the challenges of ineffective activity data extraction and the reliance on handcrafted features in existing methods, this paper proposes a WiFi-based human activity extraction and sensing method that integrates grouped multi-scale attention. First, amplitude and phase features most sensitive to human activity are extracted using antenna selection and conjugate multiplication, followed by wavelet-based denoising for preprocessing. Next, we compute sliding window factors and activity correlation factors through the variance of joint amplitude and phase features, and apply an empirical threshold to extract valid activity sequences. Finally, we design an end-to-end deep learning model (D-GMAFi) based on a grouped multi-scale attention mechanism to capture long-range temporal dependencies, multi-scale spatial information, and dynamically emphasize critical features across different dimensions. Extensive simulation experiments conducted on the self-collected WiACT dataset demonstrate that D-GMAFi achieves average recognition accuracies of 94.63% and 91.06% under single-person single-scene and multi-person multi-scene settings, respectively. The proposed method outperforms the baseline models overall and maintains stable recognition performance across different users and scenarios, indicating a certain degree of robustness and adaptability to application-oriented settings. These results suggest that the proposed approach exhibits strong potential for WiFi-based human activity recognition tasks and can provide a useful reference for the practical application of WiFi sensing technologies in real-world scenarios.

Keywords: human activity sensing; channel state information; multi-scale convolution; grouped attention mechanism; deep learning

收稿日期:2025-06-17 Received Date: 2025-06-17

^{*} 基金项目:天津市自然科学基金(22JCQNJC01100)、天津市教育委员会科研项目基金(2021KJ014)资助

0 引言

随着通信与感知一体化技术的迅速发展,人体活动识别^[1-2] (human activity recognition, HAR) 正为智能家居、身份验证、健康护理等诸多新兴物联网应用提供坚实的技术支撑。基于 WiFi 信号的 HAR 技术已引起学术界广泛关注^[3]。通过分析信道状态信息 (channel state information, CSI) 的原始幅值和相位信息,可以实现非侵入式的人体活动感知^[4]。然而,当前基于 WiFi CSI 的 HAR 仍存在诸多问题,比如有效活动数据的提取及系统鲁棒性不足和泛化能力有限,这些因素都会显著影响模型的识别性能^[5]。

采集到的原始数据往往包含大量噪声和冗余信息,这些噪声和冗余信息会掩盖真实的活动变化,导致原始幅值和相位信息不能直接使用。因此,需要从原始数据中提取出由人体活动引起的有效活动数据区间。Wu 等^[6]提出 TW-See 系统,利用 or-PCA (opposite robust PCA) 提取人体活动与其引起的 CSI 变化之间的相关性,提出了一个归一化方差滑动窗口算法来检测人体活动的起始点。Wang 等^[7]提出了 WiFall 系统,利用局部离群因子 (local outlier factor, LOF) 算法找出与人体活动相关的变化模式。Wang 等^[8]提出了 WiSDAR 系统,首先通过收集无障碍物且无志愿者动作区域内的数据,计算这段时间内的数据均值。然后,对采集到的志愿者动作的数据进行均值计算,在此基础上将原始数据与计算的均值进行差值。最后,根据设定经验功率阈值和时间阈值来提取有效活动数据。Yan 等^[9]提出了 WiAct 系统,设计了一种 AACA 算法,利用信号方差特征中活动和静止部分的差值,通过自适应阈值分割出活动数据片段。Xiao 等^[10]提出了一个基于深度学习的活动分割与识别框架 DeepSeg,通过训练一个 CNN 状态推理模型来识别活动的起始点。但是,这些方法有的是仅利用预处理后的幅值或相位信息,难以准确分割由人体活动引起的数据片段;有的则是依赖静态环境下的基准数据,增加了时间开销。

WiFi 信号不仅对人体行为敏感,也易受周围环境的影响^[11]。在面对动态环境变化时,基于 WiFi CSI 的 HAR 系统仍存在鲁棒性不足和泛化能力有限的问题。为提升系统对环境变化的鲁棒性,Wang 等^[12]提出了一种实时室内跌倒检测系统 RT-Fall,通过分析 CSI 相位差和跌倒时时间频域中的功率下降模式,以实现准确的跌倒分割和检测。Qian 等^[13]提出了 Widance 系统,从 CSI 的幅值和相位信息中提取多普勒频移信息进行人体活动识别。Zhang 等^[14]人提出了 ImgFi 系统,利用 5 种转换方法将 CSI 幅值信息转换成图像信息,通过设计的 CNN

网络进行人体活动识别。Zhang 等^[15]提出了 Widar3.0,推导出了人体活动与速度之间的关系,提取了人体坐标系下速度谱 (body-coordinate velocity profile, BVP) 的域无关特征,提高了模型的泛化能力。Li 等^[16]提出了 WIHF 系统,从 WiFi 信号中推导出与域无关的运动变化模式,设计了一个双任务深度网络模型来进行跨域手势识别和用户识别。Gu 等^[17]将 CSI 比率的幅值和相位转换成对应图片,通过自注意力模块和关系注意力模块来自适应学习手势特征。这些方法虽然表现出较好的性能,显著提升了其在环境变化下的鲁棒性,但是依赖人工特征设计,降低了系统的可扩展性,且对于 CSI 子载波之间的多尺度空间信息提取不足。

为解决上述问题,本文提出了一种融合分组多尺度注意力的 WiFi 人体活动提取与感知方法。该方法通过基于联合幅值和相位的方差阈值分割算法提取出有效的活动数据序列,然后利用基于分组多尺度注意力机制的端到端深度学习模型来捕捉序列中的前后关联、多尺度空间信息,并动态关注多个维度上的重要特征信息,以实现人体活动的提取与感知。

本文提出了一个基于方差阈值的有效活动数据区间提取算法,利用滑动窗口因子和活动相关因子从原始数据中分离活动片段,进一步提高了数据的信噪比;提出了一个基于分组多尺度注意力机制的端到端深度学习模型 D-GAMFi,该模型结合深度可分离卷积网络和门控循环单元网络,并利用分组多尺度注意力机制来捕捉不同尺度下的空间信息关联和动态关注多个维度上的重要信息,增强特征的表征能力,提高模型的鲁棒性;通过与不同的深度神经网络进行综合的性能比较,在自采数据集 WiACT 上,D-GMAFi 整体性能优于基准模型,并在不同用户与场景下保持了稳定的识别效果,体现出一定程度的鲁棒性与面向应用场景的适应能力。

1 相关理论

无线信号在从发射端到接收端的传播过程中,会受到多径、反射、衰落和干扰等影响。在不失一般性的情况下,将信号传播路径划分为静态路径和动态路径,如图 1 所示。静态路径的信号不受人体影响,动态路径的信号会受到人体和其他障碍物 (如地板、天花板、墙面等) 影响。

信号传播在频域上的表达式如式 (1) 所示。

$$Y(f, t) = X(f, t) \times H(f, t) + N \quad (1)$$

式中: $Y(f, t)$ 为接收信号; $X(f, t)$ 为发射信号; $H(f, t)$ 为信道频率状态响应 (channel frequency response, CFR); N 为环境噪声。考虑到 CSI 信号在从发射端传播到接收

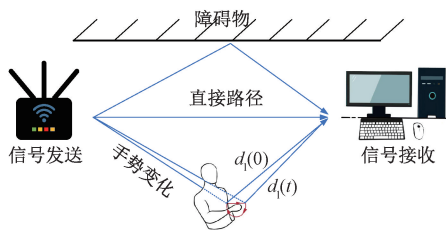


图 1 WiFi 信号传播示意图

Fig. 1 WiFi signal propagation diagram

端时会经过 L 条路径时, $H(f, t)$ 如式(2)所示。

$$H(f, t) = \sum_{l=1}^L \alpha_l(f, t) e^{-j2\pi\tau_l(f, t)} \quad (2)$$

式中: $\alpha_l(f, t)$ 和 $\tau_l(f, t)$ 分别是第 l 条路径的幅值衰减因子和传播延迟。当人处于 WiFi 信号覆盖下时, 人体对信号的反射会使相关路径信号的 $\alpha_l(f, t)$ 和 $\tau_l(f, t)$ 随着人体的运动而变化, 所以 $H(f, t)$ 就包含了人的行为活动信息。

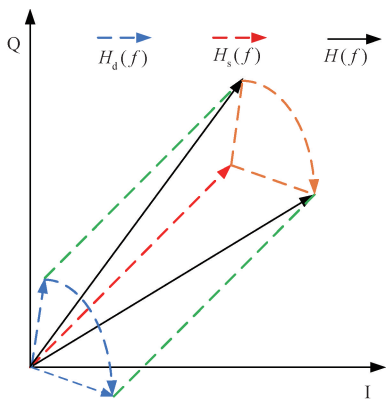


图 2 人体活动引起的多径效应

Fig. 2 Multipath effects induced by human activities

设 $H_s(f, t)$ 和 $H_d(f, t)$ 分别表示为静态频率响应和动态频率响应, 则总的 CFR 如式(3)所示。

$$H(f, t) = H_s(f, t) + H_d(f, t) \quad (3)$$

如图 2 所示, 动态部分是随时间变化的, 在静态路径不变的条件下, 来自感知目标的反射信号越强, 即 $H_d(f, t)$ 越大, CSI 对活动的感知也就越敏感。因此, 可以通过分析 CSI 数据并建立与之对应的活动映射关系, 实现对人体活动感知。

