

DOI: 10.13382/j.jemi.B2609243

CSI 时频特征融合的嵌入式细粒度手势识别研究*

崔凌云^{1,2} 赵志彪^{1,2} 石 岩^{1,2} 葛长宇^{1,2} 李 振^{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:针对嵌入式 WiFi 感知场景中 CSI 信息维度受限导致细粒度字母手势识别精度不足的问题,提出一种轻量化的 CSI 细粒度手势识别方法。为提升 CSI 数据的信号质量和序列一致性,对 CSI 序列依次进行 Hampel 异常值剔除、中值滤波、小波降噪和三次样条插值的分阶段预处理,从而抑制噪声、提高信号信噪比并稳定手势特征;随后,为改善单天线 CSI 空间信息维度受限对细粒度手势识别的影响,设计融合时频特征的双分支网络 DFANet,分别提取频域子载波相关特征和时间依赖特征,并结合树突网络实现高效分类。实验结果表明,模型在实验室、办公室和教室 3 种场景下的平均识别准确率分别为 88.51%、92.64% 和 91.60%,整体平均准确率为 91.21%;模型参数量为 147.880×10^3 ,计算复杂度为 49.534×10^6 。与 ResNet50 相比,在识别精度相近条件下,模型参数量由 21.291×10^6 降至 147.88×10^3 ,计算复杂度由 835.781×10^6 降至 49.534×10^6 。说明该方法在保持细粒度字母手势识别性能的同时具有较低的计算开销和良好的轻量化特征,可为后续资源受限嵌入式 WiFi 感知系统的部署提供方法支持。

关键词: ESP32; 细粒度手势; 时频特征; 信道状态信息(CSI); 双分支网络

中图分类号: TP183; TN83 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Embedded fine-grained hand gesture recognition based on CSI time-frequency feature fusion

Cui Lingyun^{1,2} Zhao Zhibiao^{1,2} Shi Yan^{1,2} Ge Changyu^{1,2} Li Zhen^{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:To address the problem that limited CSI dimensionality in embedded WiFi sensing scenarios leads to insufficient recognition accuracy for fine-grained alphabet gestures, a lightweight CSI-based fine-grained gesture recognition method is proposed. A staged preprocessing pipeline, including Hampel outlier removal, median filtering, wavelet denoising, and cubic spline interpolation, is applied to CSI sequences to improve signal quality and sequence consistency. Subsequently, a dual-branch network, termed DFANet, is designed to fuse time-frequency features by extracting subcarrier correlation features in the frequency domain and temporal dependency features in the time domain, and a dendritic network is employed for efficient classification. Experimental results show that the proposed model achieves average recognition accuracies of 88.51%, 92.64%, and 91.60% in laboratory, office, and classroom environments, respectively, with an overall average accuracy of 91.21%. The model contains 147.880×10^3 parameters and requires 49.534×10^6 FLOPs. Compared with ResNet50, under comparable recognition accuracy, the number of parameters is reduced from 21.291×10^6 to 147.880×10^3 , and the computational complexity is reduced from 835.781×10^6 to 49.534×10^6 FLOPs. The results demonstrate that the proposed method achieves effective fine-grained alphabet gesture recognition while maintaining low computational cost and lightweight characteristics, providing methodological support for future deployment in resource-constrained embedded WiFi sensing systems.

Keywords: ESP32; fine-grained gesture recognition; time-frequency features; channel state information (CSI); dual-branch network

0 引言

随着物联网与智能交互技术的不断发展,手势识别^[1-2]因具备非接触和自然交互等特点,在人机交互^[3]、自动驾驶^[4]及智能家居^[5]等领域受到广泛关注。现有手势识别方法主要包括穿戴式感知^[6]和视觉感知^[7]两类,其中,穿戴式方法需要用户长期佩戴,易对日常活动造成干扰;而视觉感知方法虽具有较高识别精度,但存在隐私泄露的风险,且对光照条件和环境变化较为敏感,从而限制了其在复杂场景中的应用。其中,信道状态信息(channel state information, CSI)能够细粒度表征无线信号在传播过程中的幅度与相位变化,被广泛用于手势识别^[8]。

近年来,针对 CSI 手势识别问题,已有研究主要围绕数据处理与模型设计两个方面展开。在数据预处理方面,郭梦丽等^[9]采用巴特沃斯滤波、多路径去除、离散小波变换和手势分割等方法,有效抑制环境干扰并提取有效动作片段;Ali 等^[10]采用相位校准、Hampel 异常值剔除和 DWT 去噪,改善信号质量并减弱了噪声干扰;Zhang 等^[11]在 WiNum 中利用 DWT 对 CSI 幅值序列进行去噪,并设计基于信息熵的自适应分割算法提取动作片段,提高了信号分割的有效性。这些方法有效的解决了噪声抑制和动作片段提取问题,但在样本长度统一、序列一致性和复杂场景适应性方面仍存在不足。在模型设计方面,Bastwesy 等^[12]通过 WMA 去噪和相位校正后,结合 CNN、LSTM 和 ABLSTM 进行特征学习与分类,以提升复杂环境和多用户场景下的识别能力;Jiang 等^[13]提出的 WiGAN 在活动检测、线性插值、DWT 去噪和子载波选择基础上结合 GAN 与 CNN 进行特征提取,并通过样本生成与多层特征融合增强特征表达,以缓解小样本和环境变化带来的影响。尽管取得一定进展,但大多数方法依赖多天线设备获取空间特征,增加了模型参数数量和计算复杂度,难以满足资源受限场景下对模型轻量化的需求。因此,研究者开始关注适用于低成本、低功耗感知场景的轻量化手势识别方法。Kabir 等^[14]验证了 ESP32 单天线 Wi-Fi 感知在字母手势识别中的可行性,并通过预处理、DS-Conv、特征注意和残差结构相结合的方法提升识别性能;Kocheta 等^[15]则针对 ESP32 单天线条件下 CSI 空间信息有限的问题,提出仅利用 CSI 幅值信息并结合信号处理流程与轻量化 LSTM 网络的方法,实现弱时序信号的有效感知与估计,但细粒度手势特征表达仍然受限。

针对上述问题,本文构建 ESP32 单天线 CSI 数据采集系统。随后,采用 Hampel 异常值剔除、中值滤波、小波降噪和三次样条插值相结合的分阶段预处理流程。进一

步提出融合时频特征的双分支轻量化模型 DFANet,并结合树突网络在降低模型复杂度的同时实现有效分类。

1 相关理论

WiFi 无线感知技术通过捕捉无线信号在传播过程中的特征变化实现人体活动的感知。在信号从发射端到接收端的传播过程中,会受到多径传播、反射、衰落和环境干扰等因素的影响,其传播路径可分为静态路径和动态路径两类,如图 1 所示。人体手势动作会引起信道频率响应(channel frequency response, CFR)的变化,表现为 CSI 幅值随时间的变化。

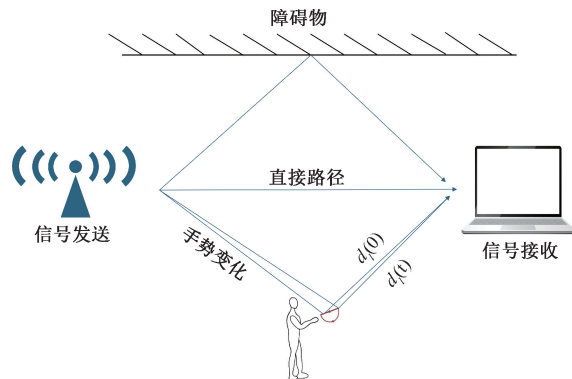


图 1 WiFi 信号传播示意图

Fig. 1 Schematic of WiFi signal propagation

CSI 是对信道频率响应的估计结果,可在频域描述每个子载波的幅度和相位特征。接收信号可表示为:

$$\mathbf{Y}(f) = \mathbf{X}(f) \times \mathbf{H}(f) + \mathbf{N}(f) \quad (1)$$

式中: $\mathbf{Y}(f)$ 为接收信号; $\mathbf{X}(f)$ 为发射信号; $\mathbf{H}(f)$ 为信道频率响应; $\mathbf{N}(f)$ 为噪声。

在实际室内环境中,无线信号通常通过多条路径传播,其信道响应应为各路径信号叠加的结果可表示为:

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

式中: L 表示传播路径数量; α_l 表示第 l 条路径的衰减退系数; f 为子载波频率; τ_l 表示传播时延; l 表示路径索引 ($1 \sim L$)。由式(2)可知,接收端 CSI 信号是多径信号叠加的结果,既反映手势动作引起的动态变化,也受到环境干扰和硬件噪声的影响。由于 ESP32 为单天线,接收的路径数量有限,空间信息维度和可用子载波数量受到限制,再加上环境扰动、载波频率偏移(carrier frequency offset, CFO)和采样频率偏移(sampling frequency offset, SFO)等因素,微小手势信号容易被噪声掩盖,从而增加细粒度手势识别的难度。

