



研究与开发

基于移动通信网信号的人体动作识别研究

钱能¹, 陆城灵², 万雨轩², 吴哲夫²

(1. 浙江邮电职业技术学院电子与通信工程学院, 浙江 杭州 312366;

2. 浙江工业大学网络空间安全研究院, 浙江 杭州 310014)

摘要: Wi-Fi 信道状态信息 (channel state information, CSI) 的人体活动识别 (human activity recognition, HAR) 方法在动作监测等领域被广泛应用, 但仍面临部署成本高和感知范围有限等挑战。为解决这些问题, 提出了一种移动通信网长期演进的注意力卷积神经网络 (long term evolution attention-guided ConvNeXt, LTE-ACN), 利用移动信号进行活动识别。首先, 基于小区参考信号提取信道状态信息有效特征; 其次, 进行噪声滤波、Savitzky-Golay 平滑、峰谷增强和格拉姆角场变换等信号处理来构建动作数据集; 最后, 将增强后的图像数据输入 LTE-ACN 模型, 通过改进的 ConvNeXt 架构有效地降低了特征信息损失, 同时引入卷积块注意力模块 (convolutional block attention module, CBAM) 提升关键特征的表达能力, 强化了空间域内的特征关联性。实验结果表明, 所提方法的 6 类动作识别平均准确率达到了 96.44%, 验证了基于 LTE 信号进行人体动作识别的可行性。

关键词: 移动通信网信号; 信道状态信息; 人体活动识别; 注意力机制; 卷积神经网络

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2026015

Research on human action recognition based on mobile communication network signals

Qian Neng¹, Lu Chengling², Wan Yuxuan², Wu Zhefu²

1. School of Electronics and Communication Engineering, Zhejiang Post and Telecommunication College, Hangzhou 312366, China

2. Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310014, China

Abstract: Wi-Fi channel state information (CSI)-driven human activity recognition (HAR) has been widely studied for applications in activity monitoring. However, challenges such as high deployment costs and limited sensing range remain. To address these, long term evolution attention-guided ConvNeXt (LTE-ACN) was proposed, which used mo-

收稿日期: 2025-05-18; 修回日期: 2025-09-29

通信作者: 吴哲夫, wzf@zjut.edu.cn

基金项目: 浙江省自然科学基金重点项目 (No.LZ22F010005)

Foundation Item: The Key Foundation of Zhejiang Natural Science (No.LZ22F010005)

bile communication LTE signals for activity recognition. Firstly, effective features of channel state information were extracted based on cell reference signals. Then, signal processing methods including noise filtering, Savitzky-Golay smoothing, peak-valley enhancement, and Gramian angular field transformation were applied to construct the motion dataset. Finally, the enhanced image data was input into the LTE-ACN model, in which the improved ConvNeXt architecture effectively reduced the feature information loss while the incorporated convolutional block attention module (CBAM) attention mechanism enhanced the expressive capability of key features and strengthened the feature correlations in the spatial domain. Experimental results demonstrate that the proposed method achieves an average accuracy of 96.44% in the six actions recognition, verifying the feasibility of LTE signal-based human motion recognition.

Key words: mobile communication network signal, CSI, HAR, attention mechanism, convolutional neural network

0 引言

随着传感技术的进步与智能应用需求的持续增长,人体动作识别技术已逐渐走向成熟。当前,该技术不仅要满足便捷性与隐私保护的要求,还须具备易部署、低成本等工业化优势。

人体动作识别技术按感知方法可分为接触式和非接触式两类:接触式感知基于专用传感器设计,用户须穿戴声学、加速度、压力等设备来采集运动数据以实现识别^[1-4],虽然数据采集相对容易且受环境影响较小,但长时间佩戴给用户带来不便;非接触式传感无须用户佩戴设备,便利性大增,主要包括基于视频图像和无线射频(radio frequency, RF)感知处理。其中,基于视频图像的技术通过摄像头采集动作帧,经处理与机器学习实现识别,采集便捷、准确率高^[5-7],但存在计算资源需求大、易受环境干扰、有隐私泄露风险等局限;基于RF的技术通过无线信号安全、非侵入地感知人员,避免了隐私侵犯和视觉障碍等问题。

无线射频感知技术包括射频识别(radio frequency identification, RFID)、超宽带(ultra wide band, UWB)、毫米波和Wi-Fi等。其中,RFID通过追踪贴附标签实现动作检测,但应用范围受限^[8];UWB可非接触监测呼吸与心跳,但依赖专用标签设备^[9-10];毫米波技术感知精度高,能快速识别人体动作^[11],然而设备成本高昂,难以普及。相比之下,Wi-Fi感知技术凭借泛在性

与强穿透性成为非接触动作识别的研究热点,不需要额外设备,成本低且易推广^[12-13]。然而,现有的基于Wi-Fi的人体动作识别仍存在感知范围小、部署成本高、识别准确率受干扰影响等缺陷。

值得关注的是,商用长期演进(long term evolution, LTE)信号在感知领域展现出独特的优势。与Wi-Fi相比,其覆盖范围更广、硬件部署成本更低。Feng等^[14]利用LTE传感完成了室内呼吸传感和室外交通监控;Sonny等^[15]提出了基于LTE信道状态信息(channel state information, CSI)和卷积神经网络(convolutional neural network, CNN)检测室外停车位占用状态的方法。此外,商用LTE信号传输速率与子载波固定,无信道争用问题,能确保感知系统稳定运行;且仅须采集基站的下行链路信号,无须部署发送端,可大幅降低硬件成本。基于此,本文提出了一种基于商用LTE信号的人体动作识别新方法,旨在实现大范围、低成本且稳定的人体动作感知。

1 系统原理

人体动作识别系统框图如图1所示,基于商用LTE信号的人体动作识别系统流程包含3个主要组成部分:CSI数据采集、数据预处理以及移动通信网的注意力卷积神经网络(LTE attention-guided ConvNeXt, LTE-ACN)模型训练与识别。其中,CSI数据采集环节由软件定义无线电USRP B210完成LTE信号的接收。