2 系统设计

如图 3 所示, 整个系统框架由数据采集、CSI 信号的预处理和 D-GMAFi 模型 3 部分构成。首先, 利用 WiFi 设备采集 CSI 活动数据, 然后对数据进行特征选择和预

处理, 最后输入联合的 CSI 幅值和相位信息到 D-GMAFi 模型进行训练和活动分类。

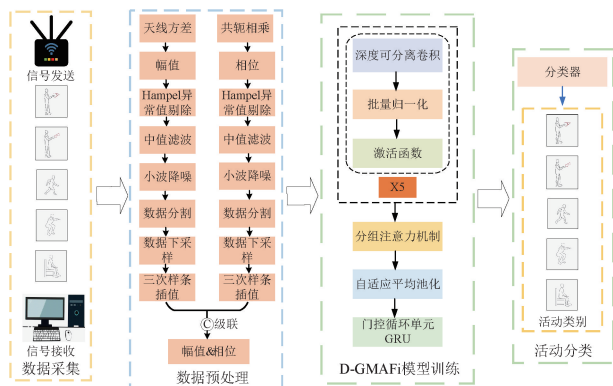


图 3 系统框架

Fig. 3 System framework

2.1 数据采集

现有开源数据集的采集环境相对比较理想, 存在明显局限。部分数据集仅在单一环境下采集, 缺乏环境的多样性; 有些数据集的采集环境部署较为繁琐, 难以在实际中应用推广; 还有些数据集在采集过程中要求志愿执行特定的起始动作或按照预设运动轨迹执行动作, 这些设计均与真实场景有较大差异, 仅利用开源数据集进行研究, 难以全面验证所提方法的鲁棒性与泛化性。为此, 本文采集并构建了名为 WiACT 的数据集。

WiACT 数据集包含 32 类人体活动, 具体信息如表 1 所示。这些活动依据移动范围的特点可划分为手势活动和躯干活动, 其中手势活动包括 26 个英文字母的手势、对号和叉号, 而躯干活动则包括坐下、起立、走路和蹲下。数据采集环境如图 4 所示, 涵盖实验室、办公室和教室 3 个场景。在这些场景中, 发射端和接收端之间的距离分别设定为 1、2 和 2 m。为确保数据的多样性, 共有 12 位身高体重各异的志愿者参与了数据收集。采集设备由两台配备英特尔 5300 网卡 (带有 3 天线) 的戴尔主机终端组成, 分别作为发射端和接收端。每个终端的系统均为 ubuntu18.04.6, 且系统中均安装了 Linux 802.11n CSI Tool 工具。在 1 kHz 的数据采样率下 (即每秒采集 1 000 个数据包), 对每个活动进行为时 4 s 的数据采集。网卡的 WiFi 信号频率设定为 5.32 GHz, 信道带宽设置为 20 MHz, 每对天线可获得 30 个有效子载波。在数据采集过程中, 每位志愿者均根据自己的书写习惯在感知区域内执行特定动作, 不对其进行任何约束。同时, 3 个场景布局在整个采集过程中保持不变。此外, 数据采集期间还引入了非志愿者人员在场内自由活动, 作为环境干扰因素, 使得整个数据采集过程更加贴近真实场景。

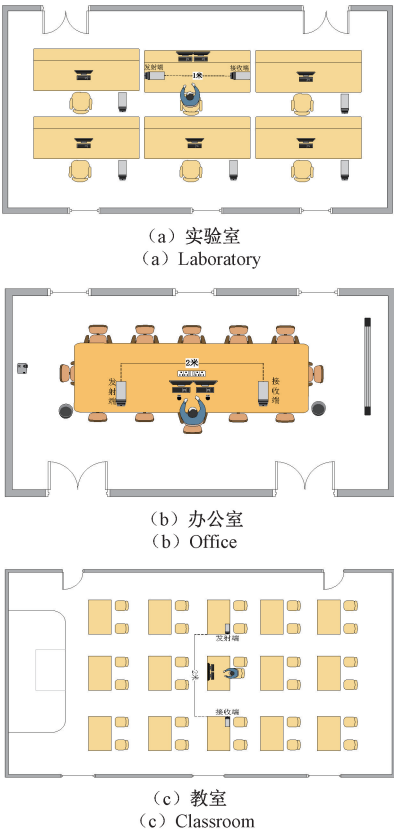


图4 数据采集环境

Fig. 4 Data collection environment

表1 WiACT数据集的详细信息
Table 1 Details of the WiACT dataset

数据集	WiACT
采集工具	Intel 5300 NIC
活动类别	32
活动名称	画 A-Z,画对号和叉号,坐下,起立,走路,蹲下
数据格式	(3,30,300)(天线,子载波,数据包)
训练/测试样本	14 218/3 555

2.2 数据预处理

人体活动数据的采集与处理是 WiFi 感知系统的基础,其数据质量直接影响感知模型的性能边界与应用场景的实际可行性。针对活动识别场景中 CSI 数据固有的多径干扰、环境敏感性和时序依赖等特点,本文提出了一种规范有效的数据预处理方法。该方法涵盖多个关键步骤,包括特征选择、Hampe 异常值剔除、中值滤波、小波降噪、有效活动区间提取、数据下采样以及基于三次样条插值的后续处理。该流程旨在提供高质量、鲁棒性更强的 CSI 数据。

1) 特征选择

WiFi CSI 的幅值信息和相位信息是反映活动变化的两个关键指标。当志愿者在感知区域内执行动作时,其

行为会改变信号的传播路径,从而引起幅值和相位的变化。由于个体朝向改变或者环境干扰,不同天线对人体活动的响应程度会有所不同。因此,本文通过天线信号的方差来量化不同天线对活动感知的敏感程度;信号的方差越大,对活动的感知越强烈。通过计算 3 根接收天线的 CSI 幅值方差,基于天线选择法^[13]选取方差最大的天线作为基准天线。某一样本手势 D 的幅值箱型图如图 5(a)所示,直观地展示了 90 个子载波 CSI 幅值数据的分布情况。从图 5(a)可以看出,红色表示的第 2 根天线的子载波幅值动态响应更为显著,静态分量相对较低,表明其对活动感知具有更高的敏感性。本文选择基准天线的幅值信息作为第 1 个特征。为了获取对活动感知更为敏感的相位信息,将其他两根天线的 CSI 信号与基准天线的 CSI 信号进行共轭乘积运算,分别取其对应的相对相位信息作为第 2 个和第 3 个特征。直接采集的原始相位信息具有周期性,往往会出现相位缠绕现象,难以直接作为特征使用。由于同一网卡上的所有天线都经历相同的相移,通过共轭相乘处理后,可以有效消除随机相位的影响^[13]。志愿者执行手势 D 时的相对相位信息如图 5(b)所示,可以看出相对相位信息 1 和相对信息 2 具有很明显的信号变化,表明其对活动感知的高敏感性。

假设 $X_D \in \mathbb{C}^{N_{ac} \times N_t \times N_{sc} \times N_{pa} \times N_{pa}}$ 为原始数据,则对于上述 3 个特征的数学表达式如式(4)所示。

$$X_{amp} = abs(X_D^{base})$$
$$X_{ang}^i = angle(X_D^i \times (X_D^{base})^*) \quad (4)$$

式中: N_{ac} 代表活动的类别数; N_t 代表发送天线数; N_r 代表接收天线数; N_{sc} 代表发送子载波数; N_{pa} 代表数据包个数; * 表示共轭; $abs(\cdot)$ 表示取幅值运算; $angle(\cdot)$ 表示取相位运算。

这种特征选择方法充分利用了 CSI 信号对活动响应的变化特性,选择出了对活动感知最为敏感的幅值和相对相位信息特征,有利于后续的模式训练。针对环境干扰、硬件设备不稳定等因素造成的个别天线数据幅值缺失,但相位可用的情况,该方法依旧可以利用该天线剩余的有效信息。

2) 小波降噪

原始数据可能出现孤立的离群值,本文采用 Hampel 滤波器来剔除离群值。经过 Hampel 离群值剔除后,进一步采用中值滤波来抑制脉冲噪声^[18]并平滑数据。然而,中值滤波后的原始 CSI 数据中仍包含着由硬件系统固有噪声以及环境微小扰动等因素引起的随机噪声^[19]。传统低通滤波器在滤除噪声的同时,也会模糊 CSI 信号中可能出现的突变特征,从而影响动作感知精度^[20]。因此,本文采用小波变换进行 CSI 降噪,以有效保留信号中的突变和边缘信息。小波变换将 CSI 信号分解为近似系数与细节系数,近似系数保留了信号的整体趋势和平滑