针对上述问题,本文从数据预处理与特征提取两个方面进行优化。

2 数据处理

2.1 数据采集

为开展嵌入式平台下的细粒度英文字母手势识别研究,本文基于 ESP32 平台搭建 CSI 数据采集系统,主要由两台配备增益天线的 ESP32 设备组成,分别作为发射端和接收端^[16],用于 CSI 数据的发射、采集与保存。ESP32 设备通过 CSI API 能够捕获 64 个子载波,经硬件筛选后可用子载波为 52 个,每个子载波的信道频率响应由实部和虚部的符号字节表示,经算术运算后可得到各子载波对应的幅值和相位信息。然而,受 ESP32 硬件缺陷与收发端同步误差影响,CSI 相位信息存在系统性偏差^[17],难以用于可靠的特征提取。因此,本文仅采用 CSI

幅值信息进行后续分析,以提高数据处理的稳定性与模型学习的可靠性。

在数据采集过程中,选取 A、B、C、D、E、F、G、J、M、P 共 10 个英文字母手势作为本次数据采集的对象。部分字母的书写顺序及对应的 CSI 幅值变化如图 2 所示,黄色箭头代表志愿者书写的运动方向。为提升数据集的场景泛化能力,实验分别在实验室、办公室和教室 3 种环境下进行采集如图 3 所示,设备之间的距离统一设置为 2 m。每个样本的采集时间为 3 s,接收端以每秒 100 个包的速率接收 WiFi 数据包,信号频率为 2.4 GHz,信道带宽为 20 MHz。为评估模型在不同用户条件下的识别能力,实验选取 5 名不同性别、身高和体重的志愿者与数据采集,构建 ESP-HAR 数据集,具体信息如表 1 所示。

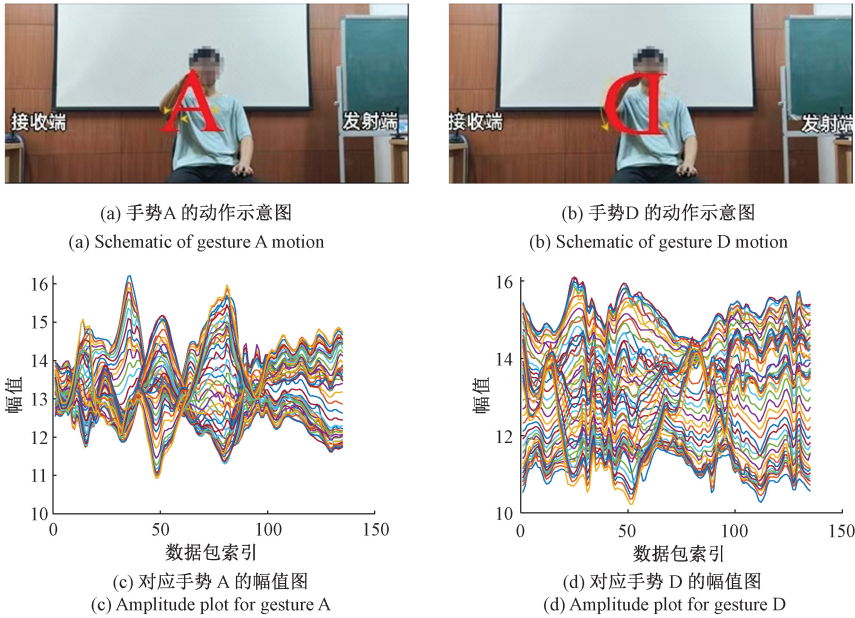


图 2 手势 A 和 D 的动作图及对应幅值图

Fig. 2 Motion schematics and corresponding amplitude plots for gesture A and D



图 3 数据采集场景

Fig. 3 Data acquisition scenes

表 1 ESP-HAR 数据集详细信息

Table 1 Detailed information of the ESP-HAR dataset

项目	内容
数据集	ESP-HAR
采集工具	ESP32 S3
活动类别	10 类
活动名称	画 A、B、C、D、E、F、G、J、M、P 共 10 类英文字母手势
数据尺寸	(1,52,135)/(天线数,子载波数,数据包数)
志愿者人数	5 名
场景数	3 种典型场景
总样本数	10 713 个

2.2 数据预处理

原始 CSI 数据存在离群值、脉冲噪声、随机噪声及序列长度不一致等问题^[18],从而影响后续特征提取的稳定性与准确性。为此,本文采用异常值剔除、中值滤波、噪声抑制以及序列对齐的分阶段预处理流程。

1) Hampel 异常值剔除

由于 CSI 幅值序列中存在孤立的离群值,降低数据的有效性,干扰后续的特征提取与模型训练。因此首先采用 Hampel 滤波器进行异常值剔除。Hampel 滤波^[10]基于滑动窗口内的中值与中位数绝对偏差 (median absolute deviation, MAD) 进行异常值检测,不依赖数据的正态分布假设,具有较强鲁棒性,适用于 CSI 这类非平稳时序信号的异常值剔除,其运算过程如式(3)所示。

$$\begin{aligned} W_i &= [X_{i-k}, \dots, X_k, \dots, X_{i+k}] \\ m_i &= \text{median}(W_i) \\ \text{MAD}_i &= \text{median}(|X_i - m_i|) \\ s_i &= c \times \text{MAD}_i \\ X &= \begin{cases} X_i, & |X_j - m_i| \leq \text{nsigma} \times s_i \\ m_i, & |X_j - m_i| > \text{nsigma} \times s_i \end{cases} \end{aligned} \quad (3)$$

式中: k 控制窗口范围,用以计算每个数据点中值和 MAD_i ; c 是校正系数,参数 nsigma 控制异常值判定的敏感度; $\text{median}(\cdot)$ 用来统计数据中的中位数; s_i 是尺度量,用于判定异常值; W_i 表示以 i 为中心的窗口数据序列; m_i 表示窗口 W_i 的中值; MAD_i 表示绝对中值偏差; i 表示数据索引; j 表示以 i 为中心的窗口内的数据点索引。以第一个子载波为例,其滤波前后的幅值变化如图 4 所示。然而,Hampel 滤波主要针对孤立的离群值,对由环境干扰引起的脉冲噪声^[19]抑制效果有限。因此,进一步采用中值滤波对 CSI 数据进行平滑处理^[20],如图 5 所示。

2) 小波降噪

在完成 Hampel 异常值剔除和中值滤波后,CSI 数据中仍存在硬件固有噪声和环境微小扰动等因素引起的随

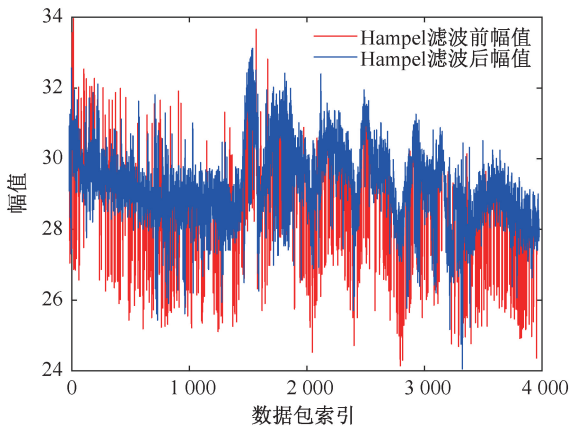


图 4 Hampel 滤波前后的幅值数据对比
Fig. 4 Amplitude data comparison before and after Hampel filtering

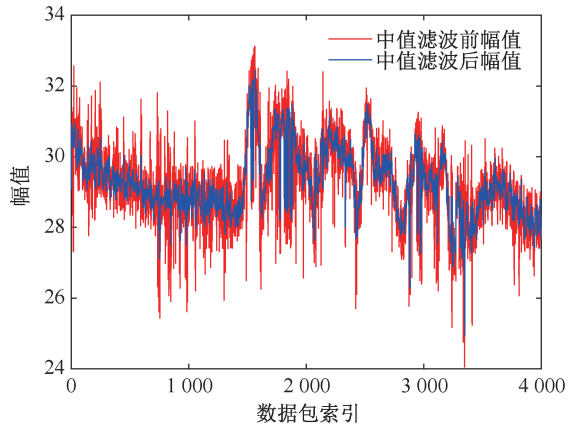


图 5 中值滤波前后的幅值数据对比
Fig. 5 Amplitude data comparison before and after median filtering

机噪声^[21],容易掩盖细粒度手势引起的微弱变化。考虑到低通滤波器在滤除噪声的同时,会模糊局部突变特征,降低动作感知精度^[22]。因此,采用基于小波变换的小波降噪方法对信号进行处理。小波变换能够对信号进行多尺度分解,并有效反映信号在时频域上的变化特征,适用于非平稳信号分析^[23]。

通过小波变化,CSI 信号可分解为近似系数与细节系数,近似系数保留信号的整体趋势和平滑特性,细节系数保留信号的局部变化及噪声成分。通过设定阈值来消除细节系数中的噪声,再将阈值化的细节系数与近似系数进行重构,恢复出原始信号如式(4)所示。

$$C_{ap} = \text{wavedec2}(X_{ap_med}, \text{db4}, \text{level}) \quad (4)$$

式中: $\text{wavedec2}(\cdot)$ 是信号分解的函数; X_{ap_med} 是经过中值滤波后的幅值; db4 表示 Daubechies 4 小波; level 表示分解层数; C_{ap} 表示分解后的小波系数。由于 CSI 信号