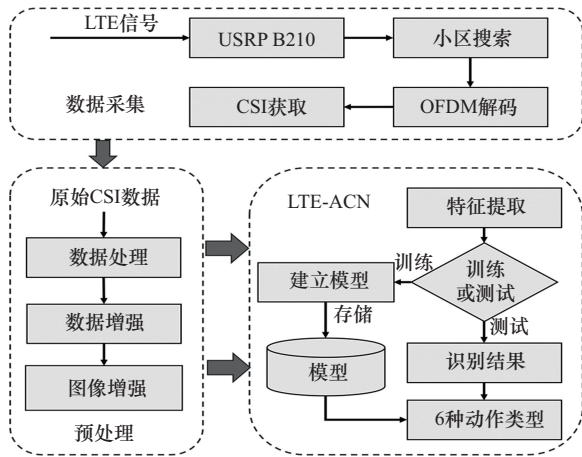


图1 人体动作识别系统框图

1.1 LTE 基础

本文主要利用LTE下行链路发送的信号进行人体动作的感知识别，LTE在下行链路上使用正交频分复用（orthogonal frequency division multiplexing, OFDM）技术来调制传输数据。从时域的角度来说，LTE基站持续传输周期为10 ms的无线电帧（frame）。系统帧格式的时域表示如图2所示。每个帧包含持续时间为1 ms的10个子帧（sub-frame），每个子帧包含两个持续时间为0.5 ms的时隙（slot）。每个时隙又由若干OFDM符号构成，其基本时长 T_u （2048点FFT配置）为66.67 μ s，系统采样间隔 T_s 为其时域基准，时隙内OFDM符号个数取决于基站的配置。对于正常的循环前缀（normal cyclic prefix, Normal CP），每个时隙包括7个OFDM符号；对于扩展的循环前缀（extended cyclic prefix, Extended CP），每个时隙包括6个OFDM符号。

LTE的时频资源网格如图3所示。网格中的资源元素（resource element, RE）可由 $RE(k, l)$ 唯一标识，其中，频率索引 k ($k=0, 1, \dots, N_{RB}^{DL} \times N_{SC}^{RB} - 1$)由下行链路资源块总数 N_{RB}^{DL} 和每资源块子载波数 N_{SC}^{RB} 共同确定；符号索引 l ($l=0, 1, \dots, N_{symbol}^{DL} - 1$)由每个时隙的OFDM符号数 N_{symbol}^{DL} 确定。从时域看，每个slot中第1个和第5个OFDM符号用于发送小区参考信号（cell reference signal, CRS）。CRS是预定义符号序列，LTE用时频资源网格中特定位

置的RE传输。从频域看，LTE无线电资源以资源块（resource block, RB）为单位调度，由频域中的 $N_{SC}^{RB}=12$ 个子载波组成，在时域持续一个slot（0.5 ms），在下行链路中（从基站到用户），根据带宽配置，基站在每个slot中可以有 $N_{RB}^{DL} \in \{25, 50, 75, 100\}$ 个RB。LTE基站在所有下行链路RB中发送CRS，而无论RB是否被调度用于数据传输。在每个RB中的4个不同位置发送CRS，分别在两个预定义符号上各有两个，且两两相隔6个子载波。因此，CRS以固定的时间和频率间隔形成密集的时频网格。以20 MHz带宽配置的LTE基站为例，因为包含100个RB，所以对于某一个包含CRS的OFDM符号来说，LTE基站的下行链路中将有200条子载波专门用于传输CRS。

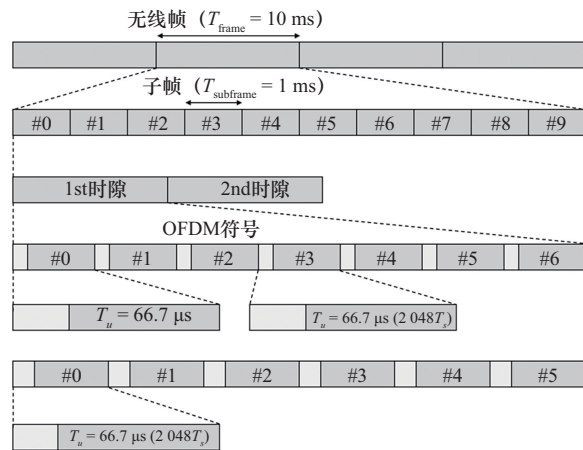


图2 系统帧格式的时域表示

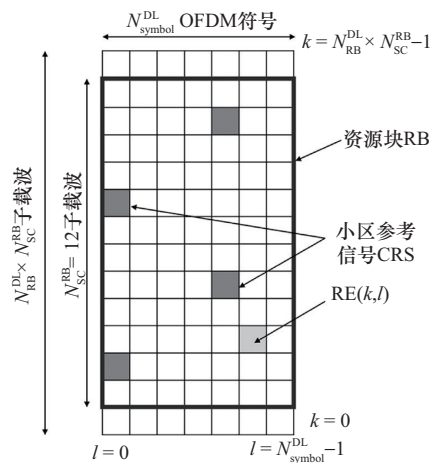


图3 LTE的时频资源网格

1.2 CRS相关知识

CRS 由一系列特定的参考符号构成，参考符号 $r_{l,n_s}(m)$ 可以通过式 (1) 来生成。

$$r_{l,n_s}(m) = \frac{1}{\sqrt{2}}(1 - 2 \cdot c(2m)) + j \frac{1}{\sqrt{2}}(1 - 2 \cdot c(2m + 1)),$$

$$m = 0, 1, \dots, 2N_{RB}^{\max,DL} - 1 \quad (1)$$

其中， n_s 表示一帧 LTE 信号中 slot 的编号，取值范围为 0~19， l 表示 slot 中 OFDM 符号的编号，根据 OFDM 循环前缀的类型，取值范围为 0~6 或 0~5， $N_{RB}^{\max,DL}$ 表示下行链路中包含的 RB 的最大数目，值为 110。另外，因为每个资源块从频域的角度来说，包含了 12 条子载波，对于包含 CRS 的 OFDM 符号，每个资源块中的 2 条子载波将用于发送 CRS，所以，对于整个下行链路，CRS 序列的最大数目为 220。此外， $c(\cdot)$ 是伪随机码序列，可以根据式 (2) 生成。

