

AM-WiGeR: 一种基于注意力机制增强残差网络的 Wi-Fi 手势识别方法



AM-WiGeR: A Wi-Fi Gesture Recognition Method Based on Attention Mechanism Enhanced Residual Network

李业深/Li Yeshen, 姚可欣/Yao Kexin, 熊轲/Xiong Ke

(北京交通大学, 中国 北京 100044)
(Beijing Jiaotong University, Beijing 100044, China)

DOI: 10.12142/ZTETJ.202604012

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20260902.0959.002>

网络出版日期: 2026-08-24

收稿日期: 2026-03-06

摘要: 提出了一种基于注意力机制增强残差网络的 Wi-Fi 手势识别方法。该方法从 Wi-Fi 设备采集的信道状态信息 (CSI) 中提取多普勒频移 (DFS) 作为特征源, 通过共轭矩阵相乘、天线对选择及相位偏移消除等方法, 从中提取能有效表征手势信息的向量参数, 并进一步通过短时傅里叶变换获得对应的 DFS 频谱图特征。然后, 以 DFS 频谱图特征作为 AM-RES 模型的输入, 进行人体手势识别。实验结果表明, 相较于现有方法, 所提 AM-WiGeR 方法对不同朝向手势识别的平均准确率达到 96.22%, 取得了良好的手势识别效果。

关键词: 手势识别; CSI; DFS; 注意力机制

Abstract: A Wi-Fi gesture recognition method based on Attention Mechanism Enhanced Residual Network (AM-RES) is proposed, termed AM-WiGeR (AM-RES based Wi-Fi Gesture Recognition). Channel state information (CSI) collected from widely deployed Wi-Fi devices is utilized, from which Doppler frequency shift (DFS) is extracted as the feature source. Through conjugate matrix multiplication, antenna pair selection, and phase offset elimination, vector parameters that can effectively characterize gesture information are extracted from the DFS, and the corresponding DFS spectrogram features are further obtained via short-time Fourier transform. The DFS spectrogram features are then fed into the AM-RES model for human gesture recognition. Experimental results demonstrate that, compared with existing methods, the proposed AM-WiGeR method achieves an average recognition accuracy of 96.22% for gestures performed in different orientations, indicating satisfactory gesture recognition performance.

Keywords: gesture recognition; CSI; DFS; attention mechanism

引用格式: 李业深, 姚可欣, 熊轲. AM-WiGeR: 一种基于注意力机制增强残差网络的 Wi-Fi 手势识别方法 [J]. 中兴通讯技术, 2026, 32(4): 79-87. DOI: 10.12142/ZTETJ.202604012

Citation: Li Y S, Yao K X, Xiong K. AM-WiGeR: A Wi-Fi gesture recognition method based on attention mechanism enhanced residual network [J]. ZTE technology journal, 2026, 32(4): 79-87. DOI: 10.12142/ZTETJ.202604012

新一代信息技术的持续演进, 推动了各类智能应用的发展, 这对人体动作及行为识别的需求日趋迫切。在诸多新兴人机交互技术中, 手势动作识别技术能够为用户提供和谐、简便、友好的操作体验^[1], 已成为当前主流的人机交互方式之一, 在智能家居、智慧商场等领域具有广阔的应用前景。

传统的手势识别方法可分为 3 类: 基于计算机视觉的方法、基于穿戴式传感器的方法和基于红外感知的方法^[2]。这 3 种技术虽各有优势, 但也均存在局限性。例如, 基于计算机视觉的方法受光照和遮挡影响较大, 且存在较高的用户隐私

泄露风险; 基于穿戴式传感器的方法要求用户持续佩戴传感器设备, 安装与穿戴不够便捷, 影响了用户的使用体验; 基于红外感知的方法需要布设专业精密设备, 增加了应用成本, 同时其感知范围有限, 难以满足大规模应用的需求^[3]。

通感一体化技术近年来发展迅速, 无线信号因同时具备通信与感知能力而受到极大关注^[4-8]。利用无线信号实现对人体动作的非接触式感知, 正逐步取代传统识别技术^[9-11], 并已引起广泛关注。

现有基于无线信号的动作识别方法, 按信号类型的不同,

可分为基于超宽带 (UWB)、基于射频识别 (RFID) 和基于 Wi-Fi 的方法等^[12]。基于 UWB 的方法需要部署大量较为昂贵的 UWB 设备, 部署成本相对较高; 基于 RFID 的方法要求识别目标随身携带标签, 而标签具有一定体积, 使用不够便捷, 且相对容易损坏; 相较而言, 基于 Wi-Fi 的动作识别方法可利用现有已广泛部署的 Wi-Fi 无线路由器进行工作, 无须额外布设专用设备; 采用非接触感知方式实现感知, 用户体验良好; 相较于基于计算机视觉的方法, 不受光照等环境条件影响, 且无须采集人体图像数据, 对用户隐私的泄露风险较低。因此, 基于 Wi-Fi 的动作识别已发展成为新的研究热点^[13]。

Wi-Fi 信号传播对环境感知灵敏, 人体动作会改变 Wi-Fi 信号的直射、反射和散射路径。当 Wi-Fi 覆盖区域内的人体手势等动作发生变化时, Wi-Fi 接收器 (如电脑、手机等) 所接收信号的强度、到达角度等参数也会发生相应变化。因此, 通过分析接收端数据变化与人体动作变化之间的作用规律, 即可实现对人体动作的识别。目前, 常见的基于 Wi-Fi 的动作识别方法通常基于接收信号强度指示 (RSSI) 或信道状态信息 (CSI)。RSSI 记录了接收端无线电信号的平均功率, 能够提供关于通信链路的相对粗粒度特征^[14]; CSI 则记录了每个 Wi-Fi 子载波信道的幅度和相位信息, 能够提供关于通信链路的相对细粒度特征^[15]。2010 年, Halperin 等通过改造 Wi-Fi 设备固件, 实现了对 IEEE 802.11n 标准无线网卡上 CSI 的获取, 从而推动了基于 CSI 的人体动作识别研究^[16-18]。