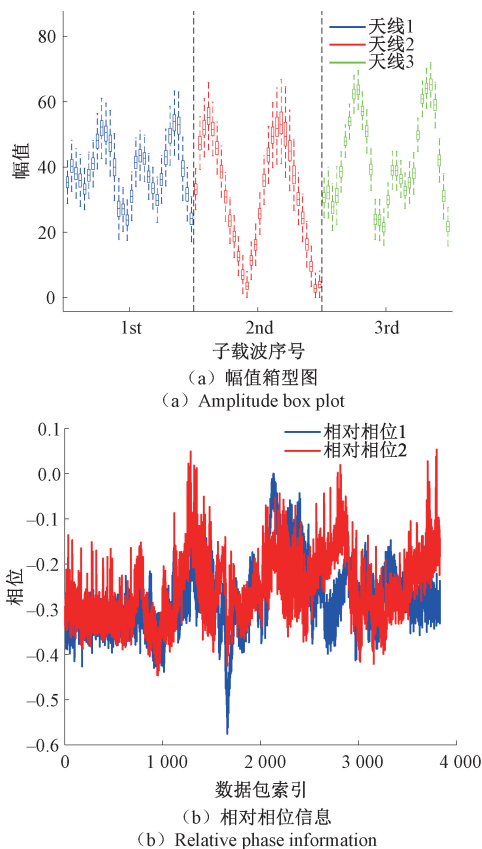


图5 特征选择方法的可视化示例图

Fig. 5 Example visualization of a feature selection method

特性,细节系数则保留了信号的局部细节(比如信号的突变、边缘、噪声等)。再将阈值化的细节系数与近似系数进行重构,恢复出原始信号。数学表达式如式(5)所示。

$$\begin{aligned} X_{amp_dwt} &= \text{wavedec2}(X_{amp_median}, db4, level) \\ X_{ang_dwt}^i &= \text{wavedec2}(X_{ang_median}^i, db4, level) \end{aligned} \quad (5)$$

式中: $\text{wavedec2}(\cdot)$ 是分解函数,用以进行信号分解; X_{amp_dwt} 和 $X_{ang_dwt}^i$ 分别是中值滤波后的幅值和相对相位。由于 CSI 信号具有非平稳性及快速变化的特点,本文使用 $db4$ [21] 小波基函数来捕捉信号中的突变和边缘信息。 $level$ 是分解层级数,这里设定为 5。信号分解完成后,采用软阈值的方法对细节系数进行降噪,其中阈值根据绝对偏差中位数自适应确定。数学表达式如式(6)~(8)所示。

$$\sigma = \frac{\text{median}(|\text{coeff}_{level}|)}{0.6745} \quad (6)$$

$$\text{uthresh} = \sigma \sqrt{2\log(n)} \quad (7)$$

$$\text{de_coeff}[i] = \text{threshold}(\text{coeff}(i), \text{uthresh}, \text{'soft'}) \quad (8)$$

式中: coeff_{level} 代表细节系数; $\text{median}(\cdot)$ 用以计算中位数; 0.6745 是将“中位数绝对偏差”归一化为标准差估计的校正因子; n 是数据中样本点的数量; uthresh 是计

算的阈值。最后利用小波变换的反变换对阈值化后的小波系数进行重构,恢复成原始信号。数学表达式如式(9)所示。

$$\begin{aligned} X_{ap} &= \text{waverec2}(X_{amp_dwt}, db4) \\ X_{ag}^i &= \text{waverec2}(X_{ang_dwt}^i, db4) \end{aligned} \quad (9)$$

其中, $\text{waverec2}(\cdot)$ 是重构函数,用以进行信号重构。重构出来的信号保留了原始数据的各部分细节信息,为后续数据的分析奠定了基础。

以某一样本的相位为例,在经过特征选择后,原始的相对相位如图 6(a) 所示,为人体活动过程中相对相位的变化。原始的相对相位数据包含大量的噪音,通过小波降噪后的相对相位如图 6(b) 所示。从图 6(b) 可以看出,小波降噪后的数据不仅保留了原始信号的基本特征,如边缘和突变信息等,整体变化趋势也变得更加明显。

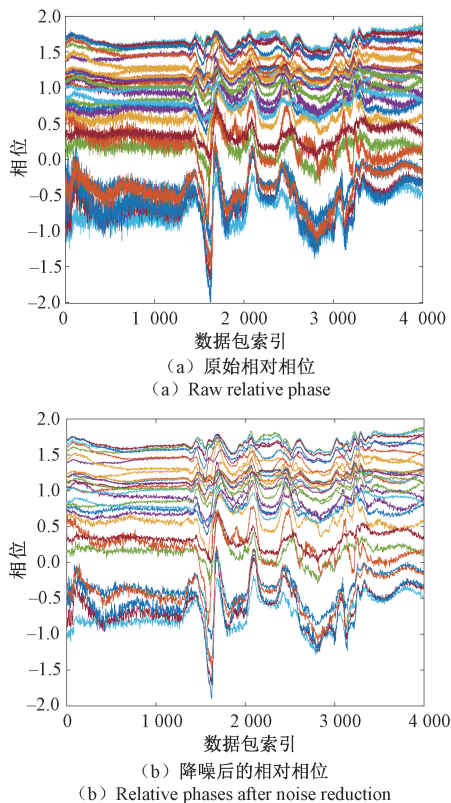


图6 降噪前后的相对相位

Fig. 6 Relative phase before and after noise reduction

3) 有效活动数据区间提取

由于硬件设备延迟、个体动作速度和习惯差异等因素,信号采集时序与动作执行时序无法完全同步,导致采集的原始数据中包含大量冗余信息,进而降低模型性能。如何有效从原始数据中提取由人体活动引起的数据区间一直是个难题。常用方法是将提取的整个数据送入模型学习,或者进行人工手动分割;前者引入大量冗余信息,

这些不相关信息会影响模型识别效果;后者则浪费大量人力资源,且由于个体主观差异,分割出数据质量参差不齐,严重影响实验结果。当志愿者处于感知区域内静止状态时,CSI 信号几乎无波动;当志愿者在感知区域内执行动作时,CSI 信号显著波动,方差明显增大^[22]。许多无线感知工作基于 CSI 信号的这种变化来训练模型,以实现相关感知任务。为解决数据分割问题,本文提出了一种基于联合幅值和相位的方差阈值分割算法,用以提取有效活动数据区间。

人体活动的发生在时序上,往往表现为局部的瞬时变化,而非序列整体统计特性的全局性变化。局部瞬时变化对于整个序列而言可能只是微小的扰动,这类扰动可能会被序列的整体方差所掩盖。因此,本文选择滑动窗口法来捕捉局部变化。该方法计算每个窗口的方差,并通过设定阈值来确定活动的起始点。对于大小为 w ($w < P$) 个数据包的窗口,每个特征的窗口数据表示如式 (10) 所示。

$$H_{w \times M} = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,M} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ H_{w,1} & H_{w,2} & \cdots & H_{w,M} \end{bmatrix} \quad (10)$$

式中: M 是子载波的个数; P 是数据包个数; $H_{i,j}$ 是每个特征对应的第 i 个数据包上的第 j 个数据值。在这个窗口里面计算每一个子载波的标准差,得到一个行向量 $std(w)$ 。

$$std(w) = [std_1 \quad std_2 \quad \cdots \quad std_M] \quad (11)$$

考虑到不同子载波对于人体活动的响应差异性 & 信道干扰可能导致后续起始点分割不准确,进一步对得到的窗口方差 $std(w)$ 进行均值操作。设定窗口变化因子 SWC , 以描述该窗口的总体局部变化。

$$SWC = \frac{1}{M} \sum_{i=1}^M std(w)_i \quad (12)$$

对每个特征的 CSI 序列进行滑动窗口处理,每次滑动步长为设定窗口大小的一半。利用式 (11) 和 (12) 可以获得每个窗口的 SWC 。假设每个特征的总体方差依次可以表示为 $std(X_{ap})$ 、 $std(X_{ag}^1)$ 、 $std(X_{ag}^2)$, 则它们可以用式 (13) 计算。

$$\begin{aligned} std(X_{ap}) &= [SWC_{ap_1} \quad SWC_{ap_2} \quad \cdots \quad SWC_{ap_{W_M}}] \\ std(X_{ag}^1) &= [SWC_{ag1_1} \quad SWC_{ag1_2} \quad \cdots \quad SWC_{ag1_{W_M}}] \\ std(X_{ag}^2) &= [SWC_{ag2_1} \quad SWC_{ag2_2} \quad \cdots \quad SWC_{ag2_{W_M}}] \end{aligned} \quad (13)$$

其中, W_M 是窗口个数,也是向量的位置索引。每一个特征的总体方差均由一个行向量表示,每个窗口计算的 SWC 按照窗口顺序依次放到行向量中。当志愿者在

做动作的时候,经过小波降噪后的幅值和相位会产生相关的波动,本文采用活动相关因子 c ^[23] 来描述联立的 CSI 幅值和相位方差的变化。将 3 个特征的总体方差进行点乘运算获得活动相关因子 c 。

$$c = std(X_{ap}) \cdot std(X_{ag}^1) \cdot std(X_{ag}^2) \quad (14)$$

其中,活动相关因子 c 也是一个行向量,如式 (15) 所示。

$$c = [c_1 \quad c_2 \quad \cdots \quad c_{W_N}] \quad (15)$$

其中, $c_j = SWC_{ap} \cdot SWC_{ag1} \cdot SWC_{ag2}$ 。当志愿者在执行动作前和执行动作结束后的一段时间内,计算所得的 c_j 数值会很小;当志愿者在执行动作的时间内,计算所得的 c_j 数值会显著增大。换句话说,活动前后的 c_j 数值会变化较大,这种变化可以用一阶差分来量化。进一步对活动相关因子 c 进行差分,放大相邻窗口间的差异。由于没有动作时的环境静态特征方差非常小,因此本文设置了一个较小的经验方差阈值 $stresh$ 来确定活动窗口的起始索引。在确定起始索引后,分别与窗口步长相乘,获得有效活动区间的起点和终点。

$$sp = c_{start} \times w/2 \quad (16)$$

$$ep = c_{end} \times w/2$$

其中, c_{start} 和 c_{end} 分别为活动窗口的起始索引, sp 和 ep 分别为有效活动数据的起点和终点。

以本实验场景中某样本的 CSI 幅值特征为例,如图 7 所示。图 7(a) 为从小波降噪后的原始数据计算出的活动相关因子的可视化直方图,红色小圆圈代表窗口起始点索引,绿色小圆圈代表窗口终点索引。从图 7(a) 可见,第 12 个窗口的活动相关因子数值较大,相较第 11 个窗口变化明显,说明第 12 个窗口捕捉到了人体活动的开始;第 26 和第 27 个窗口间的变化也较大,表明第 26 个窗口捕捉到了人体活动的结束。确定窗口起始索引后,利用式 (16) 计算出有效活动数据的起始点,分别对应图 7(b) 中红色和绿色虚线。根据起始点,从原始数据中提取由活动引起的有效数据区间。如图 7(c) 所示,提取的幅值数据片段几乎无冗余信息,表明了所提算法的有效性。

