基于双注意力增强多尺度卷积神经网络的CSI人体活动识别

CSI Human Activity Recognition Based on Dual-Attention Multi-Scale Enhanced Convolutional Neural Network

赵纯熙[1]，袁　毅[2]，刘　懿[2]（1. 中讯邮电咨询设计院有限公司，北京 100048；2. 南京大学，江苏 南京 210046）
Zhao Chunxi[1], Yuan Yi[2], Liu Yi[2]（1. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 2. Nanjing University, Nanjing 210046, China）

摘　要：
在智能安防与智慧家居等应用场景中，基于Wi-Fi信道状态信息的人体活动识别具有较高的研究价值与应用潜力。研究提出一种基于双注意力增强多尺度卷积神经网络的人体活动识别方法，通过引入子载波维度与特征通道维度的双重注意力机制，并结合多尺度卷积结构，有效提升模型对复杂CSI特征的学习与表征能力。多组公开数据集实验结果表明，该方法可实现更高的平均识别准确率，整体性能显著优于主流基线模型，尤其对行为模式相近的活动及细微动作识别，展现出更优异的鲁棒性。

关键词：
信道状态信息；人体活动识别；注意力机制；卷积神经网络
doi：10.12045/j.issn.1007-3043.2026.10.015
文章编号：1007-3043（2026）10-0087-06
中图分类号：TP18
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
In application scenarios such as intelligent security and smart home, human activity recognition based on Wi-Fi Channel State Information (CSI) holds important research value and promising application potential. A human activity recognition method based on a dual attention-enhanced multi-scale convolutional neural network is proposed. By introducing a dual attention mechanism involving the subcarrier and feature channel dimensions, and integrating a multi-scale convolutional architecture, the proposed method can effectively enhance its capability to learn and represent complex CSI features. Experimental results on multiple public datasets demonstrate that the proposed method achieves higher average recognition accuracy, and its overall performance significantly outperforms that of mainstream baseline models. In particular, it delivers superior robustness when recognizing fine-grained motions and activities with similar behavioral patterns.

Keywords：
Channel state information; Human activity recognition; Attention mechanism; Convolutional neural network

引用格式：赵纯熙，袁毅，刘懿. 基于双注意力增强多尺度卷积神经网络的CSI人体活动识别[J]. 邮电设计技术，2026(10)：87-92.

1 概述

随着智能化服务的普及，精确且无感知的人体活动识别已成为智能安防与健康监护等领域的核心技术，具有重要研究价值与广阔的应用前景[1]。当前主流识别方法存在一定局限性：计算机视觉类方法识别效果直观，但存在隐私泄露隐患，且易受光照、遮挡等环境因素干扰[2]；可穿戴设备识别精度较高，但其侵入式设计与频繁充电需求，严重制约了实际推广[3]。为克服上述局限，基于全域覆盖Wi-Fi网络的无线感知技术，凭其低成本、广覆盖、非侵入等优势备受关注[4]。

早期无线感知研究多采用接收信号强度指示（RSSI），但该指标为单一标量、信息粒度粗糙，难以应对复杂人体活动识别。因此，研究重点转向更精细的物理层信道状态信息（CSI）。在室内多径传播环境中，人体活动会通过遮挡、反射等方式扰动信号传播路径，而CSI能以多子载波粒度精准感知信道动态变化，形成可辨识的活动“指纹”，为非接触式高精度感知奠

收稿日期：2026-08-27

定物理基础[5]。初期研究借助菲涅尔区模型,多普勒频移估计等物理模型进行参数化建模[6],但此类方法对硬件设备要求严苛,在复杂多径环境中易出现信号失真,难以表征姿态相似或细微动作,实际应用受限。