传统基于 CSI 的手势识别通常采用机器学习方法, 对从 CSI 中提取的统计特征或物理特征进行分类^[9], 从而实现对人体手势的识别。例如, 文献[20]从 CSI 中提取幅值、相位和信号到达角度 3 种信息作为特征, 利用 K 近邻 (KNN) 方法进行分类, 实验结果表明该方法对 5 种动作的识别准确率达到 92.6%。传统基于 CSI 的手势识别方法的不足之处在于, 需要手动设计和提取结构化数据特征作为机器学习算法的输入, 而这类方法通常采用信号变换等传统手段提取易于表示的统计特征, 导致所提取特征的维度和适用范围较为有限。深度学习是一种具有强大特征表达能力的机器学习方法, 其输入可以直接为非结构化数据。深度学习能够通过神经网络模型自动提取输入数据的特征, 从而实现更高效、更精确的识别效果。因此, 近期越来越多的研究开始利用深度学习^[21]开展 Wi-Fi 人体手势识别研究。文献[22]通过设计手势速度关系模型, 利用卷积神经网络 (CNN) 提取空间特征, 并利用门控循环单元 (GRU) 提取时间特征, 将提取到的动作速度信息作为特征输入神经网络模型, 其识别准确率可达 92.7%。文献[23]则以 CSI 相位差矩阵作为特征输入, 采用双向门控循环网络从前后两个方向对相位差序列进行自动学习和特征提取,

实现对 4 种手势的识别, 其准确率达 94%。

综合上述相关研究发现可知, 现有利用 CSI 进行手势识别的研究虽已取得大量有价值的成果, 但仍存在若干亟待解决的问题。例如, 利用 CSI 的物理特征和统计特征表征动作时, 无法有效消除环境因素的影响, 导致特征难以准确表示手势信息; 再如, 现有深度学习模型在训练时利用输入特征的全局信息, 难以有效聚焦于与手势动作相关的特征, 从而导致识别效果不佳^[24]。

针对上述问题, 本文首先提出了一种基于注意力机制增强残差网络 (AM-RES) 模型。该模型采用具有深层网络结构且不易退化的残差网络作为骨干网络, 同时引入注意力机制以自动捕捉与手势动作相关的特征, 从而实现更高的识别准确率。在此基础上, 提出了一种基于 AM-RES 的手势识别方法, 即 AM-WiGeR。该方法利用现有广泛部署的 Wi-Fi 设备采集 CSI 数据, 通过共轭乘法、天线对选择和相位偏移消除等手段, 从中提取与动作相关性更高的多普勒频移 (DFS) 信息, 并进一步通过短时傅里叶变换得到对应的 DFS 频谱图特征。然后, 以 DFS 频谱图特征作为输入, 利用所提出的 AM-RES 模型进行人体手势识别。最后, 在自测数据集上对所提方法进行实验验证。该方法能够取得良好的手势动作识别效果, 对不同朝向手势识别的平均准确率达到 96.22%。

1 基于 CSI 手势识别原理

在室内环境中, 人的手动动作会引起 Wi-Fi 信号传播路径和传播特性的改变, 具体表现为接收信号中提取的 CSI 特征发生相应变化。不同的手动动作会对无线信道的 CSI 产生不同的影响。因此, 可利用观测到的 CSI 特征来分析引起该 Wi-Fi 信号变化的人体动作, 从而实现基于 Wi-Fi 信号的手势动作识别。本节将介绍 CSI 的基本概念及 Wi-Fi 信号的传播模型。

1.1 CSI

在无线通信系统中, CSI 是通信链路的信道属性, 能够描述信道衰落、散射和传播距离等因素的综合影响。Wi-Fi 采用多输入多输出 (MIMO) 和正交频分复用 (OFDM) 技术来发送和接收数据包, CSI 可在不同天线对和子载波上测量获得。在 t 时刻第 k 个子载波的无线信道可定义为:

$$\begin{bmatrix} Y_1[k, t] \\ \vdots \\ Y_n[k, t] \end{bmatrix} = \begin{bmatrix} H_{11}[k, t] & \cdots & H_{1m}[k, t] \\ \vdots & \ddots & \vdots \\ H_{n1}[k, t] & \cdots & H_{nm}[k, t] \end{bmatrix} \begin{bmatrix} X_1[k, t] \\ \vdots \\ X_m[k, t] \end{bmatrix} \quad (1),$$

其中, X 为发射端信号, Y 为接收端信号, H 是表示 CSI 信息的信道矩阵, m 表示发射天线数量, n 表示接收天线的数量。

对于每个子载波其CSI值是一个复数,实部表示CSI幅度信息,虚部表示相位信息。因此,当一个人在Wi-Fi信号的传播环境中做出动作时,会对接收端的信号特性产生一定影响,并反映在CSI的幅值和相位信息中。

1.2 Wi-Fi信号传播模型

为了有效地描述由手势动作引起的Wi-Fi信号变化,接下来我们将介绍Wi-Fi信号如何传播以及动作影响无线信道的原因。如图1所示,手势动作识别系统由发射端(X)、接收端(Y)和用户组成。

Wi-Fi信号在室内传播时,由于受到墙壁、地面、人体等的阻碍,发生反射、散射、衍射等现象,这使得接收端收到来自不同路径的多个信号,称为多径效应。在多径效应下接收端的CSI表示为:

$$h(f, t) = \sum_{i=1}^L A_i(t) e^{-j2\pi f \tau_i(t)} \quad (2),$$

其中, f 是信号的原始载波频率, L 表示总的路径条数, $A_i(t)$ 是第 i 条路径的幅度衰减, $\tau_i(t)$ 是第 i 条路径的传播延迟。

多普勒效应是指当发射端和接收端之间存在相对运动时,接收端接收到信号的频率与发射端发射信号的频率不同,两者之差称为多普勒频移。当人体执行手势动作时,会间接改变Wi-Fi信号发射端与接收端之间的相对距离,从而产生多普勒效应。DFS的大小与路径长度变化速率相关,可表示为:

$$f_D = f \frac{v_{\text{path}}}{c} \quad (3),$$

其中, v_{path} 是路径长度变化的速度, c 是Wi-Fi信号的传播速度。

因此,CSI可以用DFS表示为:

$$h(f, t) = \sum_{i=1}^L A_i(t) e^{-j2\pi f \left(\frac{v_i t}{c} \right)} = \sum_{i=1}^L A_i(t) e^{-j2\pi f_{D(i)} t} \quad (4),$$

其中, v_i 表示第 i 条路径长度变化速度, $f_{D(i)}$ 表示第 i 条路径下由动作产生的多普勒频移。如图1所示,在多径效应下,Wi-Fi信号到达接收端的路径可划分为静态路径和动态路径。静态路径是指由LoS路径以及静止物体反射的路径,其DFS值为0。动态路径是指由手势动作反射的路径,动作会引起路径长度随时间发生变化,导致DFS值不为0。因此,CSI值可以分为静态分量和动态分量:

$$h(f, t) = h_s(f, t) + \sum_{i \in G} A_i(t) e^{-j2\pi f_{D(i)} t} \quad (5),$$

其中, $h_s(f, t)$ 为静态分量, G 表示动态路径。因此,可以从CSI中提取动态分量,去除环境干扰信息,并通过时频分析进一步提取出DFS频谱图特征,更有效地表示手势动作。