在提取有效活动片段后,先对其进行下采样,再通过三次样条插值对信号进行对齐。在压缩数据量的同时,既保留了信号的主要变化趋势,也实现了不同信号序列的对齐。最后,将处理好的幅值和相位进行级联,为后续模型训练提供维度一致的输入。

2.3 基于分组多尺度注意力的 D-GMAFi 模型

预处理后的原始数据包含与活动相关的多尺度空间信息和时序信息,完整反映了有效活动变化的趋势。为有效学习其中固有的特征模式,本文提出了基于分组多尺度注意力机制的端到端 D-GMAFi 模型,通过聚

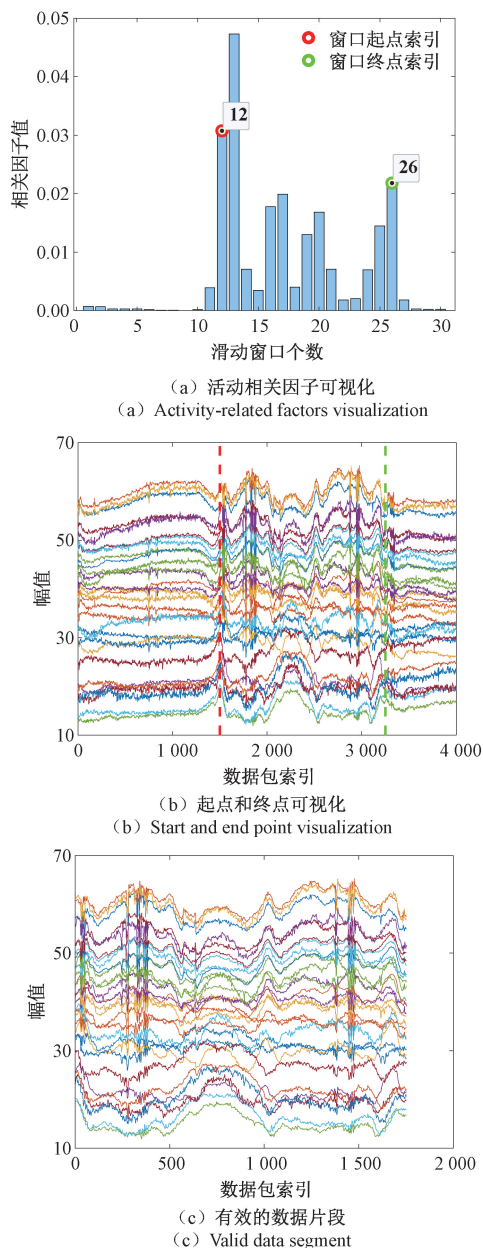


图 7 有效活动数据提取

Fig. 7 Valid activity data extraction

焦不同尺度下的空间信息关联,捕捉不同时序之间的关联,并动态关注多个维度的重要特征信息,从而提升模型的鲁棒性和泛化能力。D-GMAFi 模型由特征提取模块和分类识别模块两部分组成,如图 8 所示。特征提取部分由深度可分离卷积模块(DS-Conv block)、分组多尺度注意力模块(grouping multi-scale attention block, GMAB)和自适应平均池化模块及门控循环单元 GRU 3 个模块组成。深度可分离卷积模块由 5 个独立的深度可分离卷积(DSC)、批量归一化(batch normalization, BN)和 ReLU 激活函数组成。深度可分离

卷积模块用于对输入数据进行权重分配和捕捉人体活动的空间特征,并保持数据的时序特性不变。分组多尺度注意力模块对特征沿着通道维度进行分组,聚焦不同尺度下的空间信息关联,提取出更具有表征能力的抽象特征。自适应平均池化模块用以降低空间特征图维度,进一步平滑特征,增加特征的鲁棒性。门控循环单元 GRU 捕捉输入数据中不同时刻之间的关联,生成具有时空特性的特征。分类识别模块则由丢弃率(dropout)、全连接层(fully connected, FC)和 softmax 函数 3 部分构成,利用提取到的时空特征信息实现 32 类活动类别。其中,Dropout 的丢弃率设置为 0.5,用以缓解模型过拟合;FC 层用来进行类别映射,softmax 函数将 FC 层输出转换成各类别的概率分布,给出概率最大的类别及其对应的索引。

1) 深度可分离卷积

CSI 数据包含丰富的时空信息,相比于传统的 CNN, DSC 能够更有效地提取到数据中的空间信息,实现跨通道信息融合。深度可分离卷积将卷积过程分为深度卷积(depthwise convolution, DC)和点卷积(pointwise convolution, PC)。DSC 首先进行 DC 操作,每个通道对应一个卷积核,每个卷积核仅与对应通道的特征进行卷积,提取空间特征信息,生成新的特征图。卷积核的数量与输入数据的通道数一致,新的特征图通道数与卷积核数保持一致。DC 操作没有建立通道之间的联系,也就是没有充分利用相同空间位置不同通道之间的特征信息,PC 操作解决了上述问题。为了扩展特征图,PC 操作使用了 N 个与上一层通道数一样的 1×1 卷积核,即 $1 \times 1 \times C$, N 为输出通道数,也就是输出特征图的通道数。 $1 \times 1 \times C$ 的卷积核会与上一层特征图进行卷积,然后在通道维度上进行加权组合,生成新的输出特征图,起到了跨通道的信息融合作用。

2) 分组多尺度注意力模块

为进一步捕捉 DSC 输出特征图中通道信息和空间信息的微小差异,并动态关注重要的特征信息,设计了一个分组多尺度注意力机制。GMAB 首先采用分组策略,将输入特征沿通道维度划分为多个互不重叠的子特征图组,这样可以减少计算复杂度。然后对每个子特征组使用不多尺度卷积网络单元进行通道注意和空间注意的局部跨信息融合。最后,在通道维度上拼接各子特征组的局部信息,实现信息交互。

对于输入特征图 $X \in \mathbf{R}^{C \times L}$,其中 C 和 L 分别为通道数和序列长度,GMAB 首先将 X 沿着通道维度分为 g 组,即 $X = [X_1, X_2, \dots, X_g]$, $X_m \in \mathbf{R}^{C/g \times L}$,其中 $m = 1, 2, \dots, g$,每个子特征组 X_m 在训练过程中都会捕捉到当前通道下特定的语义信息。然后通过选择 3 个不同卷积核大小的深度可分离卷积网络来提取对应局部视野范围的特征,