$$c(k) = (x_1(k + N_c) + x_2(k + N_c)) \% 2,$$

$$N_c = 1600, k = 0, 1, \dots, 4N_{RB}^{\max,DL} - 1 \quad (2)$$

其中， N_c 是伪随机序列的移位寄存器长度参数。

$$x_1(t + 31) = (x_1(t + 3) + x_1(t)) \% 2,$$

$$t = 0, 1, \dots, 4N_{RB}^{\max,DL} + N_c - 32 \quad (3)$$

其中， $x_2(\cdot)$ 的前 31 个值满足如下条件：

$$c_{init} = \sum_{i=0}^{30} x_2(i) \cdot 2^i \quad (4)$$

$$c_{init} = 2^{10} \cdot (7 \cdot (n_s + 1) + l + 1) \cdot (2 \cdot N_{ID}^{cell} + 1) +$$

$$2 \cdot N_{ID}^{cell} + N_{CP} \quad (5)$$

$$N_{CP} = \begin{cases} 1, & \text{正常循环前缀;} \\ 0, & \text{扩展循环前缀} \end{cases} \quad (6)$$

其中， N_{ID}^{cell} 表示 LTE 基站的物理小区 ID (physical cell ID, PCI)， N_{CP} 表示 OFDM 循环前缀的类型。

综上所述，CRS 的生成与 slot 编号、OFDM 符号的编号、LTE 基站的物理小区 ID 以及 OFDM 循环前缀类型有关。

1.3 CSI 获取

获取 CSI 的方式称为信道估计，4G 信道估计主要依靠下行 CRS 进行，LTE CSI 提取过程的分阶段处理框图如图 4 所示。

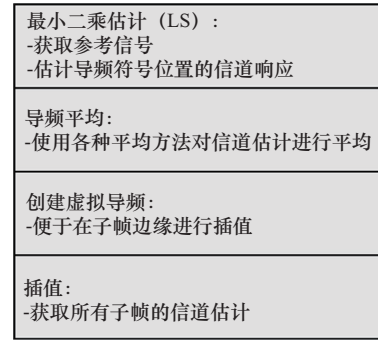


图4 LTE CSI提取过程的分阶段处理框图

1.3.1 最小二乘估计

从接收端提取发射-接收天线对的参考信号后，为了使实际接收数据与估计的接收数据之间的误差平方和最小，本文利用最小二乘 (least square, LS) 法估计准确的信道导频，并根据信道导频估计完整的信道矩阵。

LS 信道估计的结果可以表示为：

$$\hat{H}_{LS} = (X^H X)^{-1} X^H Y = X^{-1} Y \quad (7)$$

其中， X 表示原始发射信号矢量， H 表示信道响应矢量， Y 表示接收信号矢量。

对于 OFDM 系统，可以对每个子载波进行 LS 信道估计， $X[k]$ 和 $Y[k]$ 为复数，且 $X[k]$ 已进行幅值归一化，由于本文使用的 LTE 基站的下行链路中有 200 条子载波用于传输 CRS，每个子载波上的 LS 信道估计结果可以表示为：

$$\hat{H}_{LS}[k] = \frac{Y[k]}{X[k]}, k = 0, 1, \dots, 199 \quad (8)$$

1.3.2 导频平均

LTE CSI 的获取受到多径效应、移动速度和周围环境干扰等因素的影响，信道响应在时域和频域上可能出现快速变化和不稳定的情况。因此，须对多个导频信号进行平均，减小瞬时波动与噪声干扰，以获得更稳定的信道估计结果。

导频平均包含了时域平均和频域平均。时域平均包含了交织等操作，频域平均包含了卷积、滤波操作，经过上述操作后，在导频处的 CSI 值已经初步具备了整个物理资源块 (physical re-



source block, PRB) 时频域的相关噪声、多普勒偏移等信息。

1.3.3 创建虚拟导频

利用 LTE 进行人体动作识别时, 由于应用场景存在导频信号分布不均、频率资源有限的问题, 某些时刻或频段信道估计不够精确或完整。通过创建虚拟导频, 可以在原导频位置外插入模拟导频信号, 从而增加信道估计的密度和覆盖范围。

1.3.4 插值

在人体动作识别过程中, 通过信道估计得到的 LTE CSI 通常包含大量噪声和不规则的数据点。离散傅里叶变换 (discrete Fourier transform, DFT) 插值能够有效地平滑信道响应, 并填补缺失的数据点, 同时可以有效地缓解多径效应带来的信道衰落问题。

设 M 为导频数目, N 为插值得到的样点数 (即 DFT 点数), 且 $F=N/M$ 为整数, 则基于 DFT 的插值算法描述如下。

(1) 通过离散傅里叶逆变换 (inverse discrete Fourier transform, IDFT) 计算 $H_p(k^*)$ (导频位置 k^* 处的 CSI 估计) 变换域 $h_p(n)$ 为:

$$h_p((n)_M) = \text{IDFT} \{h_p(k)\}_M \triangleq \text{IDFT} \{H_p(k^*)\}_M \quad (9)$$

其中, $k^* = mF + l_p$, $l_p \in [0, F-1]$, $m \in [0, M-1]$, 下标 M 表示按 M 点的运算, $h_p(k)$ 指用于 IDFT 的频域序列, 等于 $H_p(k^*)$ 的中间符号。

(2) 对 $h_p(n)$ 作 N 点 DFT 即可得到插值序列为:

$$\hat{H}(k) = \text{ROR} \left\{ \text{DFT} \{h_p^*((n)_N)\}_N, l_p \right\} \quad (10)$$

其中, $h_p^*((n)_N)$ 为 $h_p((n)_M)$, 补充了 $(N-M)$ 零后的序列, $\text{ROR}(\cdot, l_p)$ 表示循环右移 l_p 。

至此, 可以获取所有子帧的信道估计, 其数据量为一个子载波在一个子帧内包含一个 CSI 值。

1.4 数据预处理

由于 LTE 基站与终端之间的传播距离较长 (200 m~10 km), 信号信噪比 (signal to noise ra-

tio, SNR) 较 Wi-Fi 低 25~40 dB, 目标移动引起的波动极易被噪声掩盖; 此外, 长传播路径还易受树木、车辆、行人等移动物体的干扰。针对上述挑战, 采用异常值去除、滤波去噪、Savitzky-Golay 平滑、峰谷值增强、格拉姆角场 (Gramian angular field, GAF) 及图像增强等技术对数据进行预处理, 有效地改善数据质量, 从而提升后续特征提取的准确率与效率, LTE CSI 数据预处理流程如图 5 所示。