2 AM-WiGeR 方法

本文提出的AM-WiGeR手势识别方法具有以下两个特点:在特征选取方面,提取DFS频谱图特征,能够降低环境因素的干扰,更有效地表征手势动作;在深度学习模型方面,提出结合注意力机制的残差网络模型AM-RES进行手势识别,从而有效提升识别准确率。

如图2所示,所提AM-WiGeR方法由4个功能部分组成,分别为数据采集、数据预处理、特征提取和动作识别。数据采集部分利用Wi-Fi收发设备采集CSI数据;数据预处理部分通过对原始CSI数据进行滤波器去噪、相位偏移消除及主成分分析,获取预处理后的CSI;特征提取部分采用静态分量消除、天线对选择和短时傅里叶变换等方法提取DFS频谱图;动作识别部分以DFS频谱图作为输入,利用AM-RES实现手势识别。本节将介绍这4个功能部分的工作原理与具体实现。

2.1 数据采集

所提AM-WiGeR方法的第1部分是数据采集模块,负责CSI数据的获取,涉及的硬件装置包括1个Wi-Fi发射端和1个Wi-Fi接收端。需要说明的是,随着Wi-Fi采用MIMO和OFDM技术,几乎在所有支持IEEE 802.11n标准的无线网卡上都能获取到CSI。因此,Wi-Fi手势识别的发射端可采用常见的支持IEEE 802.11n标准的路由器,接收端则为配有多根天线的支持IEEE 802.11n标准的无线网卡。相较于传统基于单天线单链路的手势识别方法,采用多根天线接收CSI,所采集的数据能够更完整地保留手势特征。

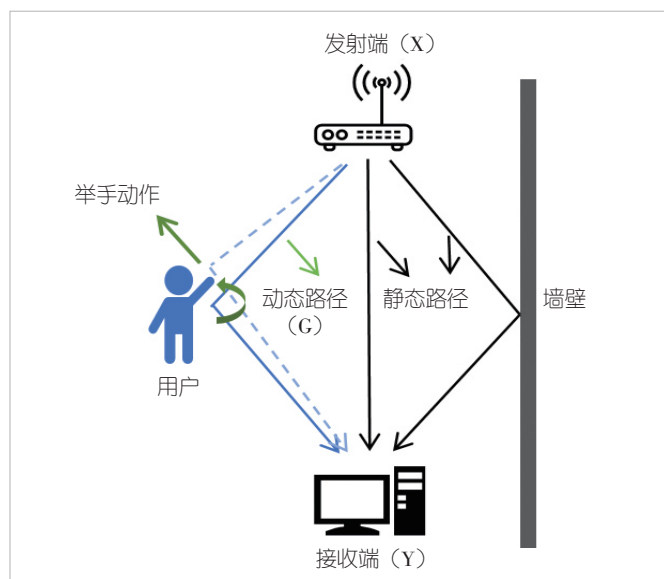


图1 室内环境Wi-Fi信号传播示意图

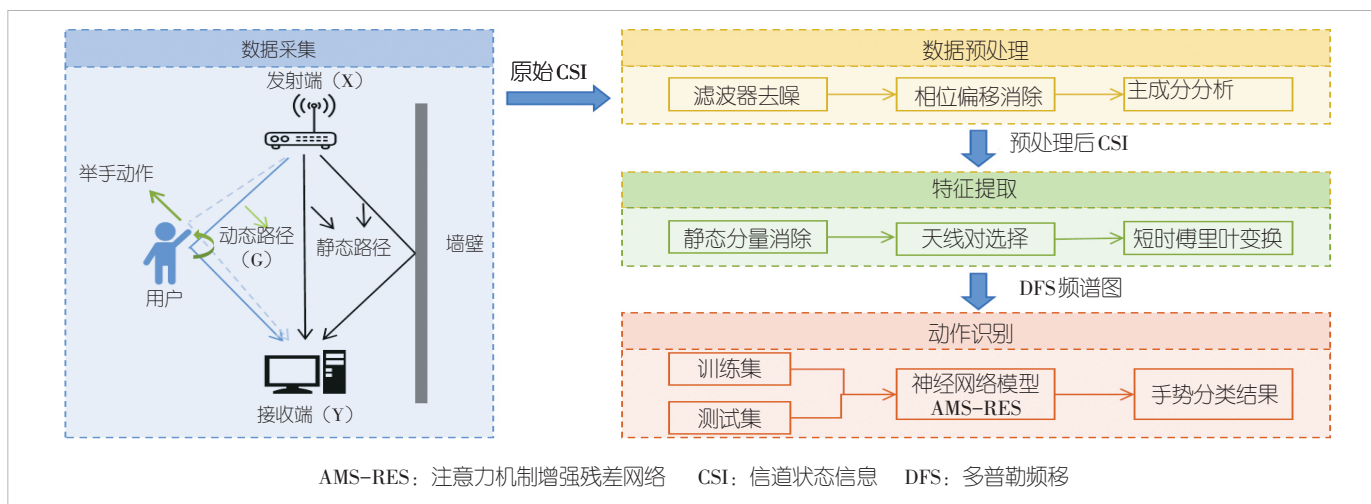


图2 AM-WiGeR整体结构图

2.2 数据预处理

1) 滤波器去噪

室内环境中存在与动作无关的背景噪声及由干扰导致的异常值,因此采集到的原始CSI数据无法直接用于动作识别,需先进行去噪处理。巴特沃斯滤波器能够平滑通频带内的频率曲线,且设计简单、性能良好,相较于其他方法,可有效消除背景噪声和异常值,同时保留手势相关信息。因此,所提AM-WiGeR方法采用巴特沃斯滤波器对数据进行去噪处理。图3给出了两根天线 r_1 、 r_2 上第15个子载波的原始CSI数据及经巴特沃斯滤波处理后的CSI数据。可以看出,经滤波处理后,波形更加平滑,噪声得到有效抑制。

2) 相位偏移消除

由于Wi-Fi收发设备之间的时钟不同步,CSI的相位存在时变随机偏移,不能直接用于DFS特征的提取。含有相位偏移的CSI可以表示为:

$$h(f, t) = e^{-j\theta_{\text{offset}}} \left(h_s(f, t) + \sum_{i \in G} A_i(t) e^{-j2\pi f D_{(i)}} \right) \quad (6).$$