具有非平稳性及快速变化的特点,因此选择 db4^[24] 小波函数来捕捉信号中的突变和边缘信息;db4 是一种紧支撑小波,具有处理非平稳信号的优势,且可以给时频域提供较好的分辨率;level 是分解层级数。在利用式(4)将信号分解完成后,获得各层级的细节系数,然后采用软阈值的方法对其进行降噪,阈值根据绝对偏差中位数来确定。通过式(5)可以实现降噪处理,提高信号质量与数据稳定。

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

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

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

(5)

式中: $\text{coeff}_{\text{level}}$ 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; $\text{denoised_coeff}[i]$ 表示去噪后的小波系数; $\text{threshold}(\cdot)$ 表示阈值函数; σ 为标准化尺度因子; soft 表示采用软阈值函数进行系数收缩。最后通过式(6)将阈值化后的小波系数进行重构,恢复原始信号。

$$\mathbf{X}_{\text{ap_dwt}} = \text{waverec2}(\mathbf{C}_{\text{ap}}, \text{db4}) \quad (6)$$

式中: $\text{waverec2}(\cdot)$ 是重构函数,用以进行信号重构; $\mathbf{X}_{\text{ap_dwt}}$ 表示重构后的信号。为了展示小波降噪后的效果,在中值滤波的幅值上进行小波变换进行对比如图 6 所示。小波降噪在保留局部突变特征的同时降低了随机噪声,为后续特征提取提供了更稳定的输入。

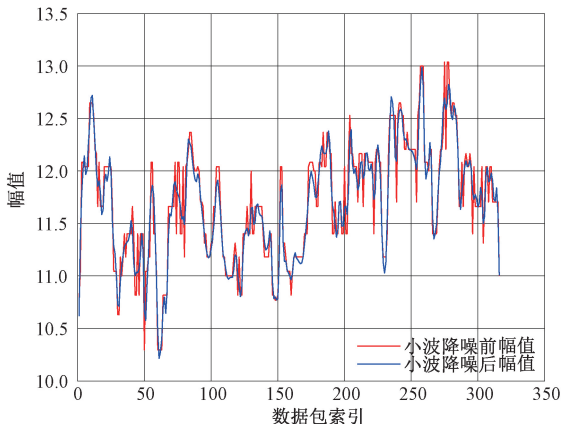


图 6 小波降噪前后的数据对比

Fig. 6 Data comparison before and after wavelet denoising

为了解不同分解层级对小波降噪的影响,绘制了 1~10 分解层级下的均方误差 (mean squared error, MSE)、平均绝对误差 (mean absolute error, MAE) 和信噪比 (signal-to-noise ratio, SNR) 曲线变化图,如图 7 所示。随着层级数的增加,MSE 和 MAE 缓慢上升而 SNR 则在下降,说明

过深分解会消弱有效信号。综合考虑降噪效果与特征保留能力,最终选取分解层级 level 为 1,在保留有效信息的同时,保持较高的信噪比。如图 8 所示,此时 MSE、MAE 和 SNR 分别约为 0.06、0.19 和 34.82 dB。

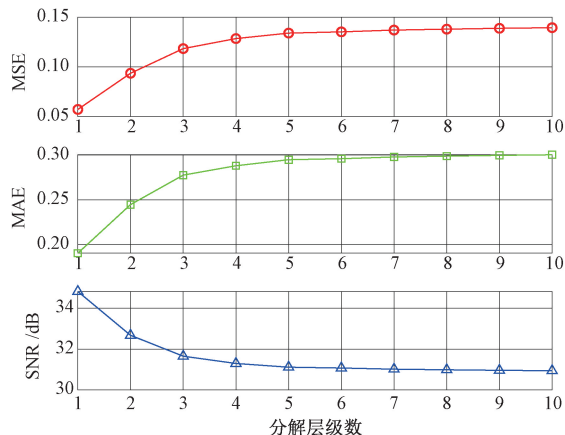


图 7 不同分解层级下的 MSE、MAE 和 SNR 图

Fig. 7 MSE, MAE and SNR at different decomposition levels

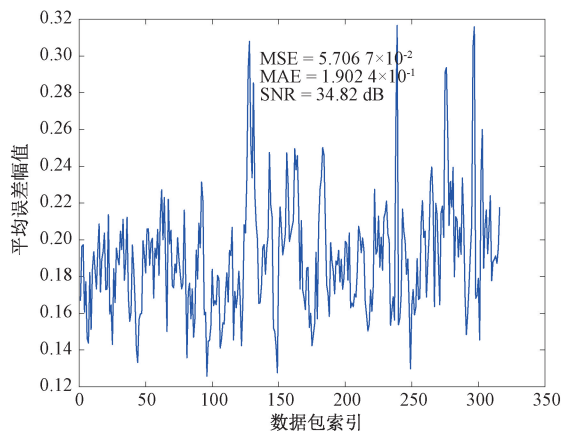


图 8 小波降噪幅值重构误差

Fig. 8 Wavelet denoising amplitude reconstruction error

以手势 P 为例,图 9 表明经过预处理后 52 个子载波的幅值数据整体变化趋势更加清晰。

3) 样条插值

由于不同样本的有效动作片段长度存在差异,需要对 CSI 序列进行对齐处理。相比零填充,三次样条插值能够在保持数据连续性的同时实现序列对齐。因此,本文采用三次样条插值方法对序列进行统一^[25]。

假设区间包含 n 个数据点,通过构建分段区间 $[x_i, x_{i+1}]$ 对应的插值函数 $S_i(x)$ 来拟合给定数据点 $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ 。分段区间对应的插值函数 $S_i(x)$ 如式(7)所示。

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (7)$$

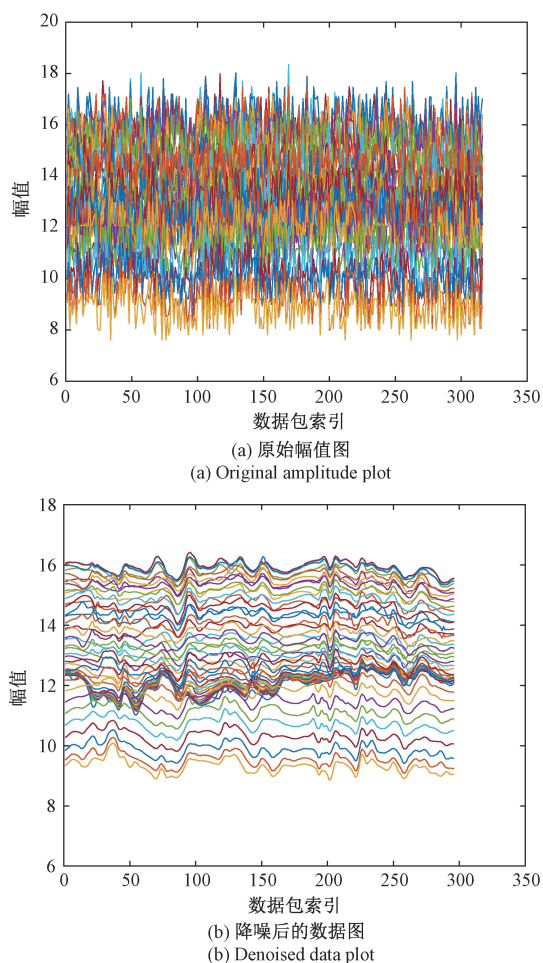


图9 原始数据降噪前后的幅值对比

Fig. 9 Amplitude comparison of original data before and after denoising

其中, $i = 0, 1, 2, \dots, n-1$; a_i, b_i, c_i, d_i 是需要根据所给定数据点和限制条件计算得来的系数; n 表示数据点的总数量; x_i 表示第 i 个数据点的横坐标; $S_i(x)$ 表示第 i 区间的三次样条插值函数。

对于 ESP-HAR 数据集, 预定义长度为 135。对训练集中下采样后有效动作片段的长度分布进行统计后发现其平均长度及主要分布区间均集中在 135 附近。综合考虑有效动作信息的保留、样本序列对齐的需求以及模型输入规模的一致性, 最终统一长度确定为 135, 以使绝大多数样本在保留主要变化趋势的前提下完成序列标准化处理。短序列采用三次样条插值补齐, 长序列进行两端裁剪, 等长序列不进行处理。依旧以上述手势 P 为例, 对提取的有效活动片段先进行下采样操作, 然后进行三次样条插值补全数据, 满足预定义长度。如图 10 所示, 经过下采样和样条插值后的有效数据片段既压缩了数据量, 也保留了信号的主要变化趋势。通过上述处理, 所有 CSI 样本均被统一为长度为 135 的序列, 从而满足后续