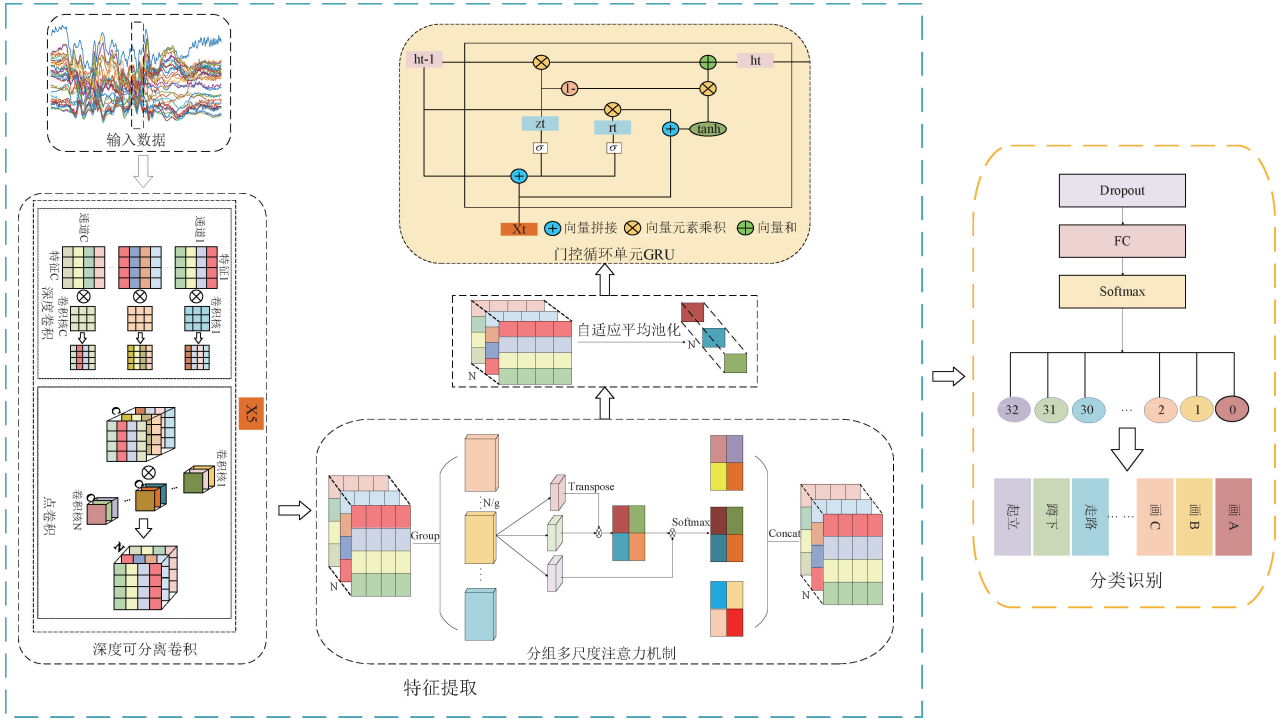


图8 D-GMAFi 模型的总体架构

Fig. 8 Overall architecture of the D-GMAFi model

分别作为自注意力机制中的 Q, K, V , 其中较小的两个卷积核提取的特征对应 Q 和 K 。最后按照式 (17) 来计算每个子特征组的重要性分数。

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (17)$$

式中: d_k 是查询矩阵和键矩阵的维度; $\sqrt{d_k}$ 是缩放因子, 防止梯度爆炸或消失。 $Q \in \mathbb{R}^{C/g \times d_k}$, $K \in \mathbb{R}^{C/g \times d_k}$, $V \in \mathbb{R}^{C/g \times d_k}$ 分别是查询矩阵, 键矩阵和值矩阵。对于 Q, K, V 可由式 (18) ~ (21) 所示。

$$\widetilde{X}_m(i, j) = \sum_{l=0}^{K-1} X_m(i, j+l-p) \times k_d(i, l) \quad (18)$$

$$Q = W_q \times \widetilde{X}_m + b_q, W_q \in \mathbb{R}^{C/g \times C/g} \quad (19)$$

$$K = W_k \times \widetilde{X}_m + b_k, W_k \in \mathbb{R}^{C/g \times C/g} \quad (20)$$

$$V = W_v \times \widetilde{X}_m + b_v, W_v \in \mathbb{R}^{C/g \times C/g} \quad (21)$$

式中: $\widetilde{X}_m \in \mathbb{R}^{C/g \times L}$; k_d 为卷积核; p 为填充量; $i = 1, 2, \dots, C/g$; $j = 1, 2, \dots, L$ 。利用式 (18) 就可以获得每个子特征组的重要性分数 $\widetilde{X}_m' \in \mathbb{R}^{C/g \times L}$, 最后沿着通道维度拼接 $\widetilde{X}_m'$ 得到注意力特征图 $X' = [\widetilde{X}_1', \widetilde{X}_2', \dots, \widetilde{X}_g']$, $X' \in \mathbb{R}^{C \times L}$ 。基于分组多尺度卷积网络单元生成的注意力特征图包含了通道信息和空间信息, 增强了特征的特征能力。

3) 门控循环单元

为进一步利用数据中的时间信息, 采用门控循环单元 GRU 来捕捉序列数据不同时刻之间的关联性。GRU 由重置门和更新门两个门组成, 重置门选择性地遗忘部分信息, 捕捉当前输入的关键特征; 更新门则主要用于在每个时间步中平衡保留历史信息与引入新输入的比例。

如式 (22) 所示, x_t 为 t 时刻的输入, h_t 为 t 时刻的输出或状态, h_{t-1} 为 $t-1$ 时刻的状态, W 为权重, σ , sigmoid , $\tanh$ 为激活函数。利用 GRU 的更新门和重置门来捕捉 CSI 数据的时序特性, 生成具有时空特性的特征图, 作为分类模块的输入。

$$\begin{aligned} z_t &= \sigma(W_{(x,z)}^T x_t + W_{(h,z)}^T h_{t-1}) \\ r_t &= \sigma(W_{(x,r)}^T x_t + W_{(h,r)}^T h_{t-1}) \\ g_t &= \tanh(W_{(x,g)}^T x_t + W_{(h,g)}^T (r_t \otimes h_{t-1})) \\ h_t &= (-z_t \otimes h_{t-1}) + z_t \otimes g_t \end{aligned} \quad (22)$$

3 实验结果与分析

为主要验证模型的有效性和泛化能力, 首先, 在开源数据集 UT-HAR 上, 验证模型的有效性。然后在实验室环境下采集的数据上分别进行消融实验和超参数实验, 验证所提模型中各模块的有效性和确定最优分组数。最后在 3 种环境下采集的数据上评估模型对未见样本的鲁

棒性和泛化能力。

3.1 实验设置

本文实验搭建在 Ubuntu 20.04 平台。该平台配备两张 NVIDIA GeForce RTX 3090 显卡,共 48 G 显存。在 Pytorch 深度学习框架下搭建模型。本文所有模型均采用 Adam 优化器,初始学习率设置为 0.001, β 参数取值为 0.9 和 0.999。同时引入正则化技术以缓解过拟合问题,并采用余弦退火学习率策略对学习率进行动态调整,训练轮次设置为 200。所有数据集均采用分层抽样,将数据按 8:2 的比例划分为训练集和测试集。

3.2 评价指标

本实验采用准确率 (accuracy, ACC)、精确率 (precision, Pr)、F1 分数 (F1-score, F1)、召回率 (recall, Re)、Params 和 FLOPs 来综合评估模型的性能。其中,ACC 用以评估模型的整体预测能力,PR 用以衡量评估模型正类预测的可靠性,F1 分数是召回率和精确率的调和平均值,Re 用以衡量模型对所有真实正例的识别能力,FLOPs 用以评估计算复杂度,Params 用以评估 GPU 内存需求。评级指标的数学表达式如式 (23) ~ (26) 所示。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)$$

$$Pr = \frac{TP}{TP + FP} \quad (24)$$

$$Re = \frac{TP}{TP + FN} \quad (25)$$

$$F1 = 2 \times \frac{Re \times Pr}{Re + Pr} \quad (26)$$

式中: TP 是模型正确预测为正类的正样本; TN 是模型正确预测为负的负样本; FP 是模型错误预测为正类的负样本; FN 是模型错误预测为负类的正样本。

3.3 平衡策略

模型均采用交叉熵损失函数 (cross-entropy loss, CEL),它是一种适用于多类别分类任务的损失函数,并且能够有效地衡量模型预测与真实标签之间的差异。WiACT 数据集中各类别样本的分布情况如图 9 所示,可以看出不同活动的样本数量存在明显的不均衡现象。在模型训练过程中,模型往往更偏向于样本数量较多的类别,从而导致样本较少类别的预测性能不理想。为缓解这一问题,实验中采用加权损失函数的方法来解决,即对每个样本的损失乘以其对应类别的权重。假设第 i 个样本的损失为 L_i ,则整体损失计算如式 (27) 所示。

$$CEL_{weighted} = \frac{1}{n} \sum_i^n w_i \times L_i \quad (27)$$

其中, w_i 是第 i 个样本所属类别的权重。将类别权

重设置为该类别样本数量与总数量比例的倒数。较少出现的类别将会被赋予较高的权重,每个样本的损失乘以其对应类别的权重,从而缓解训练过程中的样本不均衡性。

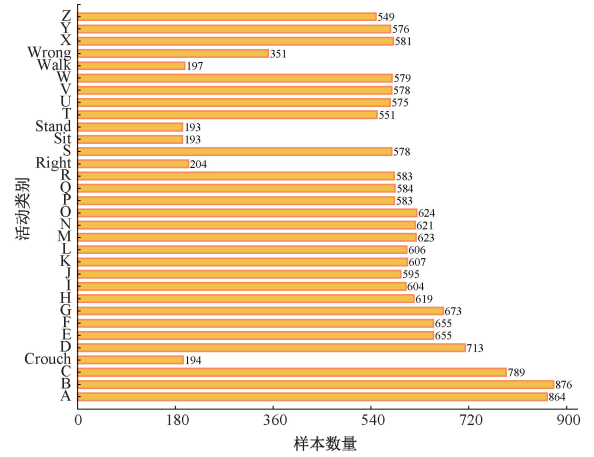


图 9 训练集中各活动样本数量分布

Fig. 9 Distribution of the number of samples for each activity in the training set

3.4 模型的有效性验证

为了验证模型的有效性,将所提模型与基准模型在开源数据集 UT-HAR^[24]上进行综合实验。该数据集是通过配备 3 对天线的英特尔 5300 网卡,并利用 Linux 802.11n CSI 工具,在同一场景下采集到的 7 种不同活动的的数据。每对天线记录 30 个子载波数据,即每个数据由 30 个复数组成,最终形成 90 维特征向量。基于公开数据集的实验结果如表 2 所示。ResNet101 实现了 99.56% 的最高准确率,BiLSTM 实现了 87.26% 的最低准确率。与以 GRU 为代表的时序模型相比,D-GMAFi 的识别准确率更高。与以残差结构为代表的 ResNet 相比,D-GMAFi 实现了 98.73% 的识别准确率,比 ResNet101 最高准确率低了 0.83%,其精确率、召回率和 F1 分数平均低了 1%,但是参数量和计算复杂度远远低于 ResNet101。