由于接收端的网卡配有多根天线,这些天线的相位偏移相同,所以通过对接收端的两根天线的CSI矩阵做共轭乘法,可有效消除相位偏移。

3) 主成分分析

IEEE 802.11n标准采用多载波传输方式,接收端可收集包含多个子载波的CSI。由于CSI数据维度较高,易导致特征信息冗余,不利于后续特征提取,因此需对数据进行降维处理。主成分分析(PCA)能够在保留绝大部分有效信息的前提下对数据进行降维,提升对样本中有效信息的处理效率。因此,所提AM-WiGeR方法采用PCA分解得到第1主成分。由于第1

主成分包含了由动作引起的主要功率变化的大部分信息,可将其用于后续DFS特征的提取。

2.3 特征提取

经过上述步骤得到的数据仍包含与手势无关的冗余信息,

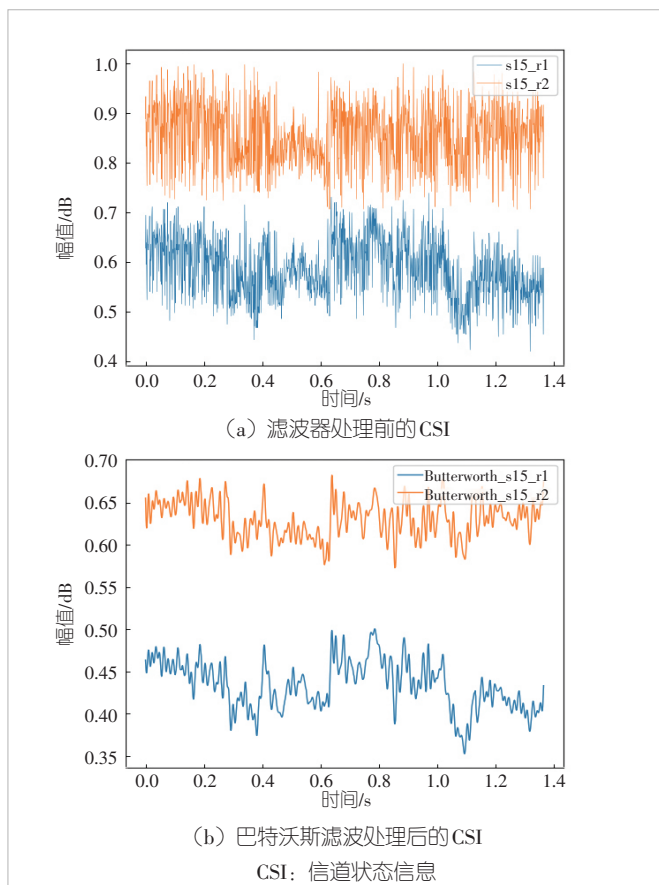


图3 巴特沃斯滤波器去噪前后CSI波形图

需进一步进行特征提取, 以利于后续手势识别。特征提取是指将原始信息转换为新的、更有效的数据描述形式, 这些形式更易于被动作识别模型有效利用。首先, 对预处理后的 CSI 进行静态分量消除和动态分量放大, 以提取 DFS 信息; 然后, 利用短时傅里叶变换将 DFS 特征信息可视化, 生成 DFS 频谱图, 作为深度学习算法的稳定输入特征。下面将详细介绍特征提取步骤。

1) 静态分量消除

由公式 (5) 知, CSI 可表示成静态分量和动态分量的和。由于 DFS 只存在于 CSI 的动态分量中, 因此需要对 CSI 进行分解, 消除静态分量, 在动态分量中提取 DFS 特征。

首先分解预处理后的 CSI 信息为静态分量和动态分量:

$$h_{cm}(f, t) = h_1(f, t)h_2(f, t)^* = \quad (7),$$

$$(h_{1,s}(f, t) + \sum_{i \in G_1} A_i(t)e^{-j2\pi f_{D(i)}})(h_{2,s}(f, t)^* + \sum_{k \in G_2} A_k(t)e^{j2\pi f_{D(k)}}) = h_{1,s}(f, t)h_{2,s}(f, t)^* + \quad (8),$$

$$h_{1,s}(f, t) \sum_{k \in G_2} A_k(t)e^{j2\pi f_{D(k)}} + \quad (9),$$

$$h_{2,s}(f, t)^* \sum_{i \in G_1} A_i(t)e^{-j2\pi f_{D(i)}} + \quad (10),$$

$$\sum_{i \in G_1, k \in G_2} A_i(t)A_k(t)e^{j2\pi t(f_{D(k)} - f_{D(i)})} \quad (11),$$

其中, 公式 (11) 为 2 根天线的动态分量之积, 其值非常小, 可忽略; 公式 (9) 和式 (10) 为 1 根天线的静态分量与另 1 根天线的动态分量之积, 其中包含所需的 DFS 信息; 公式 (8) 为 2 根天线的静态分量之积。静态分量中不含 DFS 信息, 但含有直接路径信号, 导致其功率较高。通过从预处理后的 CSI 中减去其平均值来消除静态分量, 可避免静态分量对 DFS 估计的干扰。

2) 天线对选择

公式 (9) 和 (10) 所表示的动态分量中包含所需的 DFS 信息, 但由于 2 根相邻天线具有相似的多径, 导致所包含的 DFS 信息幅值相近但方向相反。若直接利用上述 2 个动态分量计算 DFS, 会导致两者相互抵消。因此, 为从动态分量中有效提取 DFS, 需设计天线对选择策略, 以放大其中 1 个动态分量 (公式 (9)) 并缩小另 1 个动态分量 (公式 (10))。由文献[26]可知, CSI 均值较大的天线具有较高的静态分量, CSI 方差较大的天线具有较高的动态分量。因此, 在选择用于共轭矩阵相乘的 2 根天线时, 第 1 根天线选择均值与方差之比最大

的天线 $\text{Max}(\frac{E(h_i)}{\text{var}(h_i)})$, 第 2 根天线选择均值与方差之比最小

的天线 $\text{Min}(\frac{E(h_i)}{\text{var}(h_i)})$ 。通过上述天线选择步骤, 即可获得

DFS 向量参数。

3) 短时傅里叶变换

经过上述步骤, 从 CSI 中得到的 DFS 特征为大小为 $m \times n \times k \times t$ 的 4 维向量, 其中 m 代表发射天线数, n 代表接收天线数, k 代表子载波数, t 代表时间戳。为提取能够更直观体现手势信息的特征, 对 DFS 向量参数进行短时傅里叶变换 (STFT), 并生成 DFS 频谱图。传统的快速傅里叶变换 (FFT) 仅能反映信号在频域的特性, 无法在时域对信号进行分析。而 STFT 能够从时间和频率两个维度描述手势对 DFS 的影响。STFT 通过在时间轴上移动窗函数, 选用高斯窗函数截取信号片段, 并对每个片段进行 FFT 操作, 从而获得 DFS 特征, 即 DFS 频谱图。图 4 给出了两种不同动作的 DFS 频谱图, 其中, 图 4 (a) 为推拉手势对应的 DFS 频谱图, 图 4 (b) 为画圈手势对应的 DFS 频谱图。由图 4 可知, 两种动作下 DFS 的变化存在明显差异。因此, DFS 频谱图可作为动作识别的数据特征, 以用于区分不同手势。