深度学习模型的输入要求。同时, 统一的时间尺度有助于时间分支捕捉稳定的时间依赖关系, 并使频率分支学习一致的子载波结构特征, 从而提升模型的稳定性和泛化能力。

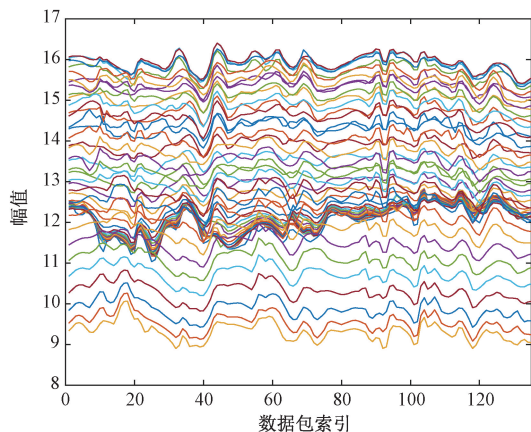


图10 下采样和样条插值后的幅值片段图

Fig. 10 Amplitude segment after downsampling and spline interpolation

经预处理后, CSI 序列的噪声得到了有效的抑制且序列长度实现统一, 为后续模型输入提供了基础。

3 DFANet 模型设计

由于 ESP32 仅配备单根天线, 无法像多天线设备一样通过多个接收天线获取空间维度信息, 因此手势判别所依赖的信息主要来源于 CSI 的频域和时域特征。其中, 频域特征反映不同子载波对人体动作的响应差异, 能够表征手势引起的频率选择性变化及子载波间的关联结构; 时域特征反映 CSI 随时间变化的动态过程, 能够描述手势运动轨迹及关键动作的演变规律。对于细粒度字母手势而言, 仅依赖频域特征难以充分反映手势动作的动态变化, 仅依赖时域特征又难以利用不同子载波之间的差异信息, 两者均存在一定的局限性。因此, 本文采用时频双分支结构, 分别从频域和时域两个角度提取互补特征, 以增强单天线 CSI 对细粒度字母手势的表达能力。基于此, 提出融合 CSI 时频特征信息的双分支手势识别模型 DFANet。其总体框架如图 11 所示, 主要由 CSI 预处理模块、特征提取模块和识别分类模块 3 部分构成。其中, 特征提取模块主要由门控分组多尺度注意力网络 (gated grouped multi-scale attention network, GMA) 和时间全局注意力网络 (temporal global attention network, TGA) 构成, 分别用于提取子载波相关特征和时间动态特征; 识别分类模块采用树突网络 (dendrite net, DD) 进行高效分类。

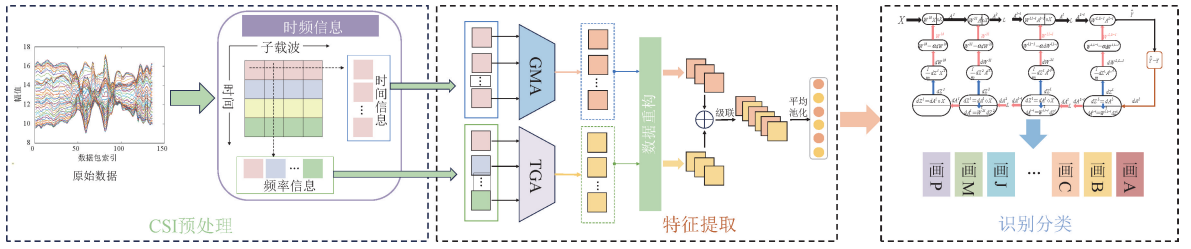


图 11 DFANet 模型的总体框架

Fig. 11 Overall architecture of the DFANet model

3.1 门控分组多尺度注意力网络 GMA

在 ESP32 单天线条件下,由于空间维度信息有限,仅依靠单一子载波难以充分的反映细粒度字母手势之间的差异。然而,CSI 子载波之间仍存在一定相关性^[26],且不同手势对各子载波产生的扰动并不完全相同。对于部分结构相近的字母手势,如‘E’和‘F’、‘P’和‘M’,其运动轨迹存在较高相似性,仅利用时域变化信息容易产生混淆,但其对应的频率响应模式仍然可能存在差异。为此,本文设计了 GMA 频域分支,旨在通过门控分组多尺度注意力机制捕捉同一时刻不同子载波之间的局部相关性,充分利用子载波间的结构化频率响

应差异,弥补单天线 CSI 缺少空间维度信息带来的不足,从而增强频域特征表达能力,提高结构相近的细粒度字母手势的区分能力。如图 12 所示,首先,对于时间 T_i 先进行一个通道维度的嵌入(embedding),增强其表征能力,嵌入后的输入序列 X_T ,如式(8)所示。

$$X_T = \text{ReLU}(\text{BatchNomal}(\text{DSConv}(T_i))) \quad (8)$$

式中:ReLU()是非线性激活函数;DSConv()是深度可分离卷积;BatchNomal()是批量归一化, T_i 为第 i 个时间步的 CSI 子载波幅值输入特征。经过维度嵌入后的输入序列 X_T 被送入门控融合网络(gated fusion network, GFN)进行特征提取。

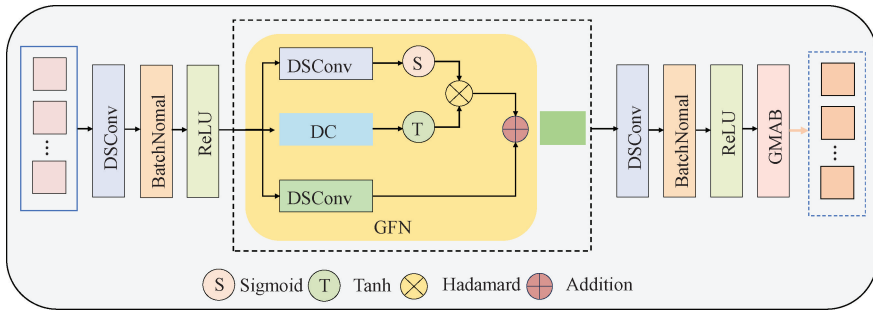


图 12 GMA 网络结构

Fig. 12 Structure of the GMA network

GFN 由门控单元及融合单元两部分组成。其中,门控单元用于提取相邻子载波的局部相关特征;融合单元用于提供更加丰富的全局语义信息。两个卷积层分别接入 Sigmoid 和 Tanh 两个非线性激活函数,以实现门控调制与非线性特征变换。通过结合 Sigmoid 和 Tanh 两个非线性激活函数,可以构成一个高效的门控单元,既增强了特征的表征能力,也抑制了冗余或无关的信息^[27]。在融合单元中,卷积核大小为 1 的深度可分离卷积对输入序列 X_T 进行线性变换,以完成维度匹配和特征融合。设 GFN 的输出为 X_{TC} ,则其信息流通过程如式(9)所示。

$$X_{TC} = \text{GFN}(X_T) =$$

$$\sigma(W_{K=3} * X_T) \cdot g(W_{d=2} * X_T) + W_{K=1} * X_T \quad (9)$$

式中: * 表示卷积操作; · 表示相乘操作; $\sigma(\cdot)$ 表示 Sigmoid 激活函数; $g(\cdot)$ 表示 Tanh 激活函数; K 表示卷

积核大小; d 表示膨胀率; W 表示卷积核权重参数; $W_{K=3}$ 表示 3×3 标准卷积的权重; $W_{d=2}$ 表示膨胀率为 2 的空洞卷积权重; $W_{K=1}$ 表示 1×1 深度可分离卷积权重。为保证两分支输出维度一致,膨胀率为 2 的空洞卷积使用了 padding 操作。为降低特征图维度,GFN 中的卷积层步长均设置为 2。

在两层 GFN 输出后,引入一层深度可分离卷积,以提取同一时间步下相邻子载波之间的局部相关性,并完成通道融合。在此基础上,引入分组多尺度注意力机制(grouping of multi-scale attention blocks, GMAB),以增强不同子载波特征之间的交互。GMAB 首先对卷积输出的通道维度进行分组,使不同特征组分别关注不同的子空间信息。随后,在每个分组内引入多尺度特征信息,以增强特征之间的全局关联信息。假设 GMA 网络的输出

为 T_{GMA} , 则其如式 (10) 所示。

$$T_{GMA} = \text{GMAB}(\text{ReLU}(\text{BatchNomal}(W_{K=1} * X_{TG}))) \quad (10)$$

式中: $\text{GMAB}(\cdot)$ 是由通道维度嵌入、两层 GFN、深度可分离卷积以及分组多尺度注意力机制组成; X_{TG} 为式 (9) 中 GFN 的输出; $W_{K=1}$ 表示卷积核大小为 1 的深度可分离卷积权重; T_{GMA} 为 GMA 网络的最终输出特征。通过上述结构, GMA 能够有效提取到同一时间步下子载波之间的相关特征, 从而增强频域特征表达能力。