为此,摒弃复杂物理建模,转而从数据中自主学习映射关系的人工智能(AI)方法成为主流。其发展路径遵循着从传统机器学习到深度学习的演进。早期研究多采用支持向量机、随机森林[7]和K近邻[8]等机器学习算法,但受限于繁琐的手动特征工程,模型泛化能力不足。深度学习的出现实现了技术突破,其端到端的学习模式可自动完成最优特征的提取与表达。针对CSI数据特性,卷积神经网络(CNN)被广泛用于提取时频图中的局部纹理特征[9],循环神经网络(RNN)及其变体(如LSTM)则擅长建模信号在时间维度上的长程依赖关系[10-11]。二者融合架构显著提升了识别精度,成为当前该领域内的主流技术路径。但如何从高维、混噪的CSI数据中,智能提取对活动敏感且鲁棒的判别性特征,仍是当前亟待突破的瓶颈。

2 AI驱动的CSI人体活动识别方法

为实现高精度、非侵入式的室内人体活动识别,本文提出了一种基于双注意力增强卷积神经网络的完整识别方法,涵盖数据采集、预处理、智能处理与分类以及结果输出等核心模块。本章将重点阐述该方法的设计思路、整体流程与关键技术。

2.1 总体框架

本文所提出的基于双注意力增强卷积神经网络的CSI人体活动识别方法的总体框架如图1所示。该方法构建了一套完整的端到端处理流程,通过多阶段精细化处理,将物理层采集的原始信道状态信息逐步转化为高层语义的人体活动识别结果。整个系统由4个核心模块依次构成:CSI数据收集、数据预处理、AI感知模块以及活动识别输出。

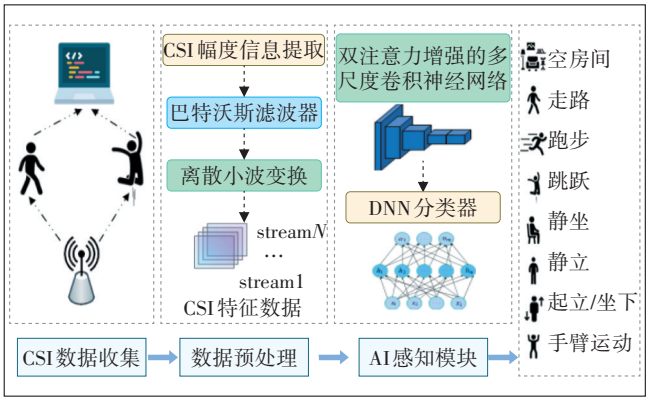


图1 CSI人体活动识别方法总体框架

首先,通过CSI数据收集模块获取能够表征人体活动变化的原始CSI数据。随后,经数据预处理对原始CSI数据进行清洗、去噪、缺失值填充及无效子载波剔除等操作,以抑制环境噪声与设备随机性干扰。在此基础上,结合信号处理与领域知识,对数据进行时频变换与统计建模,构建出能够表征人体动态的结构化特征数据。将处理后的特征数据输入AI感知模块,以双注意力机制增强的多尺度卷积神经网络等核心算法,通过深度学习技术自动学习CSI特征与人体活动之间的映射关系,实现从信号特征到语义类别的高精度映射。最终,系统输出如“走路”“静坐”“手臂运动”等具体活动类别,为各类智能感知应用提供结构化的识别信息。

2.2 CSI数据收集

本研究采用基于Nexmon CSI开源工具的数据采集方案,从商用无线接入设备的Wi-Fi芯片中提取原始CSI数据[12]。该方案无须依赖昂贵的专用硬件,显著降低部署成本,同时保障了数据采集的稳定性与可靠性。采集过程中,系统以固定采样率持续监听无线信道,确保精准捕捉人体活动引起的细微信道变化。为构建模型训练样本并支持准实时识别,本文采用滑动窗口机制对连续CSI数据流进行分段处理,将每个窗口内的数据作为独立样本,用于模型即时推断,有效提升了整体方法的响应速度与实际应用效率。

2.3 数据预处理

原始CSI数据需经一系列精细化的处理与特征提取,才能作为AI模型的有效输入。首先,考虑到相位信息易受硬件噪声干扰,从原始复数形式的CSI数据中提取稳定性更高的振幅信息作为分析基础。随后,采用巴特沃斯低通滤波器对CSI振幅序列进行处理,抑制直流分量和环境中的高频随机噪声,保留与人体活动相关的主要频带成分。最后,利用离散小波变换对滤波后的信号进行分解,深度挖掘能够表征活动细节的深层时频特征,最终得到蕴含丰富动态信息的CSI特征数据,为后续的高精度识别奠定坚实基础[13]。