不同模型在数据集上的训练损失曲线如图 10 所示。在整个训练过程中,D-GMAFi 在前 20 训练轮次内快速收敛,且训练曲线平稳无波动。在 UT-HAR 数据集上的实验中,D-GMAFi 与基准模型在性能表现上相差无异,达到了令人满意的识别结果,进一步验证了所提模型的有效性。

3.5 消融实验和超参数实验

利用实验室环境下采集的数据进行消融实验,实验结果如表 3 所示。当标准的 CNN 模块替换 DSC 模块后,模型的总体平均准确率为 86.3%,与基准模型 D-GMAFi 相比,性能下降了 3%。这表明 DSC 模型架构能够提取

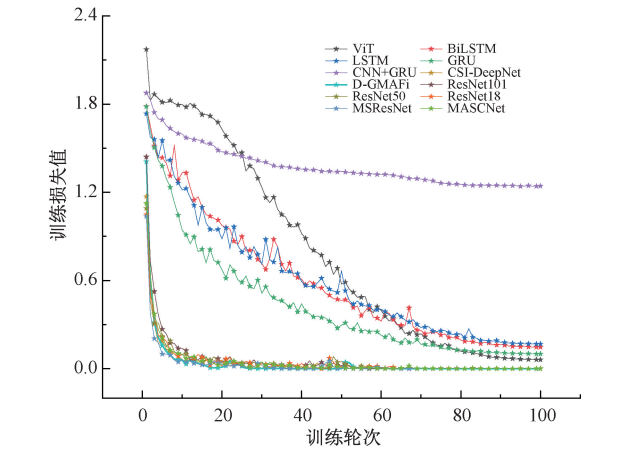


图 10 UT-HAR 数据集上不同模型的损失值

Fig. 10 Loss values of different models on the UT-HAR dataset

表 2 在 UT-HAR 数据集上不同深度神经网络上的实验结果

模型方法	ACC/%	Pr/%	Re/%	F1/%	FLOPs	Params
ResNet18	99.45	99.05	98.95	98.99	1.399×10^{-9}	11.820×10^{-6}
ResNet50	99.49	99.11	99.1	99.1	2.496×10^{-9}	23.549×10^{-9}
ResNet101	99.56	99.27	99.22	99.24	4.762×10^{-9}	42.568×10^{-6}
CNN+GRU	90.42	88.66	86.03	86.92	39.973×10^{-6}	1.430×10^{-6}
GRU	90.57	88.44	87.28	87.76	7.600×10^{-6}	30.407×10^{-3}
LSTM	89.33	87.89	86.78	87.23	10.112×10^{-6}	40.391×10^{-3}
BiLSTM	87.26	88.26	84.76	85.32	20.224×10^{-6}	80.237×10^{-3}
ViT	94.59	93.13	92.82	92.94	273.167×10^{-6}	10.551×10^{-6}
MASCNet	99.32	98.64	98.84	98.72	734.922×10^{-6}	1.573×10^{-6}
MSResNet	99.02	98.51	98.61	98.54	692.641×10^{-6}	951.503×10^{-3}
CSI-DeepNet	98.59	97.50	97.74	97.60	452.308×10^{-6}	489.291×10^{-3}
D-GMAFi	98.73 _{↓0.83}	98.10 _{↓1.17}	98.15 _{↓1.07}	98.11 _{↓1.13}	216.011×10^{-6}	454.063×10^{-3}

注:加粗表示排名第 1;下划线表示排名第 2,下同

表 3 消融实验

模型方法	ACC/%	FLOPs/ $\times10^{-6}$	Params
CNN 替换 DSC	86.38 $\pm$ 0.88	820.799	731.120×10^{-3}
注意力机制不分组	<u>86.65$\pm$1.02</u>	607.584	652.488×10^{-3}
去除分组注意力机制	85.78 $\pm$ 0.32	<u>241.248</u>	448.968×10^{-3}
标准注意力机制替换	85.29 $\pm$ 0.71	595.173	645.576×10^{-3}
分组注意力机制			
去除 GRU	74.55 $\pm$ 2.16	142.706	2.515×10^{-6}
D-GMAFi	89.30$\pm$0.28	259.220	460.488×10^{-3}

在消融实验的基础上,通过设计分组数分别为 1、4、8、16 和 32 共 5 组不同的超参数实验来确定所提注意力机制的最优分组数。如图 11 所示,可以看出随着分组数的增加,模型的准确率先下降后上升,且分组数为 32 时准确率最高,相比于分组数为 8 时,平均准确率 3.57%。在多次实验下,分组数为 32 时的误差最小,总体波动最小。因此,本文选择的最优分组数为 32,此时的 D-GMAFi 性能表现更好。

到更加有效的局部空间信息。进一步探究注意力机制的影响,在分别去除注意力机制和未分组的注意力机制后,模型的总体平均准确率分别为 85.78%和 86.65%。当标准注意力机制替换所提分组注意力机制后,模型的总体平均准确率为 85.29%,与基准模型 D-GMAFi 相比,该模型的计算复杂度由 259.220×10^{-6} FLOPs 增加到 595.173×10^{-6} FLOPs,参数量由 460 488 增加到 645 576,且性能下降了 4%,计算效率显著降低。上述结果说明所提注意力机制能够关注到重要的特征信息,且分组策略可以有效实现跨通道信息融合,进而增强特征的表征能力。当去除 GRU 模块后,模型的总体平均准确率 74.55%,性能下降了 14.75%,说明 GRU 可以有效捕捉数据中的时序关联性。综上所述,验证了所提模型中各模块的有效性。

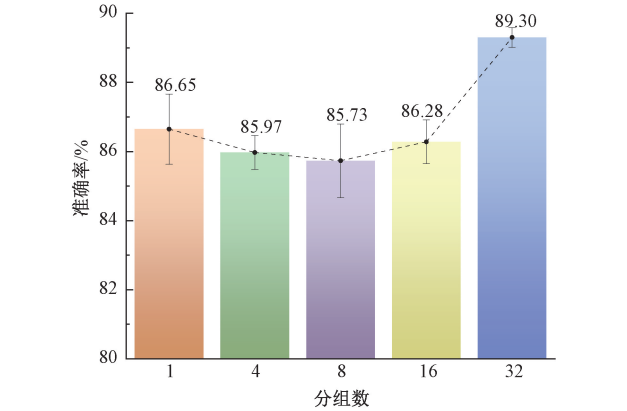


图 11 不同分组数的准确率

Fig. 11 Accuracy with different group numbers

3.6 模型的鲁棒性和泛化性验证

由于 UT-HAR 数据集是在同一环境下采集的,不能够很好地验证 D-GMAFi 在不同场景下对未见样本进行识别时的鲁棒性和泛化能力,因此本文将 D-GMAFi 与

ResNet18、ResNet50、ResNet101、EMANet^[25]、SANet^[26]、SGENet^[27]、SENet^[28]、ABLSTM^[29]、MASCNet^[30]、MSResNet^[31]和CSI-DeepNet^[32]基准模型在自采数据集WiACT上进行了综合实验对比。ResNet18/50/101等模型并未使用1D卷积核,而是将CSI数据转换为二维图像的输入格式,具体输入形式为(批次,1,时间步长,子载波个数)。SANet、SGENet、SENet与D-GMAFi的主要区别在于它们所采用的注意力机制,而网络的其他结构是相同的。EMANet、MASCNet、MSResNet和CSI-DeepNet则在基础架构上增加了GRU模块,为了更好地处理和适应数据中的时序特征。实验结果如表4所示,D-GMAFi在单人单场景和多人多场景下分别实现了94.63%和91.06%的平均准确率,从单人单场景到多人多场景的迁移过程中,D-GMAFi的性能仅下降了3.57%,依旧可以达到令人满意的效果。D-GMAFi的参数数量和计算复杂度分别为 460.488×10^{-3} 和 259.220×10^{-6} ,假设每个参数使用32位浮点数表示,即每个参数占用4 byte,则整个模型对于内存需求大约为 $460.488\times4\times1\,000=18\,411\,952\text{Bytes}\approx1.76\times10^{-6}$,按照70%的硬件利用率来算,运行模型所需要的实际需求大约 $259.220\div0.7\approx370.314\times10^{-6}$ 。为进一步评估模型的推理效率,实验中对模型先进行了100次预热迭代,然后在1 000次迭代上测量平均推理时间,D-GMAFi的推理时间为2.78 ms,有利于后续嵌入式设备的实时部署。

表 4 在 WiACT 数据集上不同深度神经网络的评估结果

模型方法	ACC/%		FLOPs	Params
	单人单场景	多人多场景		
ResNet18	95.50	90.79	1.633×10^{-9}	11.194×10^{-6}
ResNet50	94.32	90.36	2.909×10^{-9}	23.601×10^{-6}
ResNet101	94.38	88.79	5.575×10^{-9}	42.619×10^{-6}
EMANet	90.01	84.47	177.486×10^{-6}	354.198×10^{-3}
SANet	92.75	88.49	242.055×10^{-6}	448.980×10^{-3}
SGENet	92.58	88.38	241.786×10^{-6}	448.968×10^{-3}
SENet	93.78	88.11	244.244×10^{-6}	457.160×10^{-3}
ABLSTM	89.37	85.27	197.605×10^{-6}	691.105×10^{-3}
MASCNet	92.68	88.62	881.912×10^{-6}	1.578×10^{-6}
MSResNet	91.2	88.08	950.138×10^{-6}	1.353×10^{-6}
CSI-DeepNet	93.54	89.07	542.775×10^{-6}	495.716×10^{-3}
D-GMAFi	<u>94.63</u> _{↓ 0.87}	91.06 _{↑ 0.27}	259.220×10^{-6}	460.488×10^{-3}