3.2 时间全局注意力网络 TGA

与 GMA 侧重提取同一时刻不同子载波的相关特征不同, TGA 主要用于提取 CSI 序列在时间维度上的动态变化信息。虽然 GMA 能够利用频域响应差异增强子载波间特征表达, 但其更关注同一时间片段内不同子载波之间的响应关系, 难以充分反映手势执行过程中的连续的时间变化规律。对于细粒度字母手势而言, 部分字母在频域响应上可能较为相近, 但其书写过程中的先后顺序、转折位置和持续时间并不完全相同。例如, ‘E’ 与 ‘F’ 在前半段书写轨迹上较为相似, 但后续横向笔画数量和持续时间存在差异, 这会引起 CSI 幅值在时间维度的连续变化。为此, 设计了 TGA 时域分支, 通过单层深度可分离卷积提取相邻时间点的局部变化, 并结合全局注意力机制突出关键时间阶段的响应变化, 从而增强模型对细粒度动态模式的判别能力。TGA 的结构如图 13 所示, 由一层深度可分离卷积及全局平均池化 (global average pooling, GAP) 和全局最大池化 (global max pooling, GMP) 构成的轻量化全局注意力机制组成。首先, 采用卷积核大小为 3 的深度可分离卷积提取相邻时间点的局部变化信息。考虑到细粒度字母手势所对应的信号变化幅度较弱, TGA 采用单层深度可分离卷积, 以避免过度平滑。在获得局部时间特征后, 引入轻量化的全局注意力机制, 对卷积输出分别进行 GAP 和 GMP 操作。其中, GAP 用于提取整体变化趋势, GMP 用于突出局部显著响应。然后, 将两者的输出特征图进行元素相加, 并通过非线性激活函数 Sigmoid 映射到 $[0, 1]$, 得到注意力权重。最后, 将该权重与卷积层的输出相乘, 以突出关键时间片段并增强时间特征的判别能力。

对于频率信息 F_j , 以子载波个数作为通道数, 以时间步长作为特征维度, 则其在 TGA 网络中的信息流通过程为:

$$\begin{aligned} F_{DS} &= \text{ReLU}(\text{BatchNomal}(W_{K=3} * F_j)) \\ F_{TGA} &= F_{DS} \cdot \text{Sigmoid}(\text{GAP}(F_{DS}) + \text{GMP}(F_{DS})) \end{aligned} \quad (11)$$

式中: F_j 为第 j 个子载波以时间步长为特征维度的时序输入; $W_{K=3}$ 表示 TGA 分支中卷积核大小为 3 的深度可分离卷积权重; F_{DS} 是卷积层的输出; F_{TGA} 是频率信息 F_j

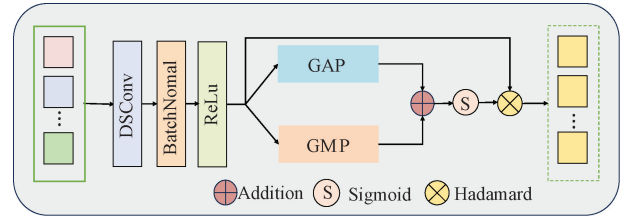


图 13 TGA 网络结构

Fig. 13 Structure of the TGA network

经过 TGA 网络的输出。在得到 GMA 和 TGA 两个网络的输出 T_{GMA} 和 F_{TGA} 后, 首先对两分支特征进行重构与匹配, 再在通道维度上将 T_{GMA} 和 F_{TGA} 在进行级联, 以融合子载波相关特征和时间动态特征。最后, 采用平均池化 (average pooling, AP) 对其进行降维, 以减少特征图尺寸和模型复杂度。设特征提取模块的输出为 X , 则上述过程如式 (12) 所示。

$$X = \text{AP}(\text{Concat}(T_{GMA}, F_{TGA})) \quad (12)$$

式中: $\text{AP}(\cdot)$ 表示平均池化的操作; $\text{Concat}(\cdot)$ 表示级联操作; T_{GMA} 表示频率分支输出; F_{TGA} 表示时间分支输出; X 表示融合后的输出结果。融合后的特征图同时包含子载波之间的相关结构信息以及时间动态信息, 为后续识别分类模块提供更有效的输入表示。

3.3 树突网络 DD

在完成 GMA 与 TGA 双分支特征提取及融合后, 所得特征同时包含 GMA 分支提取的频域关联信息和 TGA 分支提取的时域动态信息。由于二者关注的信息不同, 因此分类阶段需要进一步加强时域与频域特征之间的有效交互。传统 MLP 通常采用统一权重矩阵对输入特征进行全连接映射, 虽然具有非线性表达能力, 但对不同特征组合的处理方式相对一致, 难以突出时域特征和频域特征之间的差异性关联。相比之下, DD 网络采用树突状多分枝结构, 将不同特征组合映射到独立分支中进行处理, 并通过 Hadamard 乘积实现特征间的选择性交互, 使模型更加关注时频融合特征中的有效组合关系。在保留原始融合特征信息的基础上, DD 网络进一步增强了时域与频域特征之间的交互能力, 从而提高融合特征的判别性。此外, DD 网络仅由输入投影层、树突模块和线性分类层组成, 网络结构较为紧凑, 其主要计算开销来源于矩阵乘法运算, 未引入额外的卷积和注意力计算模块, 因此不会显著增加整体计算复杂度。相比传统 MLP, DD 网络在保持较低复杂度的同时, 能够实现特征选择性交互, 增强时域和频域之间的关联性, 更适用于资源受限场景^[28]。因此, 本文采用 DD 网络作为最终分类模块, 以提升时频融合特征的利用效率和分类性能。该网络如式 (13) 所示。

$$A^l = W^{l,l-1} A^{l-1} \circ X \quad (13)$$

式中: $\mathbf{A}^{l-1}$ 和 $\mathbf{A}^l$ 是模块的输入和输出, 特别是 $\mathbf{A}^0 = \mathbf{X}$, $\mathbf{X}$ 是 DD 的原始输入; $\mathbf{W}^{l,l-1}$ 是第 $l-1$ 个 DD 模块到第 l 个模块的权重矩阵, $l = 1, 2, \dots$ 是层序号; $\circ$ 表示的是 Hadamard 相乘。

4 实验验证与性能评估

为验证 DFANet 模型的有效性、鲁棒性和泛化能力, 设置了公开数据集验证、自采数据集消融实验和多因素条件测试 3 类实验。首先, 在开源数据集 UT-HAR^[29] 上与多种基准模型进行比较, 以验证所提模型的有效性与轻量化优势。其次, 在实验室、办公室和教室 3 种环境下采集的 ESP-HAR 数据集上开展消融实验, 以分析各组成模块对整体性能的必要性。最后, 在 ESP-HAR 数据集上通过基准模型对比以及不同环境、距离、角度和人员条件下的测试, 来验证模型的鲁棒性和泛化能力。

4.1 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。文中所有模型均采用 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策略, 将数据按 8:2 的比例划分为训练集和测试集。为保证实验结果的可重复性, 各对比模型均在相同硬件环境、训练轮次等条件下完成训练与测试。

4.2 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 用以衡量评估模型对正类预测的可靠性, 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正例的识别能力, FLOPs 用以评估计算复杂度, Params 用以衡量模型规模并从参数量的角度反映模型轻量化程度。评价指标的数学表达式如式 (14) 所示。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Pr = \frac{TP}{TP + FP}$$
$$Re = \frac{TP}{TP + FN}$$
$$F1 = 2 \times \frac{Re \times Pr}{Re + Pr}$$

(14)

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。

4.3 模型的有效验证

为了验证 DFANet 模型的有效性, 将所提模型与基准模型在开源数据集 UT-HAR 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 种不同活动的数据。每对天线记录 30 个子载波数据, 最终形成 90 维特征向量。表 2 表明, DFANet 在 UT-HAR 数据集上取得 98.82% 的分类准确率, 虽然略低于 ResNet101 但其参数量和计算复杂度显著降低, 体现出较好的轻量化优势。

表 2 在 UT-HAR 数据集上不同深度神经网络上的实验结果

Table 2 Experimental results of different DNNs on the UT-HAR dataset

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	23.549×10^6
ResNet101	99.56	99.27	99.22	99.24	4.762×10^9	42.568×10^6
MASNet	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
DFANet	98.82	98.21	98.05	98.12	159.705×10^6	149.888×10^3

实验结果表明, 尽管 DFANet 在准确率上略低于 ResNet101 0.74%, 但其参数量和计算复杂度显著降低, 说明所提模型在保持较高识别性能的同时兼顾了轻量化需求, 从而验证了所提方法的有效性。

4.4 分阶段预处理消融实验

为评估数据预处理模块的有效性, 基于 ESP-HGR 数

据集设计分阶段预处理的消融实验, 实验结果如表 3 所示。本文在原始数据基础上逐步加入 Hampel 异常值剔除、中值滤波和小波降噪, 并在相同模型结构和训练设置下进行性能对比, 为保证对比实验的一致性, 所有实验组均采用三次样条插值将 CSI 序列长度统一为 135。在此基础上对比不同预处理组合下模型的识别准确率。实验

结果表明。仅采用 Hampel 异常值剔除时模型准确率为 88.36%;再加入中值滤波后准确率提升至 90.89%;最后加入小波降噪模型准确率达到 91.21%。这说明分阶段预处理流程能够有效抑制异常点和噪声干扰,提高 CSI 序列的一致性和稳定性,从而改善模型的识别性能。

表 3 数据预处理的消融实验

Table 3 Ablation results of data preprocessing			
Hampel 异常值剔除	中值滤波	小波降噪	ACC/%
/	/	/	85.37±0.61
✓	/	/	88.36±0.56
✓	✓	/	90.89±0.71
✓	✓	✓	91.21±0.64