2.4 动作识别

1) 注意力机制

动作识别部分是 AM-WiGeR 的核心模块。目前大多数动作识别方法采用深度学习技术, 通过神经网络模型自动提取输入数据特征, 实现手势的分类识别。然而, 现有神经网络模型大多利用输入特征的全局信息进行训练, 难以有效聚焦于与手势相关的特征信息。为此, 本文提出了基于注意力机制增强的残差网络模型 AM-RES。该模型采用经典的残差网络 ResNet-18 作为主干网络, 如图 5 所示, 其扩展部分在于引入了注意力机制, 能够依据权重值判断各信息的重要程度, 自动提取与手势相关的特征, 并实现手势识别。

在神经网络的学习过程中, 注意力机制能够对当前输入特征信息的重要性进行评估, 得到对应的权重值, 并在后续网络学习过程中依据权重值判断各信息的重要程度, 从而决定保留或丢弃相应信息。图 6 给出了注意力机制的基本结构。对于输入的图像特征, 注意力机制生成对应的注意力特征:

$$W = f(P_f, C) \# (2 - 7) \quad (12),$$

$$C' = C \otimes W \quad (13),$$

其中, P_f 是参数矩阵, 通过一系列函数 f 的作用得到注意力权

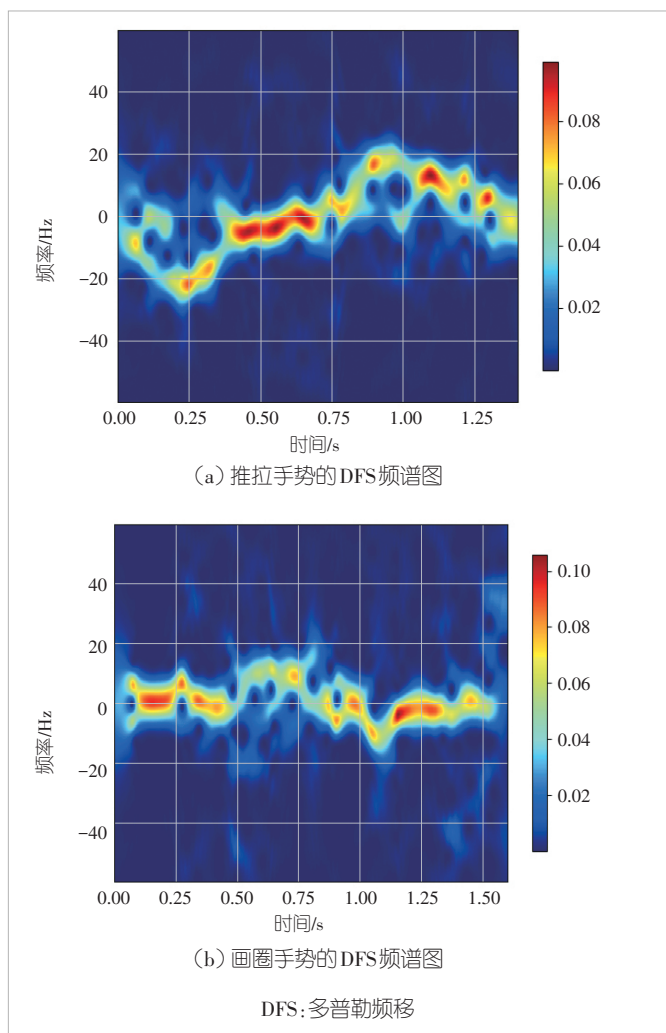


图4 推拉和画圈手势的DFS频谱图

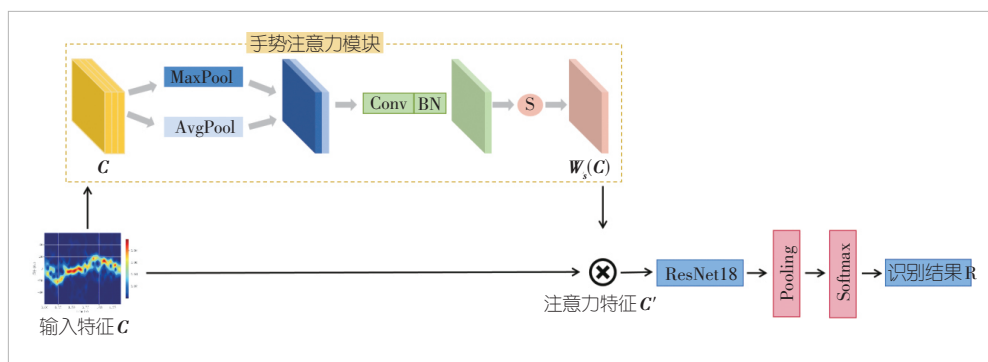


图5 本文提出的AM-RES手势识别模型网络架构

值矩阵 W , W 中的值对应不同像素在 C 中的权重。权重越大, 对应的像素越重要。然后将 C 与 W 按元素乘, 得到注意力特征 C' 。

2) AM-RES模型

模型输入特征为DFS频谱图 C , 本文引入的注意力模块为

C 生成注意力权重矩阵 $W_s(C)$, 其计算公式为:

$$W_s(C) = \text{sigmoid}\left(\text{BN}\left(\text{Conv}\left(\text{cat}\left(\text{AvgPool}(C), \text{MaxPool}(C)\right)\right)\right)\right)(14),$$

其中, Conv表示卷积运算, BN表示批归一化运算, AvgPool和MaxPool分别代表平均值池化和最大值池化操作。DFS频谱图 C 首先通过组合池化层进行特征降维。组合池化层包括最大值池化与平均值池化, 结合二者优势同时关注重要的局部特征和全局特征, 丰富了特征信息。然后进行卷积运算、批归一化运算, 通过Sigmoid激活函数计算出 $W_s(C)$ 后, 将其与 C 相乘以进行自适应特征优化, 得到注意力特征 C' 。

$$C' = W_s(C) \otimes C \quad (15),$$

然后通过预训练的ResNet18对注意力特征进行处理, ResNet18采用残差网络(ResNets)结构, 如图7所示。

在不同卷积层之间添加跨层连接, 使前面卷积层的数据可越过中间层直接传入更深的卷积层, 从而有效缓解梯度消失问题, 并提取更深层的特征信息。最后, 通过池化层和Softmax层获得识别结果, 实现对6种手势的分类识别。训练过程中, 采用交叉熵损失函数对模型分类结果与真实标签进行计算, 得到分类损失值。神经网络基于反向传播算法更新参数, 并持续迭代直至模型收敛。通过上述方式, 注意力模块能够自适应地学习如何生成准确的注意力特征, 促使网络更多地关注与手势相关的信息。同时, 预训练的ResNet-18网络可有效提取深层特征, 从而获得更优的分类效果。