为进一步分析模型对于每个活动识别的效果,绘制了单人单场景下4种最优模型的混淆矩阵图,如图12所示。由于有32个活动类别,为了混淆矩阵图的简洁,对每个类别进行了数字编号,活动类别与图中数字的具体对应关系如表5所示。ResNet18对所有活动的识别准确

率均在85%以上,D-GMAFi对所有活动的识别准确率均在83%以上,ResNet50和ResNet101对所有活动的准确率分别在75%和81%以上。以B、C和S等这类带有弧线的字母为例,当在弧线处书写速度较慢时,弧线处会近似一条直线与静态路径平行,导致CSI的动态变化很小,模型几乎学习不到这部分的有效内容,那么模型就会将两个不相关的动作进行误判。ResNet18、ResNet50和ResNet101 3个模型均将画G的动作误判成走路,误分类的概率分别为9.09%、25%和18.2%。而D-GMAFi将两个完全不相关的动作进行了区分,得益于分组多尺度注意力机制的跨通道信息融合,学习到了同一位置不同通道之间的关联性;以画对号为例,画对号时的抬手动作和落手动作会产生一个和画M相似的前部分和后部分,使得模型产生误判。4类最优模型均产生了误判,其中ResNet18误分类概率最高,达到10%,而D-GMAFi误分类概率最低。

表 5 序号对应的具体活动类别

Table 5 Activity categories corresponding to the index					
编号	活动类别	编号	活动类别	编号	活动类别
0	画 A	8	画 H	16	画 P
1	画 B	9	画 I	17	画 Q
2	画 C	10	画 J	18	画 R
3	蹲下	11	画 K	19	画对号
4	画 D	12	画 L	20	画 S
5	画 E	13	画 M	21	坐下
6	画 F	14	画 N	22	起立
7	画 G	15	画 O	23	画 T
				24	画 U
				25	画 V
				26	画 W
				27	走路
				28	画叉号
				29	画 X
				30	画 Y
				31	画 Z

为减少单次实验准确率的偶然性偏差,在多人多场景下进行了5次实验,并针对5种不同量级的模型绘制了箱型图。如图13所示,D-GMAFi的平均准确率在91%左右,实现了最高的准确率,且每次实验的结果波动较小;MSResNet的平均准确率在88%左右,整体准确率最低,且每次实验的结果波动较大。在不同量级下,所提模型实现了最优的性能。如图14所示,为进一步验证所提注意力机制的有效性,绘制了D-GMAFi与其他基准注意力网络在准确率、精确率、召回率及F1分数4项指标上的对比图。结果显示,D-GMAFi在所有评估指标上均优于基准模型,充分证明了所提注意力机制的潜力与有效性。所有模型的训练准确率曲线和训练损失曲线分别如图15(a)和(b)所示,D-GMAFi在前100训练轮次内快速收敛,随后慢慢趋于平缓,整个训练过程没有出现较大波动。为了验证所提方法可以有效地区分不同活动特征,进一步通过t-SNE^[33]的方法可视化原始数据分布和分类后的数据分布,如图16所示。图16(a)是预处理后的原始数据分布,图16(b)是经过分类器后的数据分布。

(a) D-GMAFi混淆图
(a) D-GMAFi model confusion matrix

(b) ResNet18混淆图
(b) ResNet18 model confusion matrix

(c) ResNet50 model confusion matrix

(d) ResNet101 model confusion matrix

图 12 4 种最优模型的混淆图

Fig. 12 Confusion matrices of four optimal models

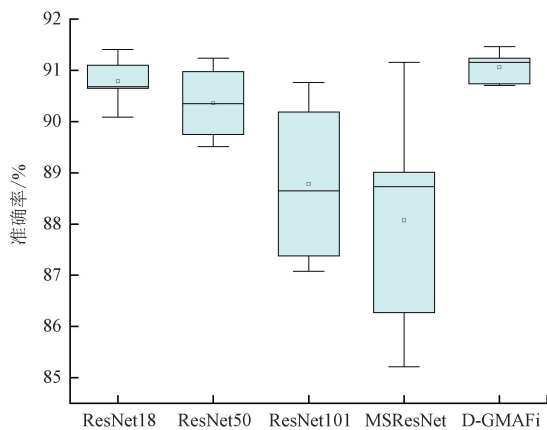


图 13 5 类不同量级模型的箱型图

Fig. 13 Box plot of five models with different magnitudes

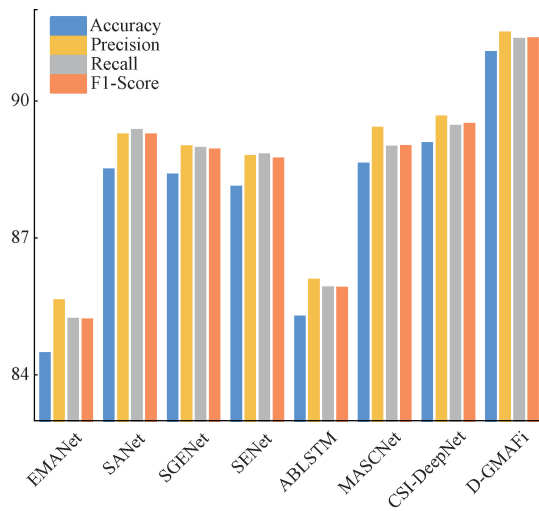
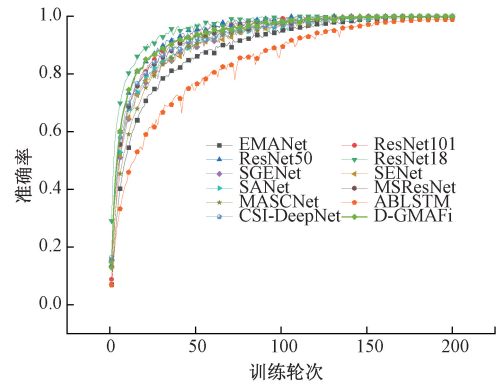
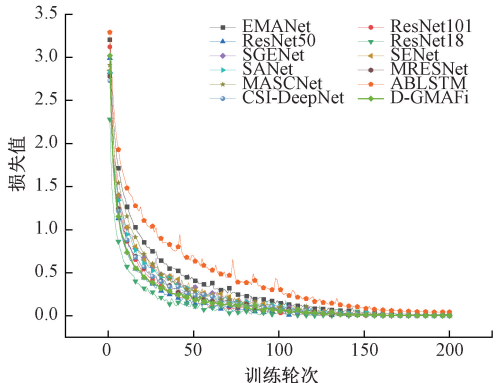


图 14 8 类注意力网络的评估图

Fig. 14 Evaluation plots for eight types of attention networks



(a) 训练准确率
(a) Training accuracy

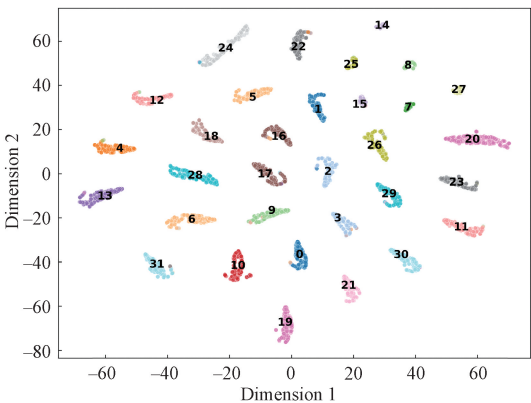


(b) 训练损失
(b) Training losst

图 15 多人多场景下的训练准确率和训练损失
Fig. 15 Training accuracy and loss in a multi-person, multi-scene scenario

预处理后数据分布比较分散,每个类别难以区分;在经过 D-GMAFi 模型进行分类识别后,数据样本呈现明显聚类,每个类别的间隔非常清晰,表明所提模型能够有效区分不同类别特征。

综上所述,所提的数据预处理方法提供了高质量的数据,对于擅长处理图像的神经网络也取得了较好地识



(b) 分类后的数据分布
(b) Data distribution after classification

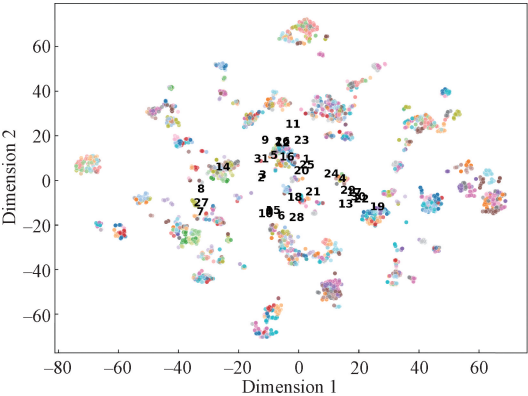
图 16 t-SNE 可视化数据分类前后的分布

Fig. 16 t-SNE visualization of data distribution before and after classification

别效果。通过在开源数据集 UT-HAR 和自采数据集 WiACT 上将所提模型 D-GMAFi 与不同基准模型进行综合实验,分别验证了模型的有效性和鲁棒性。在从单人单场景到多人多场景的迁移中,所提模型整体性能表现较优,可以稳定地识别不同类别的活动。