4.5 DFANet 消融实验

为验证 DFANet 模型中各个模块的必要性,在融合 3 种采集环境数据的 ESP-HAR 数据集上进行消融实验。实验结果如表 4 所示,当去除 TGA、GMA、GMAB 或 DD 后模型性能均出现下降,说明各模块对细粒度手势识别均具有积极作用。其中,去除 GMA 后模型准确率由 91.21% 下降至 63.78%,FLOPs 由 49.534×10^6 降至 517.828×10^3 。这是因为 GMA 分支包含通道嵌入、两层 GFN、多尺度卷积及 GMAB 等结构,是 DFANet 中主要的计算量来源;而 TGA 分支仅由单层深度可分离卷积和轻量级全局注意力机制组成,计算复杂度较低。因此,去除 GMA 后模型计算复杂度明显降低,同时准确率大幅下降,说明 GMA 对频域子载波相关特征提取和整体识别性能具有关键作用。线性层代替 DD 后准确率由 91.21% 降至 84.29%,说明简单的线性映射难以处理时频融合特征的复杂非线性关系。去除 GMAB 后,模型准确率降至 89.29%,表明 GMAB 有助于增强频域特征表达能力;去除 TGA 的注意力机制后,准确率降至 90.52%,说明注意力机制有助于增强时域特征表达能力。此外,为进一步分析 DD 网络在分类阶段的作用,本文分别采用线性层和 MLP 替代 DD 网络进行对比试验。当采用线性层代替 DD 网络时,模型准确率由 91.21% 降至 84.29%,说明

简单线映射难以有效处理时频融合特征中的复杂非线性关系。当采用 MLP 代替 DD 网络时,模型准确率为 88.64%,表明多层非线性映射相比单层线性层能够更有效地处理融合特征,但仍低于完整 DFANet 的 91.21%。相比普通 MLP,DD 网络通过树突状多分枝结构提升了特征组合与非线性表达能力,有助于挖掘时域和频域融合特征中的关键判别信息。因此,虽然 DD 网络相比 MLP 带来了一定参数数量的增加,但 FLOPs 仅有 49.461×10^6 增至 49.534×10^6 ,对整体计算复杂度影响较小,同时有效提升了识别准确率。

表 4 DFANet 模型的消融实验

Table 4 Ablation study of the DFANet model			
模型方法	ACC/%	FLOPs	Params
去除 TGA	85.69±1.37	48.967×10^6	82.072×10^3
去除 GMA	63.78±2.76	517.828×10^3	16.656×10^3
去除 GMAB	89.29±0.73	44.817×10^6	142.120×10^3
去除 TGA 中的注意力机制	90.52±0.89	49.534×10^6	147.880×10^3
线性层代替 DD 网络	84.29±0.95	49.424×10^6	37.298×10^3
MLP 代替 DD 网络	88.64±0.51	49.461×10^6	74.738×10^3
DFANet	91.21±0.64	49.534×10^6	147.880×10^3

4.6 DFANet 模型的鲁棒性和泛化性验证

在公开数据集验证 DFANet 模型的有效性之后,进一步在 ESP-HAR 数据集上通过模型对比及多因素测试评估 DFANet 的综合性能、鲁棒性和泛化能力。

1) 综合性能比较

为进一步验证 DFANet 模型的鲁棒性和泛化能力,将其与 ResNet18、ResNet50、ResNet101、ABLSTM^[30]、CSI-DeepNet^[14]、DSC-SGRU^[31] 和 D-GMAFi 等基准模型在自采数据集 ESP-HAR 上进行综合实验。实验结果如表 5 所示,DFANet 在 ESP-HAR 数据集上取得最高平均准确率 91.21%,且参数数量和计算复杂度分别为 147.880×10^3 和 49.534×10^6 ,在所有模型中属于较低水准并实现了最优的性能。

表 5 在 ESP-HAR 数据集上不同深度神经网络上的实验结果

Table 5 Experimental results of different DNNs on the ESP-HAR dataset						
模型方法	ACC/%	Pr/%	Re/%	F1/%	FLOPs	Params
ResNet18	88.78±3.31	89.03±2.94	88.89±2.98	88.88±3.01	1.639×10^9	41.363×10^6
ResNet50	89.13±3.68	88.97±3.58	88.89±3.70	88.84±3.75	835.781×10^6	21.291×10^6
ResNet101	88.77±4.58	88.92±4.13	88.73±4.32	88.76±4.27	465.568×10^6	11.83×10^6
ABLSTM	89.61±0.99	89.66±0.78	89.64±0.72	89.62±0.75	83.687×10^6	649.355×10^3
CSI-DeepNet	84.28±1.24	84.11±1.20	83.99±1.19	83.98±1.18	169.514×10^6	188.956×10^3
DSC-SGRU	87.37±1.64	87.41±1.46	87.28±1.59	87.27±1.52	17.621×10^6	78.830×10^3
D-GMAFi	88.91±1.50	88.94±1.23	88.89±1.23	88.88±1.24	93.20×10^6	454.834×10^3
DFANet	91.21±0.64	91.11±0.72	90.97±0.66	91.12±0.63	49.534×10^6	147.880×10^3

为进一步分析 8 种模型在不同活动识别任务中的分类性能,绘制了其对应的混淆矩阵。如图 14 所示,

DFANet 在所有类别上均实现了较高准确率,有的手势达到了 95%以上的识别准确率,且非对角线误判比例相对

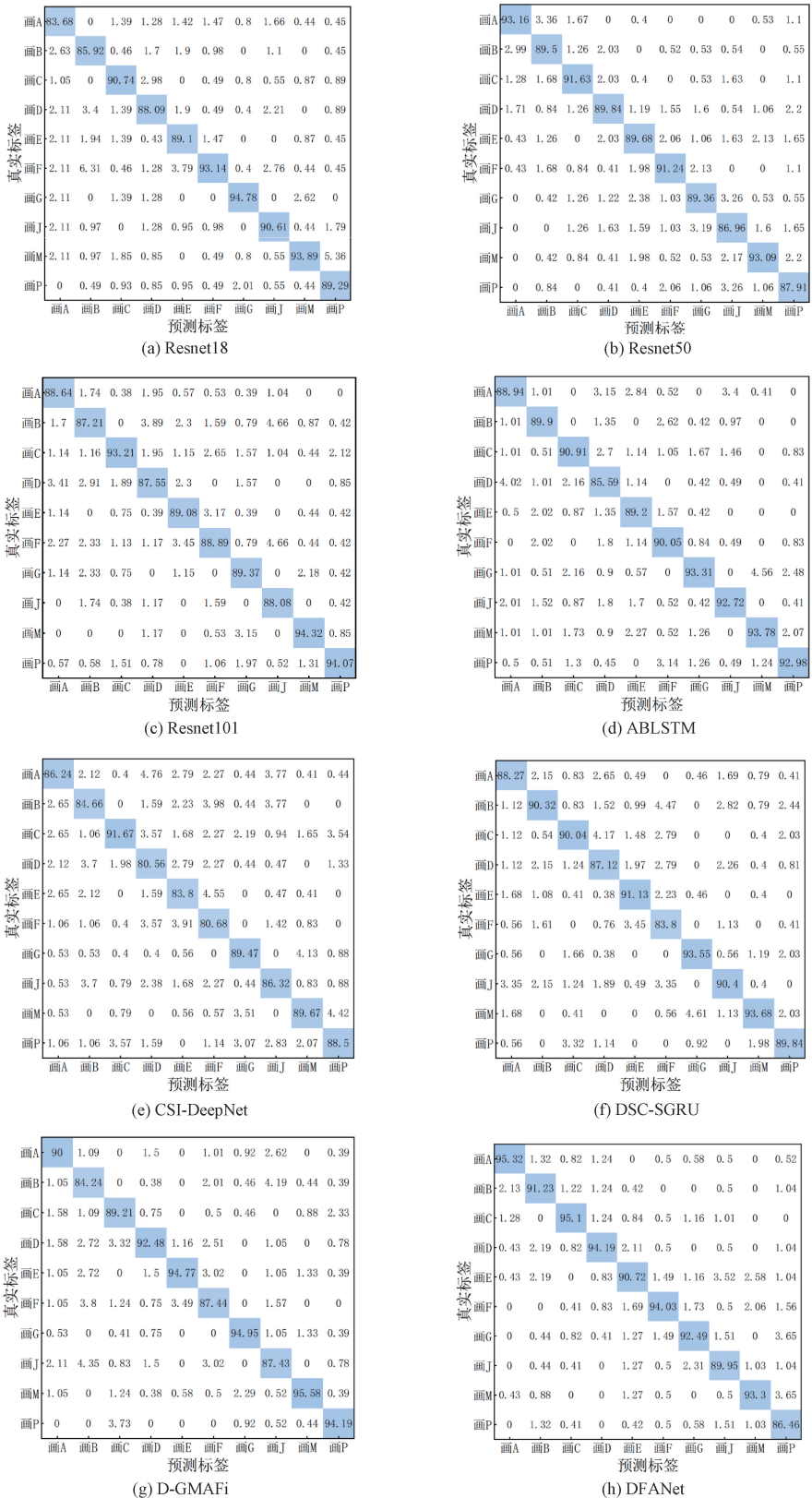


图 14 所有模型在 ESP-HAR 数据集上的混淆矩阵

Fig. 14 Confusion matrices of all models on the ESP-HAR dataset

较低。然而,有部分字母识别仍会出现偏差,例如字母‘P’被识别为‘M’,字母‘B’被识别为‘D’和字母‘E’被识别为‘F’不能够被准确区分。其原因大多在于个人手写不规范或字形相似,但 DFANet 以 91.21% 的平均准确率,已足以满足基础识别分类的需求,并在该分类任务中展现了最优的综合性能。