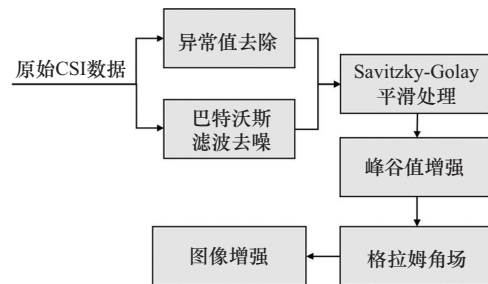


图 5 LTE CSI 数据预处理流程

1.4.1 异常值去除

基于 Hampel 滤波器的数据校验与精确修复算法可高效地完成异常值去除, 其输入为实际数据的多输入多输出 (multiple input multiple output, MIMO) 序列文件、预定义 MIMO (如 00 100 111 或 10 011 100) 以及采集的手势数据, 输出为异常值去除后的手势数据。算法执行流程如下: 首先, 读取 MIMO 序列文件; 其次, 检查该 MIMO 序列是否为指定预定义序列, 以此完成数据的初步校验; 若校验通过则输出 “校验成功”, 否则输出包含异常的错误行数; 随后, 针对须修正的行, 获取其文件编号与行号信息, 打开对应文件并读取待处理数据段, 应用 Hampel 滤波器对数据段进行异常值检测与替换操作, 再将处理后的数据段写回原文件; 最后, 输出所有修改完成的手势数据, 实现基于序列校验与滤波器优化的异常值精确修复。

1.4.2 滤波去噪

室内环境噪声会导致 LTE 原始数据出现波动,

影响特征提取与动作识别。鉴于巴特沃斯低通滤波器设计简单、去噪性能优越，且能有效地保留手势特征（手势动作多处于低频，而 LTE 数据含高频噪声），选用其对数据进行处理。滤波器原理为：

$$H(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \quad (11)$$

其中， n 为滤波器阶数， ω 为信号角频率， ω_c 为截止频率。滤波后波形平滑，噪声得到有效的抑制，异常值去除与巴特沃斯滤波器效果如图 6 所示。

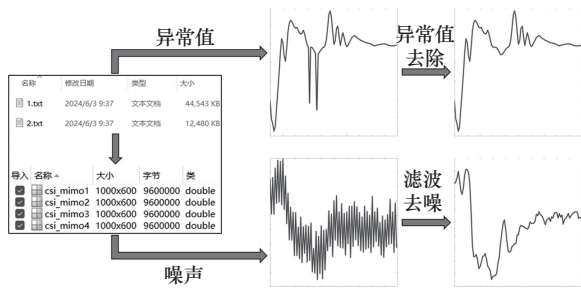


图 6 异常值去除与巴特沃斯滤波器效果

1.4.3 Savitzky-Golay 平滑处理

Savitzky-Golay 滤波器通过多项式拟合最小化拟合值与真实数据的残差。基于中心点 $n=0$ 的离散点进行 N 阶多项式拟合：

$$p(n) = \sum_{k=0}^N a_k n^k = a_0 + a_1 n + a_2 n^2 + \dots + a_k n^k \quad (12)$$

其中， N 为多项式阶次， a_k 为多项式系数， n 为该点离散序列值， $p(n)$ 为该点拟合值。

通过最小二乘法计算窗口内拟合值与原始数据的残差和 ε_N 为：

$$\varepsilon_N = \sum_{x=-M}^M (p(n) - x(n))^2 = \sum_{x=-M}^M \left(\sum_{k=0}^N a_k n^k - x(n) \right)^2 \quad (13)$$

求解 ε_N 最小值对应的多项式系数 a_k ，以窗口中点拟合值 $y(0) = p(0)$ 作为输出，通过单步平移滤波器遍历数据，实现平滑处理。

1.4.4 数据增强——峰谷值放大

本文设计了一种基于峰谷值放大的数据增强方法，旨在提升信号关键特征的显著性并增强数

据的多样性。该算法以原始信号 y 、峰值增强因子 $\text{amplification_factor}$ 、谷值减弱因子 reduction_factor 、时间窗口平移范围 shift_range 、随机缩放因子范围 scale_range 为输入，最终输出经增强处理的信号 augmented_signal 。具体执行步骤如下。

(1) 峰值与谷值检测：利用 findpeaks 函数检测信号 y 中的峰值集合 peaks 和谷值集合 valleys 。

(2) 峰谷值增强：对检测到的峰值进行增强操作 ($\text{peaks} \cdot \text{amplification_factor}$)，同时对谷值进行减弱操作 ($-\text{valleys} \cdot \text{reduction_factor}$)，突出信号关键特征。

(3) 多维度数据增强

① 时间窗口平移：生成随机平移量 $\text{shift} = \text{randi}(\text{shift_range})$ ，对经峰谷增强后的信号 enhanced_y 执行循环平移，得到 $\text{enhanced_y_shifted} = \text{circshift}(\text{enhanced_y}, \text{shift})$ ；

② 随机缩放：构造随机缩放因子 $\text{scale_factor} = 1 + (\text{scale_range}(2) - \text{scale_range}(1)) \cdot \text{rand} + \text{scale_range}(1)$ ，对平移后的信号进行缩放，得到 $\text{enhanced_y_scaled} = \text{enhanced_y_shifted} \cdot \text{scale_factor}$ ；

③ 信号合并：将缩放后的信号与噪声叠加，得到增强信号 $\text{augmented_signal} = \text{enhanced_y_scaled} + \text{noise}$ 。

Savitzky-Golay 平滑处理与峰谷值增强算法的效果如图 7 所示。

1.4.5 基于格拉姆角场的图像处理方法

GAF 可将一维时序信号转换为二维图像，便于深度学习图像网络处理以提取隐含特征。处理流程如下。

首先，将时间序列 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 归一化至区间 $[-1, 1]$ 。

$$\tilde{x}_i = \frac{(x_i - \max(\mathbf{X})) + (x_i - \min(\mathbf{X}))}{\max(\mathbf{X}) - \min(\mathbf{X})} \quad (14)$$