2.4 AI驱动的特征提取模型

为实现对CSI特征数据的深度解析与精准识别,本文设计了一种新颖的双注意力增强的多尺度卷积神经网络作为核心特征提取模型,其结构如图2所示。该模型通过引入双重注意力机制与多尺度融合架构,

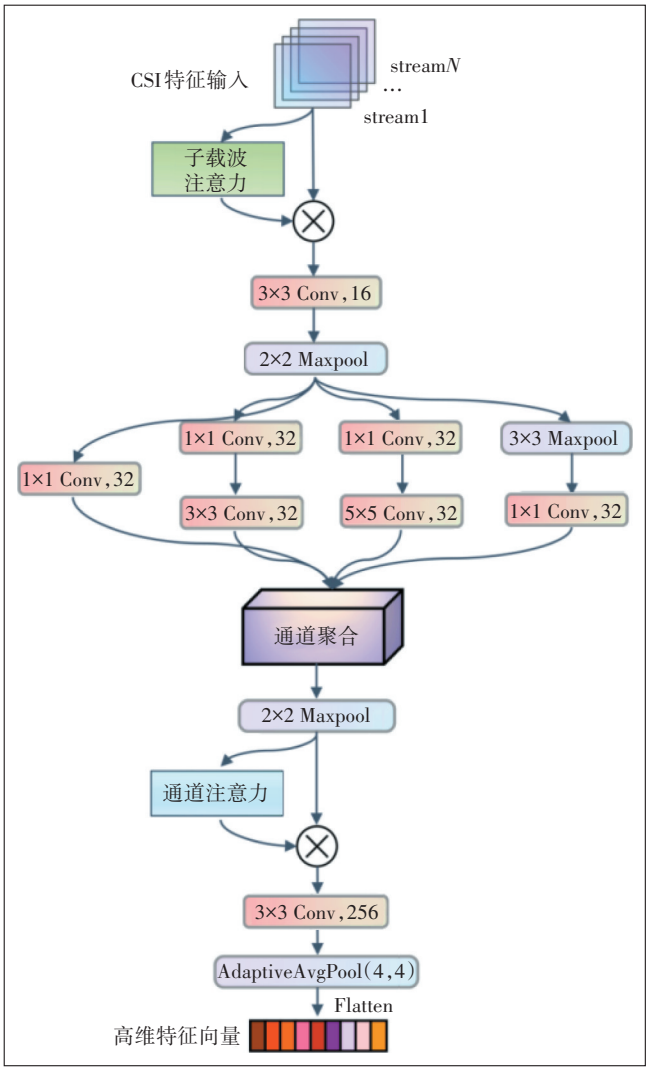


图2　双注意力增强的多尺度卷积神经网络特征提取模型

使网络能够分层次、自适应地从复杂CSI数据中提取具有高度判别性的特征。

模型以预处理后的CSI特征数据$X \in R^{A\times T\times S}$作为输入，其中$A$为天线数，$T$为时间序列长度，$S$为子载波数。首先，数据经过子载波注意力模块。该模块借鉴空间注意力机制[14]，可自适应地学习不同子载波对当前活动识别的贡献权重，并在子载波维度对特征数据进行加权，从而突出对相应活动敏感的关键频段、抑制噪声干扰。具体而言，该模块如公式(1)和公式(2)所示，通过全局平均池化压缩非子载波维度，生成每个子载波对应的全局描述符，再通过两层带激活函数的全连接层学习各子载波的权重W_s：