3 实验设置与结果分析

3.1 实验环境及设置

本文实验在北京交通大学网络实验室完成, 实验场景示意图及实际部署情况如图8所示。实验室面积约为 $6\text{ m} \times 3.5\text{ m}$, 室内摆放有桌子、椅子, 以及显示器、电脑等实验设备若干。实验硬件设备具体包括Wi-Fi信号发送端和接收端, 其中发射端为支持IEEE 802.11n协议的TP-LINK TL-WR886N型号、带有3根天线的路由器, 接收端为配备3根天线的Intel 5300网卡的主机, 操作系统为Ubuntu 16.04。Wi-Fi发射端与接收端的间距为2 m, 距离地面的高度均为0.8 m, 实验人员坐在收发链路中间执行预设

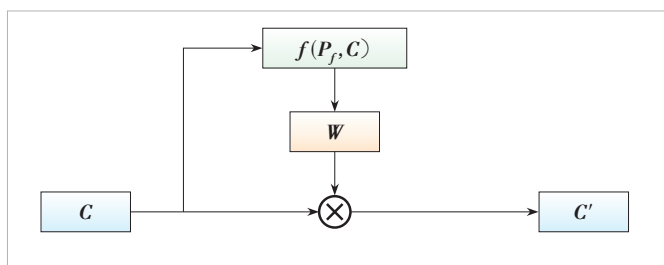


图6 注意力机制的基本结构

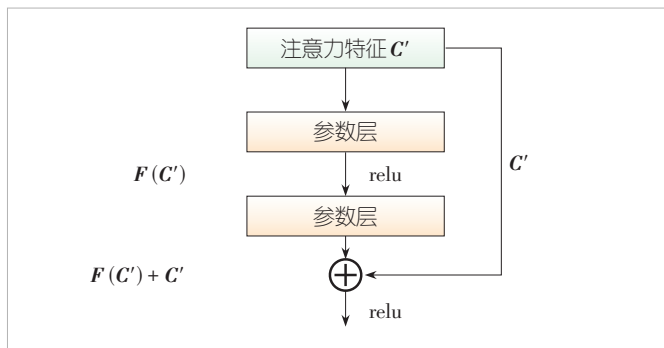


图7 残差结构示意图

手势动作。

3.2 实验数据采集

实验采集了室内环境下实验人员执行不同手势动作时接收端的CSI数据。通过Halperin等设计的CSI Tools软件采集接收端CSI数据，采样频率设置为1 kHz。实验采集的手势包括6种常用的人机交互动作，分别为拍手、举手、平扫、推拉、画圈和交叉。在采集数据时，为避免人体位于收发链路之间因朝向不同而产生的差异，实验人员分别朝向自身的正前方、右前方和左前方3个方向执行手势动作。每个动作执行时间为2s，每种动作各执行900次，最终收集5 400组数据。在深度学习模型训练过程中，80%的数据为训练集，20%为测试集。

3.3 实验结果与分析

本文使用评价指标分为识别准确率和混淆矩阵。

1) 准确率验证与对比实验

(1) 模型准确率验证

在手势识别问题中，手势识别准确率，即分类正确的样本数占总样本数的比例，是最重要的评判指标之一，能有效反映模型的识别效果。图9给出了本文所提AM-WiGeR方法在训练和测试过程中对6种手势识别的平均准确率，其中横轴表示迭代轮次，纵轴表示准确率。最终，模型在测试集上的平均识别准确率达到96.22%。

(2) 不同特征对识别准确率影响

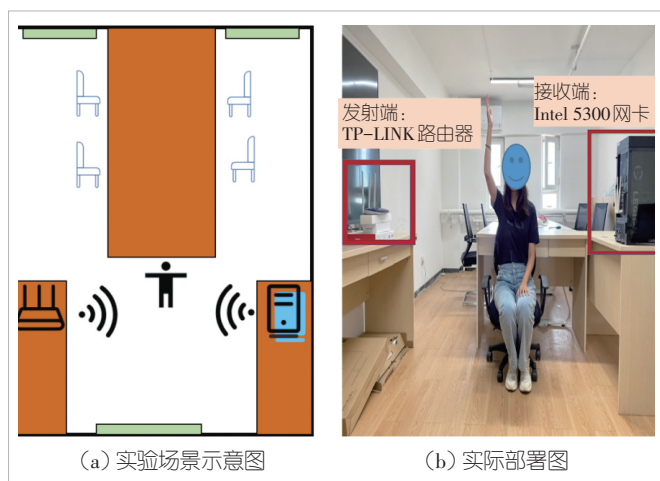


图8 Wi-Fi手势识别实验场景

为探索不同特征对识别准确率的影响，分别将CSI幅值、CSI相位、CSI幅值与相位的融合信息以及所提AM-WiGeR方法提取的DFS频谱图作为模型输入，并对比其识别准确率，结果如图10所示。

由图10所示的实验结果可知，采用DFS频谱图作为特征时识别准确率最高，这验证了所提DFS特征提取方法的有效性。采用CSI相位特征时识别效果最差，这是由于CSI相位存在固有偏移，仅依靠相位信息难以获得理想的识别效果。与CSI幅值和相位信息相比，DFS能够消除环境中静态路径的干扰，同时有效保留动态路径的CSI信息。因此，采用DFS频谱图作为输入特征时，手势识别准确率更高。

(3) 不同识别算法的准确率比较

本文还对比了不同识别算法的效果。除本文所提AM-RES模型外，还对比了未引入注意力机制的ResNet-18模型，以及文献[23]中WiHGR算法和文献[22]中Widar3.0算法。将这