接着，利用反余弦转换将归一化序列转换为极坐标表示，确保角度在 $[0, \pi]$ 内单调（一一映射）并保留时间递增信息。

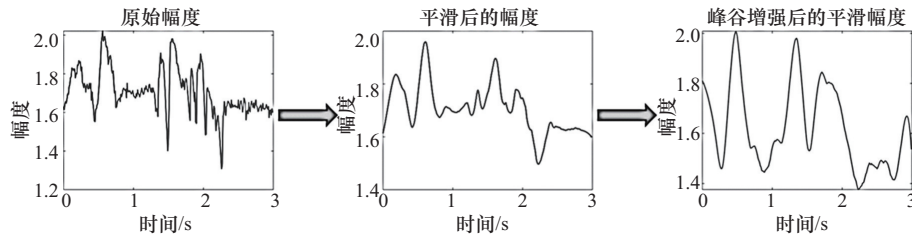


图7 Savitzky-Golay平滑处理与峰谷值增强算法的效果

$$\begin{cases} \phi = \arccos(\tilde{x}_i), -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{\mathbf{X}} \\ r = \frac{t_i}{N}, t_i \in N \end{cases} \quad (15)$$

其中, t_i 为时间戳, N 为极径坐标的正则化跨度。

最后, 利用三角函数和角公式定义时间点的相关性, 构建GAF矩阵。

$$\mathbf{G} = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} = \frac{\tilde{\mathbf{X}}' \cdot \tilde{\mathbf{X}} - \sqrt{\mathbf{I} - \tilde{\mathbf{X}}'^2} \cdot \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^2}}{\tilde{\mathbf{X}}' \cdot \tilde{\mathbf{X}} - \sqrt{\mathbf{I} - \tilde{\mathbf{X}}'^2} \cdot \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^2}} \quad (16)$$

其中, $\phi_i (i=1, 2, \dots, n)$ 是第 i 个时间点的角度, \mathbf{I} 为单位行向量, $\tilde{\mathbf{X}}'$ 表示向量 $\tilde{\mathbf{X}}$ 的转置。因 GAF 密度稀疏, 重新定义内积 $\langle x, y \rangle = x \cdot y - \sqrt{1-x^2} \cdot \sqrt{1-y^2}$ 以引入惩罚项, 消除噪声。

GAF 矩阵的每个元素都表征了不同时间点的余弦相关性, 且从左上角到右下角时间递增, 可直观地呈现数据时间依赖关系, 格拉姆角场法图像编码如图 8 所示。

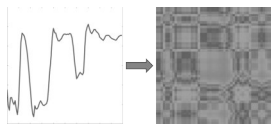


图8 格拉姆角场法图像编码

1.4.6 图像增强

限制对比度自适应直方图均衡化 (contrast limited adaptive histogram equalization, CLAHE) 是一种用于提高图像局部对比度, 尤其是增强图像细节的图像处理技术。

CLAHE 的主要步骤如下:

- (1) 将图像划分为多个小块;
- (2) 计算每个小块的直方图;
- (3) 对直方图进行裁剪和均衡化;
- (4) 使用双线性插值法平滑块之间的边界。

CLAHE 效果如图 9 所示。

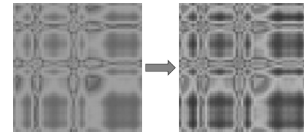


图9 CLAHE 效果

1.5 加入注意力机制的残差神经网络模型

在人体动作识别领域, 深度学习模型的选择对提升分类准确率和加速训练至关重要。ResNet 引入跳跃连接, 使梯度能在网络中直接流动, 缓解了退化问题, 并能够有效地训练更深的网络。这一架构既提高了模型的收敛速度, 也提升了准确性。

1.5.1 ConvNeXt 残差网络模型

本文前期工作中使用 ResNet 对 GAF 图像进行分类训练实现人体动作识别时, 训练结果波动大, 且准确率与收敛速度存在不足, 因此本文提出了加入注意力机制的改进 ConvNeXt^[16]。

- (1) 层归一化 (layer normalization, LN) 层

LN 层是借鉴 Transformer 网络对残差网络的一个改进, 用于取代传统残差网络中的批归一化 (batch normalization, BN) 层。LN 是一种横向的规范化, 综合考虑了一层所有维度的输入, 计算该层的平均输入值和输入方差, 然后用同一个规范化操作来转换各个维度的输入。LN 针对单个训

练样本进行，不依赖于其他数据，因此可以避免BN中受批数据分布影响的问题，可以用于小批数据规模场景。

(2) 改进的残差模块

ConvNeXt网络的残差模块优化集中在以下几方面：①将原7×7卷积（步距2）和最大池化层简化为单层4×4卷积（步距4），以匹配Transformer的降采样方式；②调整残差结构卷积顺序，将1-3-1核改为7-1-1核，并在中间层扩展通道数为4倍，形成中间宽、两端窄的棱柱结构；③取消残差层内部下采样，改为层间独立下采样层实现特征图减半；④原残差结构每卷积后接BN+ReLU，改进后仅首层卷积后加LN，次层后加高斯误差线性单元（Gaussian error linear unit, GELU）；⑤残差层结构由[3, 4, 6, 3]调整为[3, 3, 9, 3]，首层卷积通道数从64增至96。

模型的整体结构如图10所示。图中，dim代表特征维度（通道数）， k 与 s 分别标注了卷积核大小与步长，模块旁的 $\times N$ 表示该模块的堆叠次数， p 表示卷积操作的分组数。