在数据集 ESP-HAR 上,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,总体来看,多数模型在前 50~100 个训练轮次内快速收敛,但不同模型的训练稳定性存在差异。相比之下,DFANet 模型收敛较快,后期训练过程较为稳定;其损失值最终稳定在 1 附近,这是由于所采用的自适应交叉熵损失函数导致的。该策略有助于缓解模型的过拟合问题并提升模型的泛化能力。

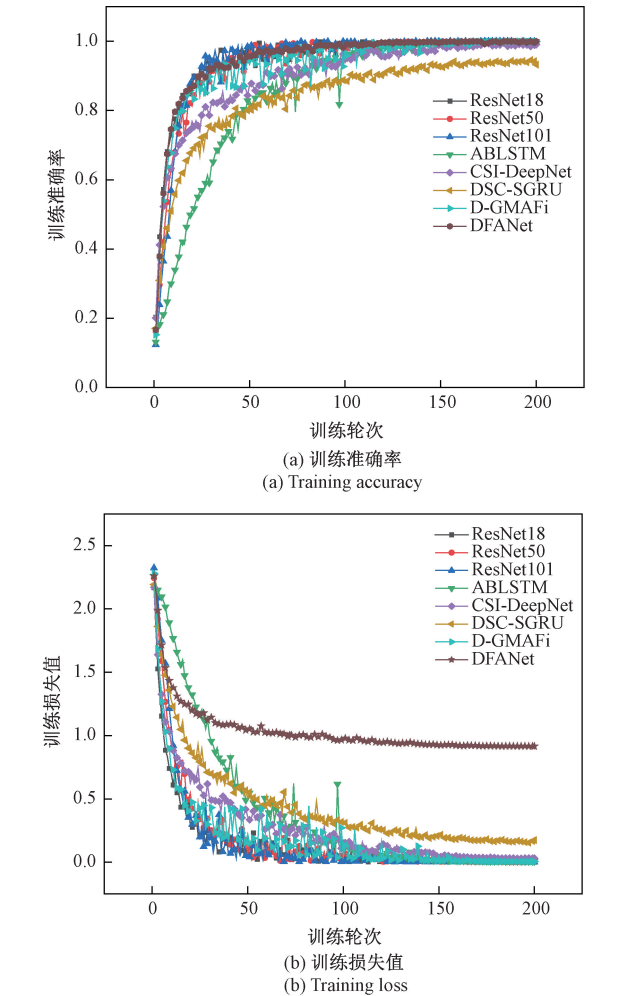


图 15 DFANet 模型的训练准确率和训练损失值曲线
Fig. 15 Training accuracy and training loss curves of the DFANet model

为验证所提方法能够有效地区分不同活动特征,采用 t-SNE 方法对原始数据分布和分类器学习到的特征分

布进行了可视化,如图 16 所示。结果表明,原始特征分布较为复杂,而经 DFANet 提取后的特征形成了更清晰的类间边界,证明了该模型能够区分不同活动类别的深层特征。

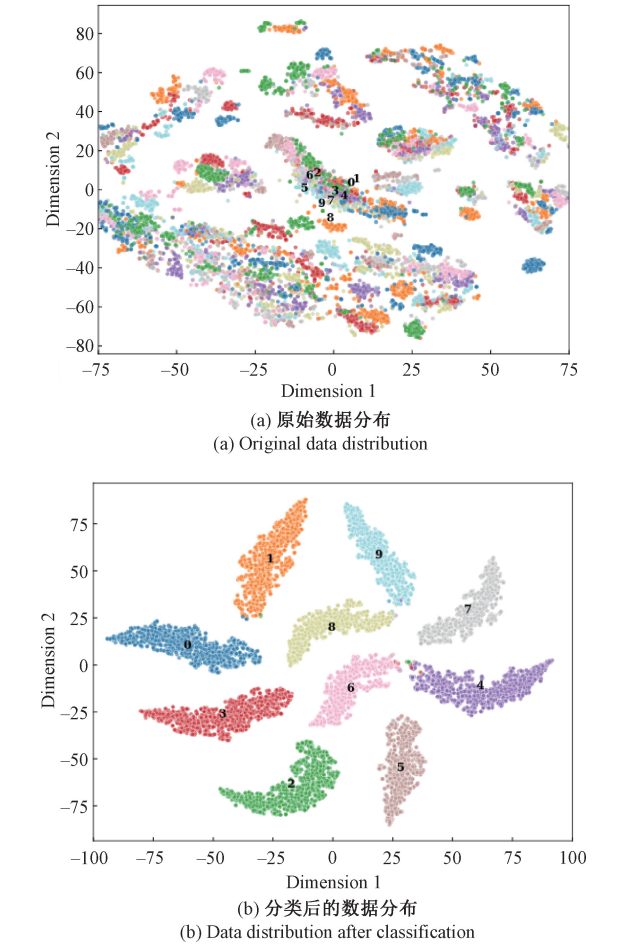


图 16 t-SNE 可视化数据分类前后的分布
Fig. 16 t-SNE data distribution visualization before and after classification

2) 不同条件下的鲁棒性与泛化分析

(1) 不同环境的影响

在控制其余实验条件一致的情况下,对不同环境进行测试,结果如图 17(a) 所示。实验室、办公室和教室的平均准确率分别为 88.51%、92.64% 和 91.60%。其中,实验室的准确率低于其他两种场景,主要原因在于实验室中各类仪器较多,对无线信号传播产生了较强干扰。

(2) 不同距离的影响

为探究发射和接收设备之间的距离对手势活动感知的影响,选择了 1.5、2.0 和 2.5 m 3 种距离进行验证。结果如图 17(b) 所示,设备间距过远或者过近都会影响模型的感知准确率。其中 2.0 m 时整体准确率最高。原

因在于该距离下信号衰减适中、干扰较小更有利于手势特征的稳定感知。

(3) 不同角度的影响

为探究用户手势运动方向对模型性能的影响,以设备 LOS 方向中心点为基准,选取 30°、60°、90°、120° 和 150° 5 个角度展开实验,结果如图 17(c) 所示。模型在 90° 时整体识别准确率最高为 93.37%;而在 30° 和 150° 的时候准确率明显下降。说明手势运动方向会影响信号传播变化路径,其中垂直于 LOS 方向时更有利于模型识别。

(4) 不同人员的影响

为验证模型对不同用户的泛化能力,邀请了 5 名志愿者参加实验,结果如图 17(d) 所示。DFANet 模型对 5 名志愿者的平均识别准确率均超过 83%,其中有 3 名志愿者的准确率超过 92%。结果表明 DFANet 在不同用户条件下具有较好的识别稳定性和一定的泛化能力。

综合不同条件下的实验结果可知,所提 DFANet 在多种实际测试条件下均保持较稳定的识别性能,具有较好的环境适应能力和跨用户泛化能力。

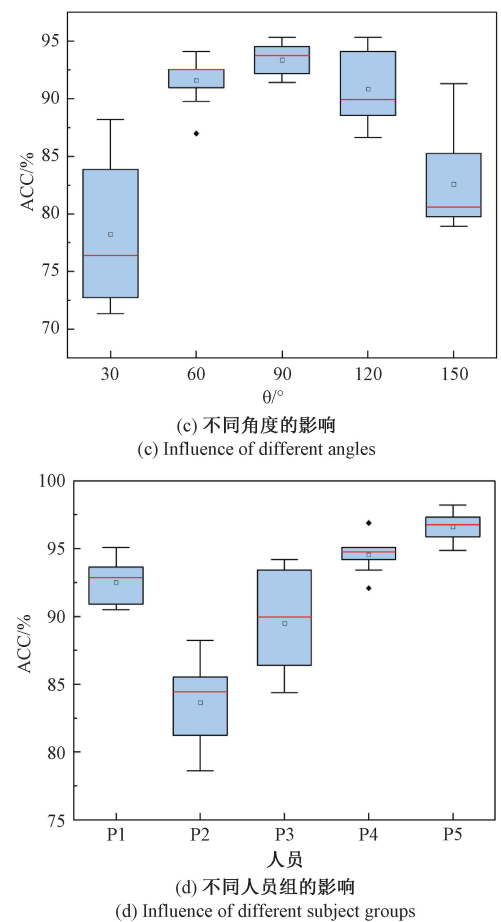
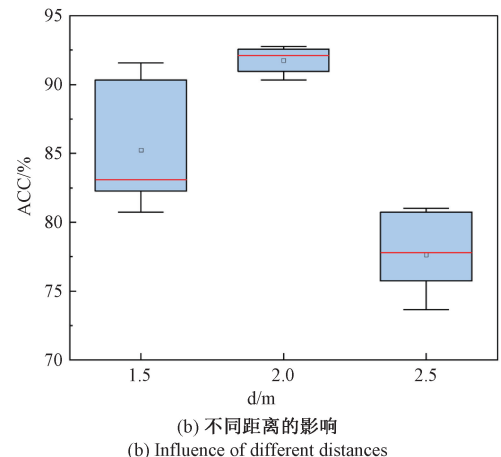
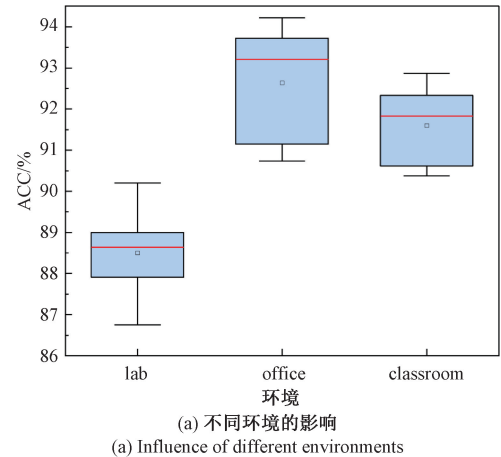