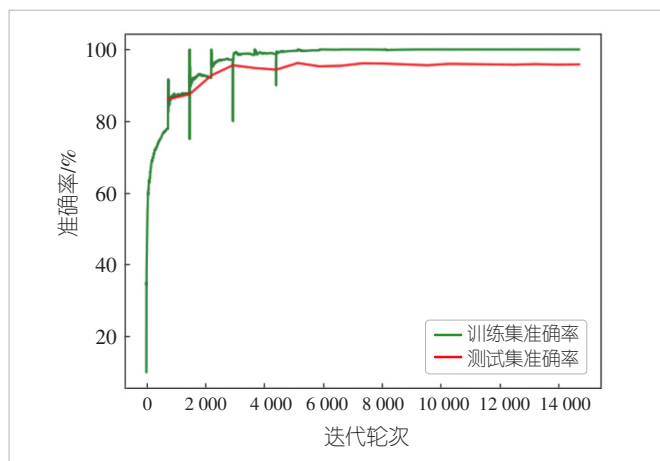


图9 AM-WiGeR方法识别手势的平均准确率

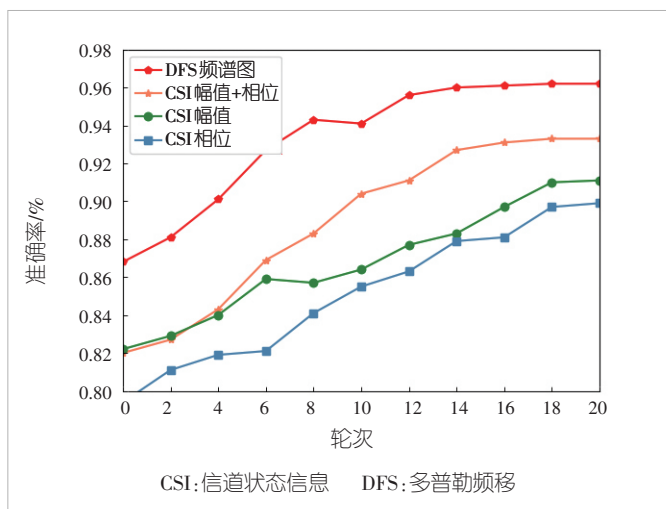


图10 不同特征的识别准确率

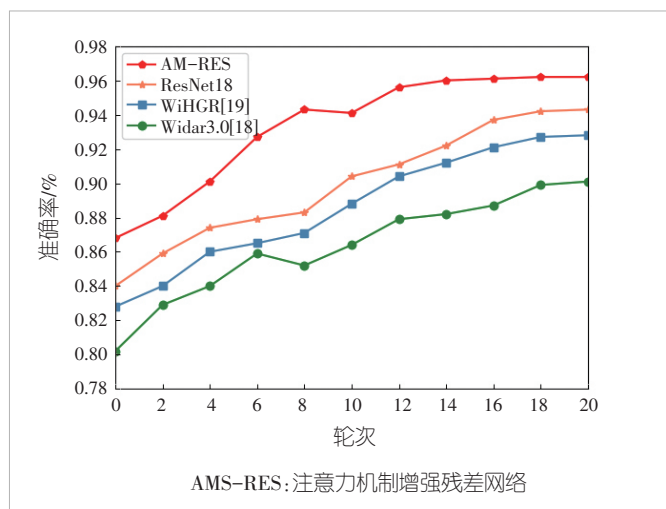


图11 不同识别方法下准确率效果

4种算法在相同数据集上进行实验,并对比其识别准确率,结果如图11所示。

由图11所示结果可知,相较于其他算法,本文所提AM-RES算法的识别准确率显著提升,且模型收敛速度更快,验证了所提算法的有效性。文献[22]中Widar3.0算法的识别准确率为89.7%。文献[23]中WiHGR算法采用双向循环网络,从前两个方向对特征序列进行学习,但该网络结构导致模型收敛速度较慢,训练耗时较长。本文所提AM-RES模型引入注意力机制,依据权重值判断各信息的重要程度,提取DFS频谱图的深层特征,相较于未引入注意力机制的ResNet-18模型,识别准确率有显著提升。同时,采用预训练的ResNet-18网络作为模型主干,其参数量相对较小,训练时间较短^[24-25]。

2) 混淆矩阵性能实验

动作识别的目的在于尽可能准确地地区分各类动作,区分度越高,识别效果越佳。混淆矩阵以矩阵形式统计不同类别数据的实际类别与模型预测类别。其中,行表示样本的实际标签,列表示模型的预测类别。对角线元素代表各类手势动作的识别准确率,对角线元素数值越大,表明该手势的区分度越高,非对角线元素则表示识别错误的概率。图12给出了采用所提AM-WiGeR方法在自测数据集上进行实验所得的混淆矩阵。图12表明,AM-WiGeR对交叉手势的识别准确率最高,达到98%,对其余5种手势的识别准确率均在94%以上。

4 结束语

在手势识别领域,传统识别技术存在易受障碍物影响、易泄露隐私、部署困难等问题,因此,利用已广泛部署的Wi-Fi设备实现手势动作的非接触式感知尤为重要。针对当前基于Wi-Fi信号的手势识别研究中,表征能力强的信号特征难

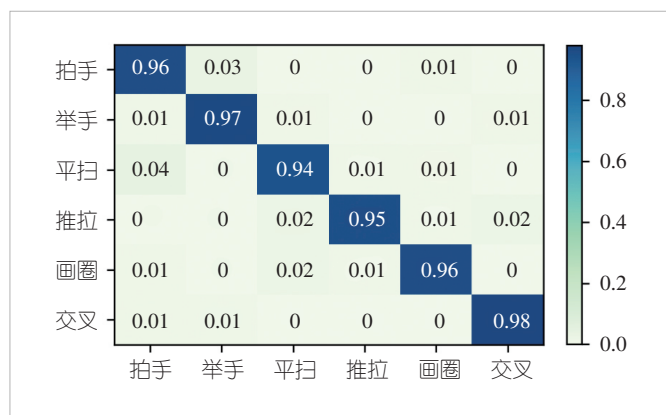


图12 6种手势识别结果的混淆矩阵

以提取、模型识别准确率提升困难等问题,本文提出了一种基于注意力机制增强残差网络模型的手势识别方法AM-WiGeR。该方法利用广泛部署的Wi-Fi设备采集CSI,从中提取DFS频谱图作为输入特征,并利用所提AM-RES模型进行分类识别。实验结果表明,相较于现有方法,所提AM-WiGeR方法有效提升了手势识别的准确率,对不同朝向手势识别的平均准确率达到96.22%。