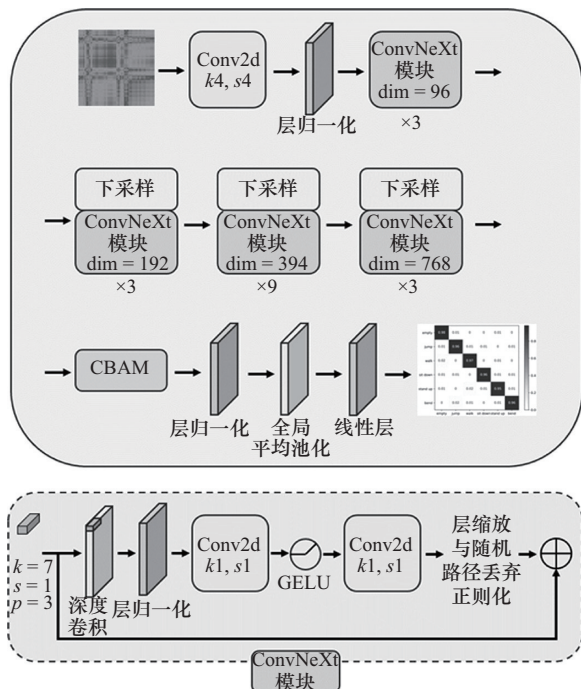


图10 模型的整体结构

1.5.2 CBAM注意力机制模块

GAF图像特有的编码方式使信号的特征左上到右下呈线性排列，而在右上到左下的方向上，图像特征的重要性从中心向两边沿线性依次递减。为使神经网络能够逐渐将注意力向信号特征密集的区域靠拢，本文在残差网络中加入注意力机制来加快特征提取的速度和提高准确性。

卷积块注意力模块（convolutional block attention module, CBAM）采用通道和空间注意力机制的组合，沿着通道和空间两个独立的维度依次推断注意力权重。CBAM结构如图11所示，该模块以前一层输出的特征图 F 作为输入，首先，通过通道注意力模块 M_c 生成通道注意力权重，并与输入特征图相乘得到通道细化后的特征 F' ；随后， F' 输入空间注意力模块 M_s 生成空间注意力权重，进一步优化特征，最终输出精炼后的特征图 F'' 。通过这种将注意力权重与输入特征图相乘以进行自适应特征细化的方式，有效地提高了模型的精度。

2 实验验证

2.1 实验设置

软硬件配置：硬件部分，利用软件无线电USRP B210接收LTE信号，使用联想Thinkbook 14+笔记本电脑（i9-12900H, 2.5 GHz, 32 GB RAM）进行信号处理和模型搭建；软件部分，对开源4G项目srsRAN的代码进行修改，连续提取LTE CSI，并用MATLAB对接收到的信号进行处理。

实验环境：实验在一个4 m×8 m的会议室中进行，包含一张多人会议桌和若干把椅子。实验环境如图12所示，实验选取与会议室相距1 km的LTE基站（LTE eNodeB, LTE-Enb）作为信号发射端，在室内部署LTE用户设备（LTE user equipment, LTE UE）作为接收终端以采集信号

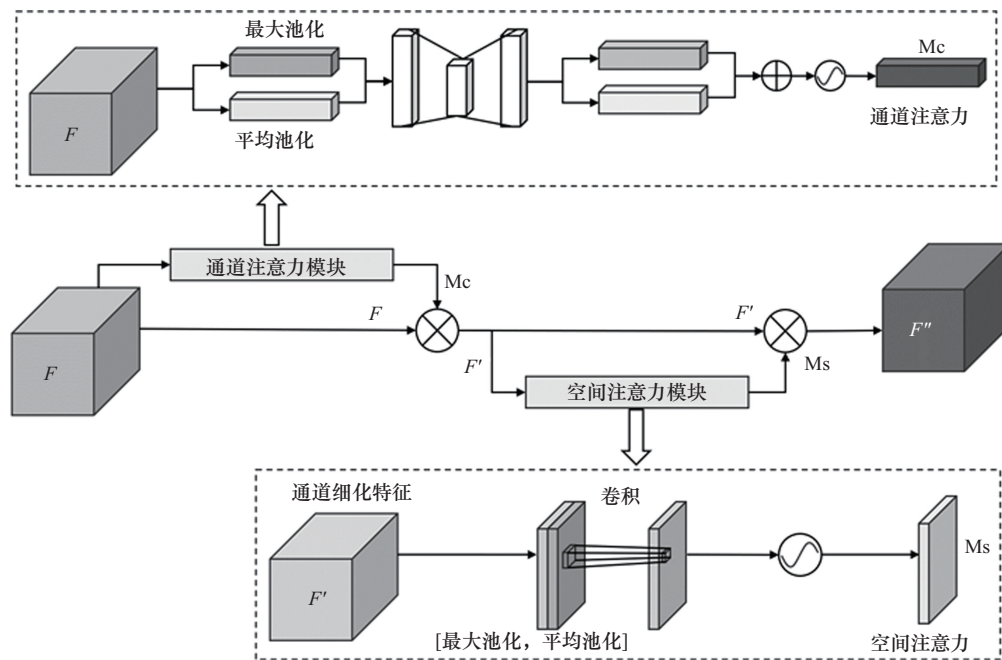


图11 CBAM结构

并生成动作数据集。志愿者在距该 UE 终端 2 m 处执行指定动作。

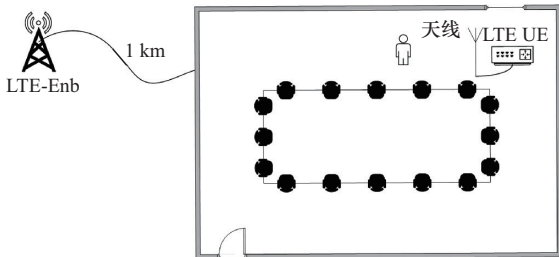


图12 实验环境

活动类别：采集了空、跳跃、行走、坐下、起立、弯腰6类动作，每类动作采集3 000个以上样本，每个样本持续3 s。手势动作数据集概要信息见表1。

表1 手势动作数据集概要信息

手势动作类别	样本数量
空	3 207
跳跃	3 184
行走	3 328
坐下	3 056
起立	3 145
弯腰	3 260

实验设置：本文使用 Python3.6 和 Tensorflow 库进行人体动作识别分类模型的验证，训练过程中，批次大小 (batch-size) 设置为 10，最大迭代次数 (epoch) 设置为 60，学习率设置为 0.000 5，损失函数采用 softmax 交叉熵损失函数，优化器使用 Adam 优化器，评估模型的性能指标主要为混淆矩阵、准确率与 Loss 值。

2.2 人体动作分类识别的评估与分析

本文将 70% 的数据用于训练模型，30% 用于测试。人体动作分类识别混淆矩阵如图 13 所示。从结果中可以看出，模型在大多数类别上表现良好，如“行走”和“起立”，这些类别的精确度和召回率较高。然而，一些类别如“弯腰”和“坐下”表现略有下降，可能受到数据不平衡或者特定动作特征的影响。综合评估指标显示，模型的准确率达到 96.33%、精确度为 97.85%，表明模型在多类别动作识别任务中展现出显著的整体性能和稳定性。