$$z_s = \frac{1}{A\times T}\sum_{a=1}^{A}\sum_{t=1}^{T}X(a,t,s) \tag{1}$$

$$W_s = \mathrm{Sigmoid}(W_2 \cdot \mathrm{ReLU}(W_1 \cdot z_s)) \tag{2}$$

$$X' = W_s \otimes X \tag{3}$$

其中，z_s为第s个子载波池化压缩后的全局描述符，W_1和W_2为全连接层权重。接着根据式(3)，通过权重W_s对原始输入进行加权得到特征张量X'。

随后，模型采用并行多尺度特征提取结构，通过不同尺寸的卷积核(如1×1、3×3、5×5)与池化分支，融合多粒度时频特征，增强模型对不同尺度活动模式的感知能力。其中，1×1卷积主要用于特征降维与跨通道信息整合，3×3和5×5卷积能够提取更宽范围内的空间-时序结构信息，池化操作则提升特征平移并减少冗余。如式(4)所示，各分支输出的特征在通道维度拼接，形成融合多尺度信息的特征张量F_{multi}：

$$F_{\mathrm{multi}} = \mathrm{Concat}[\mathrm{Conv}_{1\times1}(X'),\mathrm{Conv}_{3\times3}(X'),\mathrm{Conv}_{5\times5}(X'),\mathrm{MaxPool}(X')] \tag{4}$$

在多尺度特征融合的基础上，模型进一步引入特征通道注意力模块[15]，用于显式建模通道间的依赖关系。该模块与子载波注意力类似，通过全局平均池化和全连接层学习各特征通道权重，强化高信息含量通道，优化整体特征表示。

通过子载波注意力的前端筛选、多尺度卷积的并行提取以及后端通道注意力增强，网络逐步提炼出高判别性深度特征。最终特征被展平为高维特征向量，用于后续的高精度活动分类。

2.5 分类器

经核心特征提取模型处理后得到的高维特征向量，由基于深度神经网络(DNN)的分类器完成最终活动分类，其结构如图3所示。该分类器以前序模块输出的高维特征向量作为输入，依次通过包含2 048和512个神经元的全连接层，配合非线性激活函数进行

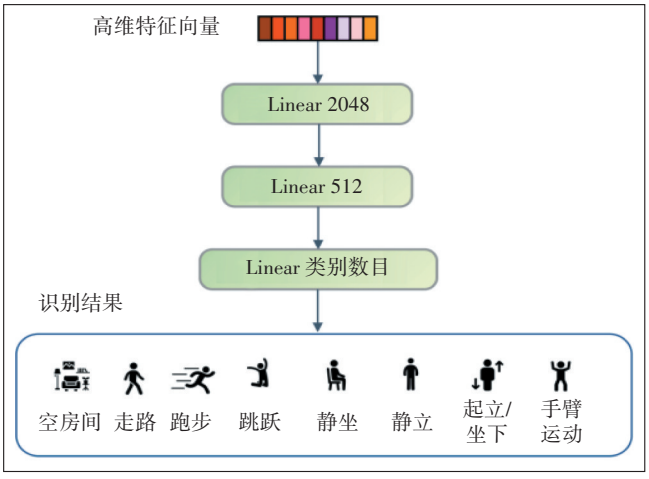


图3　基于DNN的分类器

变换,逐步将特征映射至判别性更强的低维表示。最后通过与活动类别数目相同的输出层,并利用Softmax函数将输出映射为类别概率分布,将概率最高类别作为最终识别结果。

2.6 模型训练与优化

在模型训练阶段,本文将数据集按80%、15%和5%的比例划分为训练集、验证集和测试集。采用交叉熵损失与中心损失(Center Loss)联合监督,在扩大类间差异的同时减小类内距离,提升特征判别力。优化器选用带有权重衰退的AdamW,并结合余弦退火调度器动态调整学习率,以实现高效、稳定的模型收敛。

3 多活动分类方法性能评估

3.1 实验设置

为验证所提方法的有效性和准确率,采用了NTU-Fi[16]、UT-HAR[17]和80 M-CSI[18]3个广泛使用的开源数据集。其中,80 M-CSI数据集基于IEEE 802.11 ac标准和80 MHz信道带宽采集,包含13名受试者在不同环境、时段及硬件配置下的多场景CSI数据,具有数据丰富度高、领域覆盖全面的特点,能更充分地体现本文算法的性能优势,因此,会针对该数据集的实验结果进行详细分析。实验硬件平台配置为AMD EPYC 9554 CPU、NVIDIA GeForce RTX 4090D 24 GB GPU和1 024 GB内存的计算服务器。算法基于Pytorch2.4.1框架实现,初始学习率设置为1×10^{-3},训练批次大小为64,测试批次大小为32。