参考文献

- [1] Zou H, Zhou Y X, Yang J F, et al. WiFi-enabled device-free gesture recognition for smart home automation [C]//Proceedings of IEEE 14th International Conference on Control and Automation (ICCA). IEEE, 2018: 476-481. DOI: 10.1109/ICCA.2018.8444331
- [2] 鲁勇, 吕绍和, 王晓东, 等. 基于WiFi信号的人体行为感知技术研究综述 [J]. 计算机学报, 2019, 42(2): 1-21. DOI: 10.11897/SP. J.1016.2019.00231
- [3] Khan R Z, Ibraheem N A. Hand gesture recognition: a literature review [J]. International journal of artificial intelligence & applications(IJAIA), 2012, 3(4): 161. DOI: 10.5121/ijaia.2012.3412
- [4] Venkatnarayan R H, Mahmood S, Shahzad M. WiFi based multi-user

- gesture recognition [J]. IEEE transactions on mobile computing, 2021, 20(3): 1242–1256. DOI: 10.1109/TMC.2019.2954891
- [5] 吴彤, 李业深, 黄镇煌, 等. 一种行人遮挡下的UWB非视距传播识别方法[J]. 物联网学报, 2023, 7(4): 63–71. DOI: 10.11959/j. issn. 2096–3750.2023.00348
- [6] Liu Z H, Li H, Gao B, et al. DQ-WiAPDoM: a DQN-based AP deployment optimization method for Wi-Fi FTM positioning [C]// Proceedings of 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). IEEE, 2023: 1–5. DOI: 10.1109/ICSTSN57873.2023.10151521
- [7] Wu T, Li H, Gao B, et al. NM-LUD: a NLOS mitigation method with low-cost UWB devices for indoor positioning [C]// Proceedings of 8th International Conference on Computer and Communication Systems (ICCCS). IEEE, 2023: 204–208. DOI: 10.1109/ICCCS57501.2023.10151336
- [8] Kexin Yao, Han Li, Ming Liu, et al. ARES-WiGR: An Attention-enhanced ResNet based Wi-Fi Gesture Recognition [C]// Proceedings of IEEE INFOCOM. IEEE, 2023
- [9] Gao Q H, Wang J, Ma X R, et al. CSI-based device-free wireless localization and activity recognition using radio image features [J]. IEEE transactions on vehicular technology, 2017, 66(11): 10346–10356. DOI: 10.1109/TVT.2017.2737553
- [10] Zhang J, Wu F X, Wei B, et al. Data augmentation and dense-LSTM for human activity recognition using WiFi signal [J]. IEEE Internet of Things journal, 2021, 8(6): 4628–4641. DOI: 10.1109/JIOT.2020.3026732
- [11] 周祥, 常俊, 王颖颖, 等. 基于信道状态信息的无接触人体行为识别方法[J]. 现代电子技术, 2021, 44(17): 80–85. DOI: 10.16652/j. issn. 1004–373x.2021.17.015
- [12] Ma Y S, Zhou G, Wang S Q. WiFi sensing with channel state information: a survey [J]. ACM computing surveys, 2020, 52(3): 1–36. DOI: 10.1145/3310194
- [13] Zhang J, Wei B, Wu F X, et al. Gate-ID: WiFi-based human identification irrespective of walking directions in smart home [J]. IEEE Internet of Things journal, 2021, 8(9): 7610–7624. DOI: 10.1109/JIOT.2020.3040782
- [14] Yang Z, Zhou Z M, Liu Y H. From RSSI to CSI: Indoor localization via channel response [J]. ACM computing surveys, 2013, 46(2): 1–32. DOI: 10.1145/2543581.2543592
- [15] Farid Z, Nordin R, Ismail M. Recent advances in wireless indoor localization techniques and system [J]. Journal of computer networks and communications, 2013, 2013: 185138. DOI: 10.1155/2013/185138
- [16] Wang F X, Gong W, Liu J C. On spatial diversity in WiFi-based human activity recognition: a deep learning-based approach [J]. IEEE Internet of Things journal, 2019, 6(2): 2035–2047. DOI: 10.1109/JIOT.2018.2871445
- [17] Zhu A C, Tang Z H, Wang Z X, et al. Wi-ATCN: attentional temporal convolutional network for human action prediction using WiFi channel state information [J]. IEEE journal of selected topics in signal processing, 2022, 16(4): 804–816. DOI: 10.1109/JSTSP.2022.3163858
- [18] Zhang X, Tang C P, Yin K, et al. WiFi-based cross-domain gesture recognition via modified prototypical networks [J]. IEEE Internet of Things journal, 2022, 9(11): 8584–8596. DOI: 10.1109/JIOT.2021.3114309
- [19] Feng C H, Arshad S, Zhou S W, et al. Wi-multi: a three-phase system for multiple human activity recognition with commercial WiFi devices [J]. IEEE Internet of Things journal, 2019, 6(4): 7293–7304. DOI: 10.1109/JIOT.2019.2915989
- [20] Zhu D L, Pang N, Li G, et al. WiseFi: activity localization and recognition on commodity off-the-shelf WiFi devices [C]// Proceedings of IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2016: 562–569. DOI: 10.1109/HPCC-SmartCity-DSS.2016.0085
- [21] Nirmal I, Khamis A, Hassan M, et al. Deep learning for radio-based human sensing: recent advances and future directions [J]. IEEE communications surveys & tutorials, 2021, 23(2): 995–1019. DOI: 10.1109/COMST.2021.3058333
- [22] 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, 2022, 44(11): 8671–8688. DOI: 10.1109/TPAMI.2021.3105387
- [23] Meng W, Chen X C, Cui W, et al. WiHGR: a robust WiFi-based human gesture recognition system via sparse recovery and modified attention-based BGRU [J]. IEEE Internet of Things journal, 2022, 9(12): 10272–10282. DOI: 10.1109/JIOT.2021.3122435
- [24] Pu Q F, Gupta S, Gollakota S, et al. Whole-home gesture recognition using wireless signals [C]// Proceedings of the 19th Annual International Conference on Mobile Computing & Networking – MobiCom '13. ACM, 2013: 27. DOI: 10.1145/2500423.2500436
- [25] Qian K, Wu C S, Yang Z, et al. WIdar: decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi [C]// Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2017: 1–10. DOI: 10.1145/3084041.3084067

作者简介



李业深, 北京交通大学计算机科学与技术学院在读博士研究生; 主要研究方向为人工智能、无线感知、夹持天线等; 已发表论文9篇。



姚可欣, 北京交通大学计算机科学与技术学院在读硕士研究生; 主要研究方向为人工智能、无线感知等; 已发表论文2篇。



熊轲, 北京交通大学计算机科学与技术学院教授、博士生导师, 副院长; 主要研究方向为人工智能+5G/6G网络、无线大数据分析处理、AI赋能的移动网络优化设计、绿色智慧物联网、网络大数据分析、云计算边缘计算、无线感知等; 已发表论文200余篇。