2.3 接收天线与监测目标之间不同间距的影响

为进一步研究动作分类准确性受 LTE 终端接

收天线与目标之间距离的影响, 本文以接收天线与监测目标之间的距离为变量, 使目标以 0.5 m 的步长改变与接收天线的距离, 分别采集了 0.5~3.0 m 距离下的人体动作数据。并且, 分别处理各种距离下采集到的人体动作数据, 不同距离下人体动作的准确率如图 14 所示。

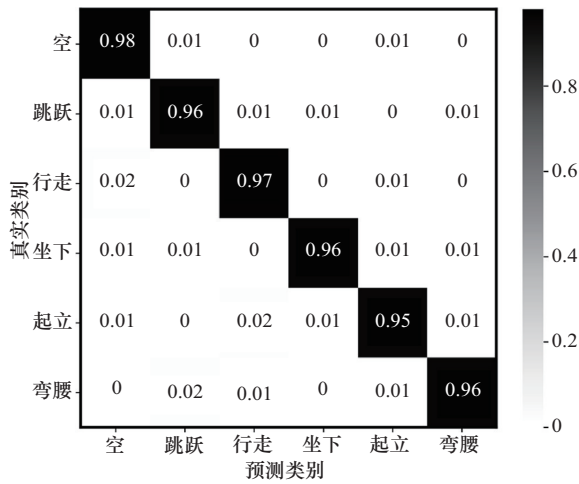


图 13 人体动作分类识别混淆矩阵

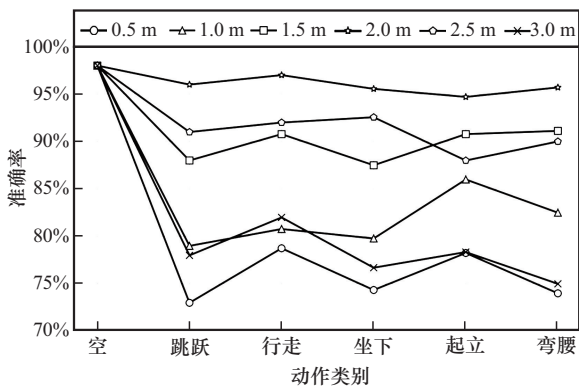


图 14 不同距离下人体动作的准确率

由图 14 可知, 接收天线与监测目标之间距离为 2.0 m 时, 人体动作分类的准确率最高。接收天线与监测目标之间距离太近时, 呼吸起伏等干扰对信道的影响较大, 从而降低了对人体动作分类的准确率。接收天线与监测目标之间距离稍远时, 由于路径损耗, 信号在传播过程中会随着距离的增加而逐渐衰减, 分类的准确率降低。

2.4 不同 LTE 基站对识别准确率的影响

上述实验采用距离实验场地 1 km 的时分双工 (time division duplexing, TDD) 制式 LTE 宏基站 1 (PCI 为 189) 发射的 2 330 MHz、带宽 20 MHz 的信号。本节更换为距接收端 2.5 km 的频分双工 (frequency division duplexing, FDD) 制式宏基站 2 (PCI 为 384), 其 LTE 信号中心频率为 1 845 MHz, 带宽同为 20 MHz。不同 LTE 基站对应的动作识别准确率如图 15 所示。

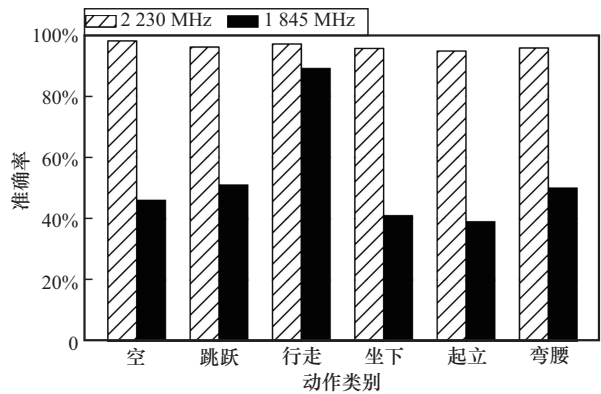


图 15 不同 LTE 基站对应的动作识别准确率

从图 15 可见, 使用宏基站 2 的识别准确率远低于宏基站 1。原因一方面是宏基站 2 的 LTE 信号传输距离更远, 衰减更严重; 另一方面是其传播环境更复杂, 多径效应增强, 掩盖了人体动作引起的信号变化。因此, 仅“行走”动作的识别准确率略高, 其他动作识别效果明显下降。

2.5 与现有方法比较

为了验证本文所提的 LTE-ACN 在人体动作识别中的准确性, 将模型分别与直接使用 ConvNeXt 网络以及基于 Wi-Fi CSI 的一些先进模型进行比较, 包括使用了 CNN 的 SignFi^[17]、LSTM^[18]、使用了 CNN 和 LSTM 的 Wi-Sensing^[19]、采用了 CCFE 方法的 MatNet-eCSI^[20]、结合了 CBAM 和 Resnet50 的 CSI-GDAM^[21]。不同模型准确率比较见表 2。



表2 不同模型准确率比较

模型	空	跳跃	行走	坐下	起立	弯腰	平均准确率
SignFi ^[17]	82.56%	80.87%	83.27%	79.33%	78.51%	80.06%	80.77%
LSTM ^[18]	84.04%	84.43%	85.64%	82.74%	82.75%	83.83%	83.91%
Wi-Sensing ^[19]	92.42%	92.65%	91.73%	89.83%	91.62%	90.26%	91.42%
MatNet-eCSI ^[20]	93.48%	90.47%	94.63%	92.93%	88.62%	93.29%	92.24%
CSI-GDAM ^[21]	95.49%	95.83%	94.88%	93.85%	93.71%	93.69%	94.58%
LTE-ACN	98.24%	96.48%	97.12%	95.75%	94.69%	96.37%	96.44%