3.2 实验结果分析

本文以识别准确率作为模型性能的评估指标,并在3个开源数据集上将所提的方法与多种现有识别算法(包括基于MLP的识别算法[19]、基于BiLSTM的识别算法[19]以及基于DeepCNN的识别算法[20])进行对比。实验结果如表1所示,本文方法在3个开源数据集上的识别准确率均优于其他基线方法,且整体保持在较高水平,初步验证了所提方法的优越性。

此外针对任务更为复杂多样的80 M-CSI数据集进行重点分析。所提方法在该数据集上取得了97.85%的平均识别准确率,优于所有基线模型。作为对比,基于DeepCNN的方法准确率为95.12%,BiLSTM方法的准确率为94.94%,而MLP方法的准确率仅为80.74%。该结果初步证明了本文模型整体架构的先进性与有效性,表明其能够从高维复杂的CSI数据中学习到更具判别力的深层特征。

为深入分析各模型对不同活动的具体分类性能及易混淆情况,图4至图7分别展示了本文方法、MLP、BiLSTM和DeepCNN在测试集上的混淆矩阵。其中,

表1 所提方法与基线方法在3种开源数据集上准确率对比

方法	NTU-Fi	UT-HAR	80 M-CSI
本文所提方法准确率/%	100.00	99.20	97.85
MLP方法准确率/%	94.32	91.16	80.74
BiLSTM方法准确率/%	98.48	94.88	94.94
DeepCNN方法准确率/%	98.86	93.88	95.12

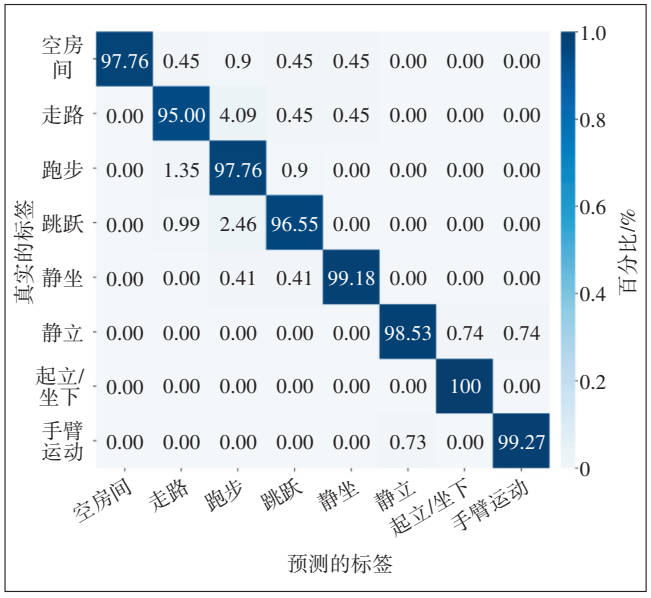


图4 本文所提方法混淆矩阵(平均准确率为97.85%)

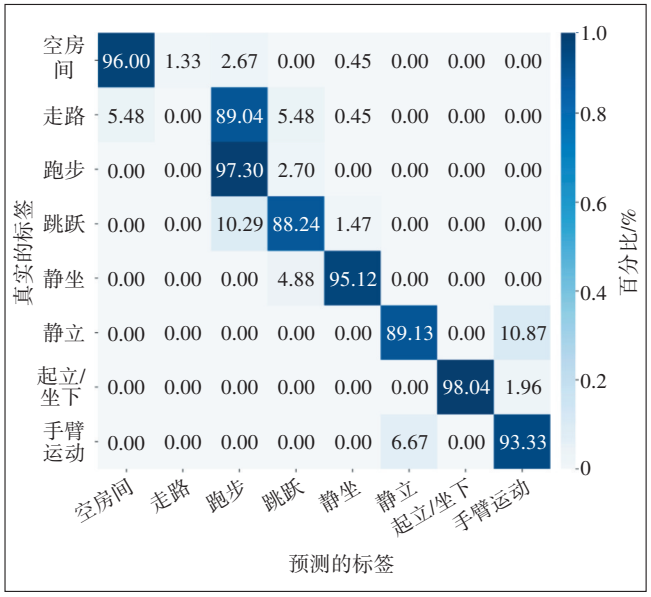


图5 基于MLP识别方法混淆矩阵(平均准确率为80.74%)