4 结 论

本文提出了一种融合分组多尺度注意力的 WiFi 人体活动提取与感知方法,旨在通过提取 CSI 数据中的有效活动区间和信道相关特征,提升活动识别的准确性与鲁棒性。本方法通过天线选择和共轭相乘法分别提取对活动感知最为敏感的 CSI 幅值和相对相位信息特征,然后采用小波降噪等方法对数据进行降噪和平滑处理。其次,设计了一种基于方差阈值的数据提取算法来提取有效的活动数据区间,避免了人工裁剪或引入不相关数据。再次,采用联合的 CSI 幅值和相位信息作为数据特征,摒弃繁琐的人工特征设计,显著减少人为干预。最后,提出一种基于分组多尺度注意力机制的端到端神经网络模型 D-GMAFi 来捕获数据特征之间的深层关联性,并实现对不同活动类别的精确分类。其中,D-GMAFi 模型中的深度可分离卷积模块对特征进行权重分配和空间信息的有效提取;分组多尺度注意力机制捕捉通道和空间之间的局部信息关联及动态关注多个维度上的重要信息,并进行跨通道的信息融合;门控循环单元用于捕捉数据在不同时间点之间的关联性。实验结果表明,在单人单场景到多人多场景的迁移学习任务中,D-GMAFi 整体性能优于基准模型,并在不同用户与场景下保持了稳定的识别效果,体现出一定程度的鲁棒性与面向应用场景的适应能力。考虑到真实环境的复杂性,未来研究重点在于将



(a) 原始数据分布
(a) Raw data distribution

所提方法部署到实际应用场景中,进一步提升其鲁棒性与适用性。

参考文献

- [1] 王颖,杨志家,谢闯,等. 基于惯性传感器的 HAR 数据采集系统设计[J]. 电子测量技术, 2023, 46(23): 146-152.
- WANG Y, YANG ZH J, XIE CH, et al. Design of a HAR data acquisition system based on inertial sensors [J]. Electronic Measurement Technology, 2023, 46(23): 146-152.
- [2] 高闻霏,杨波,闵琳孜. 低分辨率红外感知系统的人体行为数据增强[J]. 仪器仪表学报, 2023, 44(1): 163-171.
- GAO W P, YANG B, MIN L Z. Data augmentation for human behavior in low-resolution infrared sensing systems [J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 163-171.
- [3] CHEN C, ZHOU G, LIN Y. Cross-Domain Wi-Fi sensing with channel state information: A survey [J]. ACM Computing Surveys, 2023, 55(11): 1-37.
- [4] ZHANG Y, WANG G, LIU H, et al. Wi-Fi-Based indoor human activity sensing: A selective sensing strategy and a multi-level feature fusion approach[J]. IEEE Internet of Things Journal, 2024, 11(18): 29335-29347.
- [5] XIAO J, LI H, WU M, et al. A survey on wireless device-free human sensing: Application scenarios, current solutions, and open issues[J]. ACM Computing Surveys, 2022, 55(5): 1-35.
- [6] WU X, CHU Z, YANG P, et al. TW-See: Human activity recognition through the wall with commodity Wi-Fi devices [J]. IEEE Transactions on Vehicular Technology, 2018, 68(1): 306-319.
- [7] WANG Y, WU K, NI L M. Wifall: Device-free fall detection by wireless networks[J]. IEEE Transactions on Mobile Computing, 2016, 16(2): 581-594.
- [8] WANG F, GONG W, LIU J. On spatial diversity in Wi-Fi-based human activity recognition: A deep learning-based approach [J]. IEEE Internet of Things Journal, 2018, 6(2): 2035-2047.
- [9] YAN H, ZHANG Y, WANG Y, et al. WiAct: A passive Wi-Fi-based human activity recognition system[J]. IEEE Sensors Journal, 2019, 20(1): 296-305.
- [10] XIAO C, LEI Y, MA Y, et al. DeepSeg: Deep-learning-based activity segmentation framework for activity recognition using Wi-Fi [J]. IEEE Internet of Things Journal, 2020, 8(7): 5669-5681.
- [11] SHENG B, HAN R, CAI H, et al. CDFi: Cross-Domain action recognition using Wi-Fi signals [J]. IEEE Transactions on Mobile Computing, 2024, 23(8): 8463-8477.
- [12] WANG H, ZHANG D, WANG Y, et al. RT-Fall: A real-time and contactless fall detection system with commodity Wi-Fi devices [J]. IEEE Transactions on Mobile Computing, 2016, 16(2): 511-526.
- [13] QIAN K, WU C, ZHOU Z, et al. Inferring motion direction using commodity Wi-Fi for interactive exergames [C]. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017: 1961-1972.
- [14] ZHANG C, JIAO W. ImgFi: A high accuracy and lightweight human activity recognition framework using CSI image[J]. IEEE Sensors Journal, 2023, 23(18): 21966-21977.
- [15] ZHANG Y, ZHENG Y, QIAN K, et al. Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 8671-8688.
- [16] LI C, LIU M, CAO Z. WiHF: Gesture and user recognition with Wi-Fi [J]. IEEE Transactions on Mobile Computing, 2020, 21(2): 757-768.
- [17] GU Y, YAN H, ZHANG X, et al. Attention-based gesture recognition using commodity Wi-Fi devices [J]. IEEE Sensors Journal, 2023, 23(9): 9685-9696.
- [18] HUSSAIN A, CHEN Y, ULLAH A, et al. WiSigPro: Transformer for elevating CSI-based human activity recognition through attention mechanisms [J]. Expert Systems with Applications, 2024, 258: 124976.
- [19] SHI Z, CHENG Q, ZHANG J A, et al. Environment-robust Wi-Fi-based human activity recognition using enhanced CSI and deep learning [J]. IEEE Internet of Things Journal, 2022, 9(24): 24643-24654.
- [20] HE Y, CHEN Y, HU Y, et al. Wi-Fi vision: Sensing, recognition, and detection with commodity MIMO-OFDM Wi-Fi [J]. IEEE Internet of Things Journal, 2020, 7(9): 8296-8317.
- [21] 师浩伟,王薇,裴东兴,等. 基于改进小波变换的火炮膛压测试降噪方法[J]. 国外电子测量技术, 2022, 41(3): 41-44.
- SHI H W, WANG W, PEI D X, et al. Denoising method for gun chamber pressure test based on improved wavelet transform [J]. Foreign Electronic Measurement Technology, 2022, 41(3): 41-44.
- [22] ARSHAD S, FENG C, LIU Y, et al. Wi-chase: A Wi-Fi based human activity recognition system for sensorless environments [C]. 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks. IEEE, 2017: 1-6.

- [23] GUO Z, XIAO F, SHENG B, et al. WiReader: Adaptive air handwriting recognition based on commercial Wi-Fi signal [J]. IEEE Internet of Things Journal, 2020, 7(10): 10483-10494.
- [24] YOUSEFI S, NARUI H, DAYAL S, et al. A survey on behavior recognition using Wi-Fi channel state information[J]. IEEE Communications Magazine, 2017, 55(10): 98-104.
- [25] LI X, ZHONG Z, WU J, et al. Expectation-maximization attention networks for semantic segmentation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9167-9176.
- [26] ZHANG Q L, YANG Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2021: 2235-2239.
- [27] LI X, HU X, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks[J]. ArXiv preprint arXiv:1905.09646, 2019.
- [28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [29] CHEN Z, ZHANG L, JIANG C, et al. Wi-Fi CSI based passive human activity recognition using attention based BLSTM[J]. IEEE Transactions on Mobile Computing, 2018, 18(11): 2714-2724.
- [30] CHEN W, SHI K. Multi-scale attention convolutional neural network for time series classification[J]. Neural Networks, 2021, 136: 126-140.
- [31] WANG F, HAN J, ZHANG S, et al. Csi-Net: Unified human body characterization and pose recognition[J]. ArXiv preprint arXiv:1810.03064, 2018.
- [32] KABIR M H, HASAN M A, SHIN W. CSI-DeepNet: A lightweight deep convolutional neural network based hand gesture recognition system using Wi-Fi CSI signal[J]. IEEE Access, 2022, 10: 114787-114801.
- [33] 陈剑,程明. 基于 tSNE-ASC 特征选择和 DSMT 融合决策的滚动轴承声振信号故障诊断[J]. 电子测量与仪器学报, 2022, 36(4): 195-204.
- CHEN J, CHENG M. Fault diagnosis of rolling bearing acoustic-vibration signals based on tSNE-ASC feature selection and DSMT fusion decision [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(4): 195-204.

作者简介



李振, 2021 年于安徽科技学院获得学士学位, 现为天津职业技术师范大学研究生, 主要研究方向为无线感知。

E-mail: 0521231027@tute.edu.cn

Li Zhen received his B. Sc. degree from Anhui Science and Technology University in 2021. He is currently a M. Sc. candidate at Tianjin University of Technology and Education. His main research interest includes wireless sensing.



赵志彪(通信作者), 2012 年于燕山大学获得学士学位, 2019 年于燕山大学获得博士学位, 现为天津职业技术师范大学讲师, 主要研究方向为工业智能控制、机器学习。

E-mail: zhaozhibiao@tute.edu.cn

Zhao Zhibiao (Corresponding author) received his B. Sc. degree from Yanshan University in 2012 and Ph. D. degree from Yanshan University in 2019, respectively. He is currently a Lecturer at Tianjin University of Technology and Education. His main research interests include industrial intelligent control and machine learning.