研究发现,传统的深度学习模型,如CNN和LSTM在整体识别准确率上并不理想,并且容易将数据特征相似的动作变化相混淆。比如,CNN较难区分具有相反时域特征的活动,如“坐下”和“起立”,因为CNN很难保留时间顺序的上下文特征关系;由于经过了对传统网络的改进,Wi-Sensing、MatNet-eCSI和CSI-GDAM这3种方法在本数据集的表现相对良好,识别准确率分别达到91.42%、92.24%和94.58%;另外,改进后的ConvNeXt网络架构能够更高效地从GAF图像中提取判别性特征,同时通过引入CBAM注意力机制,显著优化了网络在通道和空间维度上的特征关联性,因此,所提模型LTE-ACN识别准确率达到96.44%,性能表现最好。

3 结束语

本文提出了一种基于LTE CSI和GAF的人体动作识别系统。首先,通过修改开源4G项目srsRAN提取各种人体动作信号的LTE CSI数据,并经过预处理后将时序信号数据编码为GAF图像;然后,改进优化残差神经网络ConvNeXt构建LTE-ACN模型以适配GAF图像特有的空间分布特性,使模型的准确率获得提升;最后,通过实验分析模型性能,并与其他算法进行对比分析。实验结果表明,本文所提的人体动作分类识别方法优于现有的先进方法。

后续工作将通过进一步优化LTE-ACN模型结

构,降低计算复杂度,开发轻量级、高效的深度学习模型,从而提高系统的实用性和可扩展性。

参考文献:

- [1] Nikpour B, Sinodinos D, Armanfard N. Deep reinforcement learning in human activity recognition: a survey and outlook[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(3): 4267-4278.
- [2] Chen Z, Zhu Q, Soh Y C, et al. Robust human activity recognition using smartphone sensors via CT-PCA and online SVM[J]. IEEE Transactions on Industrial Informatics, 2017, 13(6): 3070-3080.
- [3] Ertin E, Stohs N, Kumar S, et al. AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field[C]//Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2011: 274-287.
- [4] Politi O, Mporas I, Mega V. Human motion detection in daily activity tasks using wearable sensors[C]//Proceedings of the 22nd European Signal Processing Conference (EUSIPCO). Piscataway: IEEE Press, 2014: 2315-2319.
- [5] Aggarwal J K, Ryoo M S. Human activity analysis: a review[J]. ACM Computer Survey, 2011, 43(3): 1-43.
- [6] Alrashdi I, Siddiqi M H, Alhwaiti Y, et al. Maximum entropy Markov model for human activity recognition using depth camera[J]. IEEE Access, 2021, 9: 160635-160645.
- [7] Abbas M, Jeannès R L B. Exploiting local temporal characteristics via multinomial decomposition algorithm for real-time activity recognition[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70(11): 1-11.
- [8] Xu W, Liu J, Zhang S, et al. RFace: anti-spoofing facial authentication using COTS RFID [C]//Proceedings of the IEEE INFOCOM 2021. Piscataway: IEEE Press, 2021: 1-10.
- [9] Zheng T, Chen Z, Zhang S, et al. Catch your breath: simultane-

- ous RF tracking and respiration monitoring with radar pairs[J]. IEEE Transactions on Mobile Computing, 2023, 22(11): 6283-6296.
- [10] Zhang S, Zheng T, Chen Z, et al. Can we obtain fine-grained heartbeat waveform via contact-free RF-sensing?[C]//Proceedings of the IEEE INFOCOM 2022. Piscataway: IEEE Press, 2022: 1759-1768.
- [11] Cha J, Choi I, Yoo K. Human activity classification with user-held fmcw radar using deep recurrent neural networks[J]. IEEE Sensors Letters, 2025, 9(5): 1-4.
- [12] Wu Z, Xu Q, Li J, et al. Passive indoor localization based on csi and naive bayes classification[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 8(9): 1566-1577.
- [13] Pan Y, Zhou Z, Gong W, et al. SAT: a selective adversarial training approach for Wi-Fi-based human activity recognition[J]. IEEE Transactions on Mobile Computing, 2024, 23(12): 12706-12716.
- [14] Feng Y, Xie Y, Ganesan D, et al. LTE-based pervasive sensing across indoor and outdoor[C]//Proceedings of the SenSys 2021. New York: ACM Press, 2021: 138-151.
- [15] Sonny A, Rai P K, Kumar A, et al. Deep learning-based smart parking solution using channel state information in lte-based cellular networks[C]//Proceedings of the International Conference on Communication Systems & Networks (COMSNETS). Bengaluru, 2020: 642-645.
- [16] Liu Z, Mao H, Wu C Y, et al. A ConvNet for the 2020s[C]//Proceedings of the CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 11966-11976.
- [17] Ma Y, Zhou G, Wang S, et al. SignFi: sign language recognition using Wi-Fi[J]. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 1-21.
- [18] Yousefi S, Narui H, Dayal S, et al. A survey on behavior recognition using wifi channel state information[J]. IEEE Communications Magazine, 2017, 55(10): 98-104.
- [19] Lee H, Ahn C R, Choi N. Fine-grained occupant activity monitoring with Wi-Fi channel state information: practical implementation of multiple receiver settings[J]. Advanced Engineering Informatics, 2020, 46(10): 101147.

- [20] Shi Z, Zhang J A, Xu R Y, et al. Environment-robust device-free human activity recognition with channel-state-information enhancement and one-shot learning[J]. IEEE Transactions on Mobile Computing, 2022, 21(2): 540-554.
- [21] Zhang Y, Chen Y, Wang Y, et al. CSI-based human activity recognition with graph few-shot learning[J]. IEEE Internet of Things Journal, 2022, 9(6): 4139-4151.

[作者简介]



钱能 (1982-), 男, 浙江邮电职业技术学院电子与通信工程学院副教授, 主要研究方向为普适感知和信息融合技术等。



陆城灵 (1999-), 女, 浙江工业大学网络空间安全研究院硕士生, 主要研究方向为无线感知和深度学习技术等。



万雨轩 (1999-), 女, 浙江工业大学网络空间安全研究院博士生, 主要研究方向为通感一体化和信号处理技术等。



吴哲夫 (1971-), 男, 浙江工业大学网络空间安全研究院副教授、博士生导师, 主要研究方向为无线感知、信号处理和深度学习技术等。