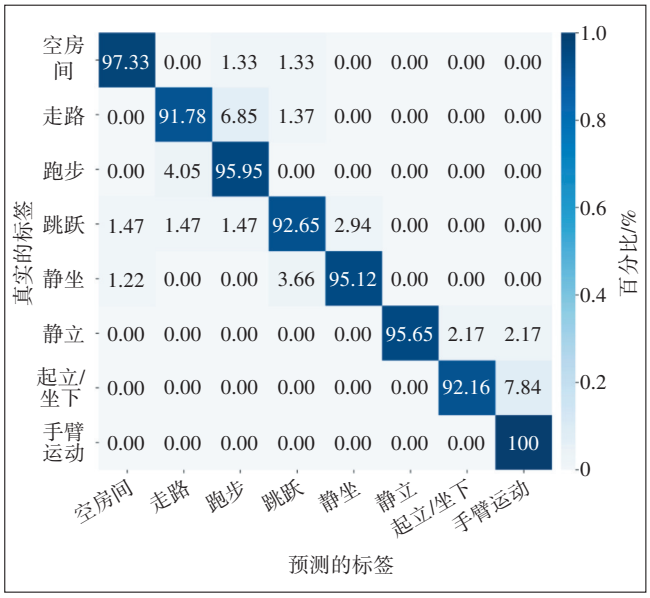


图6　基于BiLSTM识别方法混淆矩阵（平均准确率为94.94%）

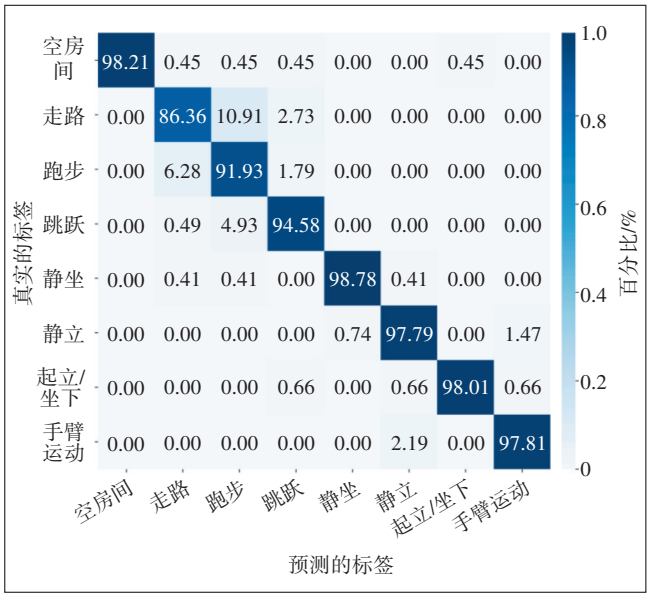


图7　基于DeepCNN识别方法混淆矩阵（平均准确率95.12%）

混淆矩阵主对角线元素表示模型对该类别活动的识别准确率，非对角线元素则反映了类别间误判分布情况。图8对比了4种方法在8类活动上的准确率，可更直观地考察各模型在不同类别上的识别能力差异。

在“走路”“跑步”“跳跃”等动作模式相似、识别难度较高的任务中，本文方法表现最优（见图4），其对上述3类活动的识别准确率分别达到95.00%、97.76%和96.55%，显著优于各基线模型。例如，图7中DeepCNN模型将10.91%的“走路”样本误判为“跑步”，图6中BiLSTM模型也存在类似混淆现象。而图5中性能最差的MLP模型更是在多个类别间出现了严重混淆，将89.04%的“走路”误判为“跑步”，说明其难以捕捉CSI信号中不同活动间的细微差别。