图 17 不同因素对 DFANet 模型的影响
Fig. 17 Influence of different factors on the DFANet model

5 结 论

本文围绕 ESP32 单天线条件下的细粒度字母手势识别问题,完成了 CSI 数据采集系统搭建、预处理流程设计以及轻量化识别模型构建。通过融合时间动态特征与子载波相关特征,所提 DFANet 在多场景实验中表现出较好的识别性能,并兼顾了较低的参数量与计算复杂度。研究表明,在资源受限的低功耗 WiFi 感知设备上,基于时频协同特征提取的轻量化方法能够有效提升细粒度手势识别能力。该研究为单天线嵌入式 WiFi 感知设备上的手势交互应用提供了方法参考。当前工作主要通过参数量与计算复杂度验证 DFANet 的轻量化特性,模型参数量为 147.88×10^3 ,计算量为 49.534×10^6 FLOPs。实验结果表明,在保证识别性能的同时,DFANet 具有较低的参数规模和计算开销,为资源受限场景下的细粒度手势识别提供了一种轻量化解决方案。未来将进一步结合动态轨迹信息、多模态特征融合等策略,持续提升复杂细粒度手势的识别精度与场景适应能力。

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 Ying, Yang Zhijia, Xie Chuang, et al. Design of HAR data acquisition system based on inertial sensor[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 Wenpei, Yang Bo, Min Linzi. Human behavior data augmentation for the low-resolution infrared perception system[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 163-171.
- [3] 刘鑫. 基于虚拟现实技术的人机交互三维手势识别研究[J]. 现代电子技术, 2025, 48(20): 40-44.
Liu Xin. Research on human-computer interaction 3D gesture recognition based on virtual reality technology[J]. Modern Electronics Technique, 2025, 48(20): 40-44.
- [4] Kim K M, Choi J I. Passengers' gesture recognition model in self-driving vehicles: Gesture recognition model of the passengers' obstruction of the vision of the driver[C]//Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS). Singapore: IEEE, 2019: 239-242.
- [5] 张应辉, 陈博文, 曹进, 等. 具有隐私保护的细粒度智能家居远程数据安全更新方案[J]. 电子与信息学报, 2023, 45(3): 810-818.
Zhang Yinghui, Chen Bowen, Cao Jin, et al. Fine-grained remote data security update scheme for smart home with privacy protection[J]. Journal of Electronics & Information Technology, 2023, 45(3): 810-818.
- [6] 张鋆豪, 何百岳, 杨旭升, 等. 基于可穿戴式惯性传感器的人体运动跟踪方法综述[J]. 自动化学报, 2019, 45(8): 1439-1454.
Zhang Yunhao, He Baiyue, Yang Xusheng, et al. A review on wearable inertial sensor-based human motion tracking[J]. Acta Automatica Sinica, 2019, 45(8): 1439-1454.
- [7] Varshini A, Bhavani G, Vithya M, et al. Real-time hand gesture recognition for robotic arm and home automation[C]//Proceedings of the 2021 International Symposium on Electrical, Electronics and Information Engineering. New York: Association for Computing Machinery, 2021: 218-223.
- [8] Wu C S, Yang Z, Zhou Z M, et al. Non-invasive detection of moving and stationary human with WiFi[J]. IEEE Journal on Selected Areas in Communications, 2015, 33(11): 2329-2342.
- [9] 郭梦丽, 许勇, 何昕, 等. 基于 PHY 层的手势身份识别技术研究[J]. 计算机与数字工程, 2022, 50(10): 2252-2258.
Guo Mengli, Xu Yong, He Xin, et al. Research on gesture identity recognition technology based on PHY layer[J]. Computer & Digital Engineering, 2022, 50(10): 2252-2258.
- [10] Ali H A H, Seytnazarov S. Human walking direction detection using wireless signals, machine and deep learning algorithms[J]. Sensors, 2023, 23(24): 9726.
- [11] Zhang Y, Xu K L, Wang Y J. WiNum: A WiFi finger gesture recognition system based on CSI [C]//Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City. New York: Association for Computing Machinery, 2019: 181-186.
- [12] Bastwesy M R M, ElShennawy N M, Saidahmed M T F. Deep learning sign language recognition system based on Wi-Fi CSI[J]. International Journal of Intelligent Systems and Applications, 2020, 12(6): 33-45.
- [13] Jiang D H, Li M Q, Xu C L. WiGAN: A WiFi-based gesture recognition system with GANs [J]. Sensors, 2020, 20(17): 4757.
- [14] 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.
- [15] Kocheta P, Bhatia N S, Obraczka K. Pulse-Fi: A low-cost system for accurate heart rate monitoring using Wi-Fi channel state information [C]//2025 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT). IEEE, 2025: 226-230.
- [16] Macedo D, Loureiro M, Martins Ó G, et al. From CSI to coordinates: An IoT-driven testbed for individual indoor localization[J]. Future Internet, 2025, 17(9): 395.
- [17] Das Chagas A O S, Lovisolo L, Tcheou M P, et al. A proposal for fall-risk detection of hospitalized patients from wireless channel state information using Internet of Things devices[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108628.
- [18] Zhao C J, Wang J, Huang W, et al. Communication under mixed Gaussian-impulsive channel: An end-to-end framework [C]//2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN). IEEE, 2024: 119-125.
- [19] Hussain A, Chen Y S, Ullah A, et al. WiSigPro:

- Transformer for elevating CSI-based human activity recognition through attention mechanisms [J]. Expert Systems with Applications, 2024, 258: 124976.
- [20] Wang Z J, Jiang K K, Hou Y S, et al. A survey on human behavior recognition using channel state information [J]. IEEE Access, 2019, 7: 155986-156024.
- [21] Shi Z G, Cheng Q Q, Zhang J A, et al. Environment-robust WiFi-based human activity recognition using enhanced CSI and deep learning [J]. IEEE Internet of Things Journal, 2022, 9(24): 24643-24654.
- [22] He Y, Chen Y, Hu Y, et al. WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi [J]. IEEE Internet of Things Journal, 2020, 7(9): 8296-8317.
- [23] Kim Y, Jeon W S, Jeong D G. AmFall: WiFi CSI amplitude-based fall detection using denoised scalograms [J]. IEEE Internet of Things Journal, 2025, 12(18): 37988-38003.
- [24] Omidvar M, Zahedi A, Bakhshi H. EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers [J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(11): 10395-10403.
- [25] Kim H, Moon N. TN-GAN-based pet behavior prediction through multiple-dimension time-series augmentation [J]. Sensors, 2023, 23(8): 4157.
- [26] Meng W, Liu Z C, Li B, et al. GrapHAR: A lightweight human activity recognition model by exploring the sub-carrier correlations [J]. IEEE Transactions on Wireless Communications, 2024, 23(4): 2755-2770.
- [27] Tay Y, Luu A T, Zhang A, et al. Compositional De-Attention Networks [C]//Advances in Neural Information Processing Systems 32. Red Hook: Curran Associates, Inc., 2019: 6132-6142.
- [28] Liu G, Wang J. Dendrite net: A white-box module for classification, regression, and system identification [J]. IEEE Transactions on Cybernetics, 2022, 52(12): 13774-13787.
- [29] 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.
- [30] Chen Z H, Zhang L, Jiang C Y, et al. WiFi CSI-based passive human activity recognition using attention-based BLSTM [J]. IEEE Transactions on Mobile Computing, 2019, 18(11): 2714-2724.
- [31] 何育浪, 赵志彪, 李振, 等. 基于 DSC-SGRU 模型的 Wi-Fi 手势识别系统研究 [J]. 电子测量与仪器学报, 2024, 38(10): 97-108.
- He Yulang, Zhao Zhibiao, Li Zhen, et al. Research on Wi-Fi gesture recognition system based on DSC-SGRU model [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 97-108.

作者简介



崔凌云, 2024 年于天津职业技术师范大学获得学士学位, 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: 2934352248@qq.com

Cui Lingyun received her B. Sc. degree from Tianjin University of Technology and Education in 2024. She is now a M. Sc. candidate at Tianjin University of Technology and Education. Her 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 now a lecturer at Tianjin University of Technology and Education. His main research interests include industrial intelligent control and machine learning.