同时，本文方法的优势同样体现在对“静坐”“静立”“起立坐下”以及“手臂运动”等细微或静态活动的精准识别上。如图4所示，该方法对过渡性动作“起立坐下”的识别准确率达到100%，证明了模型对短时动态变化的捕捉能力极强。对于“静立”和“手臂运动”这类易受环境噪声干扰的微小动作，本文方法也分别取得了98.53%和99.27%的高准确率。相比之下，图6中BiLSTM模型将7.84%的“起立坐下”误识别为“手臂运动”，图5中MLP模型则将10.87%的“静立”误判为“手臂运动”，上述结果均表明基线模型在细微动作区分上存在明显短板。

综上，无论从整体平均准确率，还是从对各类活动的细粒度识别能力来看，本文所提方法均全面优于MLP、BiLSTM和DeepCNN等基线模型。实验结果表明，该方法能够更有效地从CSI数据中提取具有判别性的时空耦合特征，在区分动作模式相似的活动时表现出更强的鲁棒性与优越性。

4　总结

针对现有基于Wi-Fi CSI的人体行为活动识别方法在特征判别能力上的不足，提出一种融合双注意力机制与多尺度卷积神经网络的识别方法。通过系统化的数据采集与预处理流程，结合子载波注意力和通道注意力机制，强化了模型对关键频段和高判别性特征的提取能力，并利用多尺度卷积融合不同粒度时空信息，有效提升了复杂活动和细微动作的辨识性能。在多个公开多类别活动数据集上的实验结果表明，所提方法在多种活动识别任务中的平均准确率显著优于主流基线模型，尤其在动作相似度高和易受噪声干扰的静态动作识别中表现出优越的鲁棒性，验证了所提方法在特征学习和分类性能上的先进性与有效性。

参考文献：

[1]　王杨，许佳炜，王傲，等. 融合交叉序列预测和一致性对比的Wi-Fi人体活动识别[J]. 计算机工程与科学，2025，47(1)：160-170.

[2]　Beddiar D R，Nini B，Sabokrou M，et al. Vision-based human activity recognition：a survey[J]. Multimedia Tools and Applications，2020，79(41)：30509-30555.

[3]　Francisco M. Calatrava-Nicolás，Oscar Martinez Mozos. Light Residual Network for Human Activity Recognition using Wearable Sen-

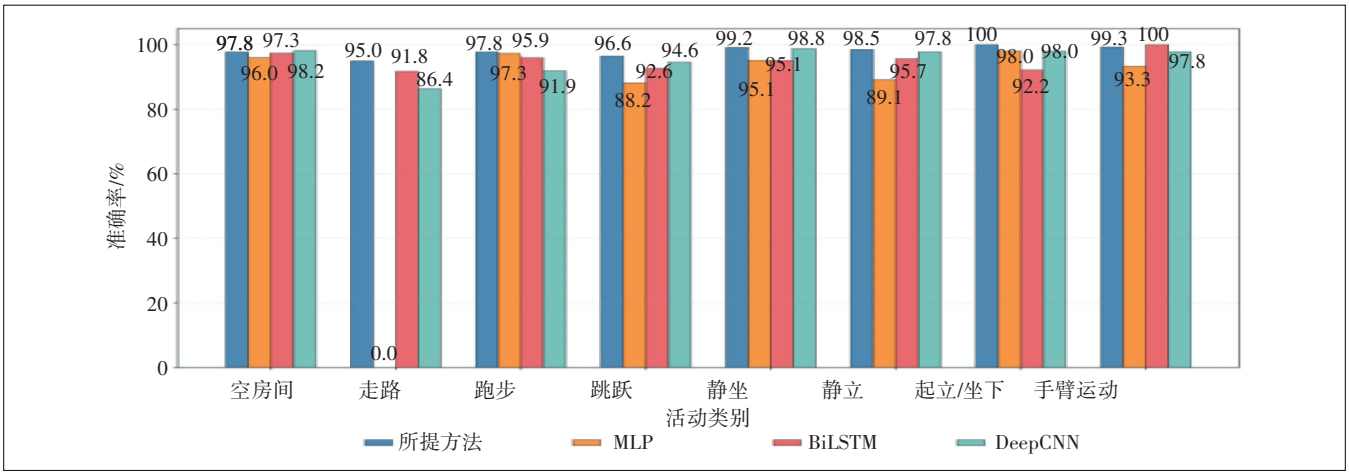


图8　4种识别方法在每个分类任务下的准确率对比

sor Data[J]. IEEE Sensors Letters, 2023, 7(10): 1-4.

[4] Ma Y, Zhou G, Wang S. Wi-Fi Sensing with Channel State Information: A Survey[J]. ACM Computing Surveys, 2020, 52(3): 1-36.

[5] Yang Z, Zhou Z, Liu Y. From RSSI to CSI: Indoor localization via channel response[J]. ACM Computing Surveys, 2013, 46(2): 1-32.

[6] Kotaru M, Joshi K, Bharadia D, et al. SpotFi: Decimeter Level Localization Using WiFi[J]. Computer Communication Review: A Quarterly Publication of the Special Interest Group on Data Communication, 2015, 45(4): 269-282.

[7] Jawad S K, Musaab Alaziz M. Human Activity and Gesture Recognition based on Wi-Fi[C]//2022 Iraqi International Conference on Communication and Information Technologies (IICCIT), Basrah, Iraq, 2022: 268-273.

[8] Liu X, Chen H, Jiang X, et al. Wi-CR: Human Action Counting and Recognition with Wi-Fi Signals[C]//2019 4th International Conference on Computing, Communications and Security (ICCCS), Rome, Italy, 2019: 1-8.

[9] Danista Khan, Ivan Wang-Hei Ho. Deep Learning of CSI for Efficient Device-free Human Activity Recognition[C]//2021 IEEE 7th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2021: 19-24.

[10] Ding J, Wang Y. Wi-Fi CSI-Based Human Activity Recognition Using Deep Recurrent Neural Network[J]. IEEE Access, 2019, 7: 174257-174269.

[11] Niu J, He X, Fang B, et al. CSI-F: A Human Motion Recognition Method Based on Channel-State-Information Signal Feature Fusion[J]. Sensors, 2024, 24(3): 23.

[12] Gringoli F, Schulz M, Link J, et al. Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets[C]//Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH '19), New York, NY, USA, 2019: 21-28.

[13] Xin Q, Hu S H, Liu S Q, et al. An Attention-Based Wavelet Convolution Neural Network for Epilepsy EEG Classification[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 957-966.

[14] Yang H, Yuan C, Zhang L, et al. STA-CNN: Convolutional Spatial-Temporal Attention Learning for Action Recognition[J]. IEEE Transactions on Image Processing, 2020, 29: 5783-5793.

[15] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 11531-11539.

[16] Yang J, Chen X, Zou H, et al. EfficientFi: Toward Large-Scale Lightweight Wi-Fi Sensing via CSI Compression[J]. IEEE Internet of Things Journal, 2022, 9(15): 13086-13095.

[17] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, et al. A Survey on Behavior Recognition Using Wi-Fi Channel State Information[J]. IEEE Communications Magazine, 2017, 55(10): 98-104.

[18] Meneghello F, Nicolò Dal Fabbro, Garlisi D, et al. A CSI Dataset for Wireless Human Sensing on 80 MHz Wi-Fi Channels[J]. IEEE Communications Magazine, 2023, 61(9): 146-152.

[19] Yang J, Chen X, Wang D, et al. SenseFi: A library and benchmark on deep-learning-empowered Wi-Fi human sensing[J]. Patterns, 2023, 4(3): 2666-3899.

[20] Meridja Karim, Sekkiou Imène, Gammoudi Aymen. Real-Time Human Activity Recognition Using Wi-Fi CSI: A Deep CNN Architecture Approach[C]//2024 IEEE Thirteenth International Conference on Image Processing Theory, Tools and Applications (IPTA), Rabat, Morocco, 2024: 1-6.

作者简介：

赵纯熙，工程师，硕士，主要从事网络创新产品、云内监控产品的研发设计工作；袁毅，学士，主要从事6G无线通信系统设计与优化方面研究工作；刘懿，学士，主要从事6G无线通信系统设计与优化方面研究工作。