

WiGNet: 一种适用于无线感知场景的手势识别模型

马凯凯, 段鹏松, 孔金生

(郑州大学网络空间安全学院, 450000, 郑州)

摘要: 针对现有 Wi-Fi 感知技术的手势识别研究中存在因感知特征辨识度较低及难以提取而导致识别准确率不高的问题, 在对感知数据进行多维重构并施以二维离散小波变换的基础上, 提出了一种基于卷积神经网络的高精度手势识别模型 WiGNet。首先, 在对信道状态信息影响因子分析的基础上, 提取振幅数据作为手势识别的基础数据; 其次, 将感知数据重构为多维矩阵形式, 以充分提取手势动作的时空特征; 然后, 对重构数据在时间和空间维度上进行二维离散小波变换, 以实现感知数据的降噪和平滑; 最后, 使用神经架构搜索技术优化网络深度以及卷积核个数的合理化配置, 并提出较好适配重构数据的手势识别模型 WiGNet。实验结果表明: 对感知数据进行二维离散小波变换后, 特征辨识度和模型整体运算速度均有显著提升; WiGNet 在自建数据集上和公共数据集上的平均识别准确率分别达到 98.1% 和 96.0%, 均优于同类模型。此外, WiGNet 在实现高精度手势识别的同时, 还能保持较快运算速度, 并兼具一定的鲁棒性。

关键词: 信道状态信息; 手势识别; 二维离散小波变换; 卷积神经网络

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202305019 **文章编号:** 0253-987X(2023)05-0194-10

WiGNet: A Gesture Recognition Model for the Wireless Sensing Scenario

MA Kaikai, DUAN Pengsong, KONG Jinsheng

(School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450000, China)

Abstract: Existing research suggests low accuracy of gesture recognition based on the WiFi sensing technology because of difficulties in extracting hard-to-recognize perceptual features. To solve this problem, this paper, through multidimensional reconstruction of perceptual data and two-dimensional discrete wavelet transform, proposes a high-precision gesture recognition model WiGNet based on the convolutional neural network. Firstly, based on the analysis of influence factors of channel state information, the amplitude data is extracted as the base data for gesture recognition; secondly, the perception data is reconstructed into a multidimensional matrix to fully extract the spatio-temporal characteristics of gestures; next, the two-dimensional discrete wavelet transform is applied to the reconstructed data in the temporal and spatial dimensions to reduce noise and smooth the perceptual data; finally, a neural architecture search technique is used to optimise the network depth and rationalise the number of convolutional kernels. Thus WiGNet, a gesture recognition model that can be better adapted to reconstructed data, is proposed. The ex-

收稿日期: 2022-07-13。 作者简介: 马凯凯(1999—), 男, 硕士生; 段鹏松(通信作者), 男, 博士, 讲师。 基金项目: 国家自然科学基金资助项目(61972092); 郑州市协同创新重大专项资助项目(20XTZX06013); 河南省高等学校重点科研计划资助项目(21A520043)。

网络出版时间: 2022-12-14

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1069.T.20221213.1647.002.html>

perimental results show that the model significantly improves the feature identification accuracy and has a higher overall operation speed after the perception data is processed by two-dimensional discrete wavelet transform. WiGNet achieves an average recognition accuracy of 98.1% and 96.0% based on the self-built dataset and the public dataset respectively, outperforming similar models. In addition, WiGNet can not only achieve high-precision gesture recognition, but also maintain a fast computing speed and have a certain degree of robustness.

Keywords: channel state information; gesture recognition; two dimensional discrete wavelet transform; convolutional neural network

随着物联网及人工智能技术的快速发展,人机交互技术(HCI)的使用场景越来越多^[1],手势识别技术也成为学术界和工业界的研究热点之一^[2-3]。利用手势识别技术,人与设备的交互更为便捷,其相关技术及应用在万物互联背景下潜力巨大。例如,在智能家居中,可以使用手势控制电视、空调、冰箱等家用设备,而无需额外的控制器,能极大提升用户的使用体验和操作效率。

目前,实现手势识别的主流技术可以分为基于计算机视觉^[4-5]、专用传感器^[6]和射频信号^[7]三类。其中,基于计算机视觉的手势识别技术通过收集大量的手势照片,利用机器学习等技术学习和识别具体手势信息。该技术研究成果较多,识别精度也较高,但存在弱光照条件受限及低隐私保护等缺点。基于专用传感器的手势识别技术通过专用传感器(如可穿戴设备)采集手势动作,再通过内部自带的识别系统感知手势类别^[6]。该技术在手势识别粒度及精度上有较大优势,但安装和维护成本较高,难以实现大规模应用^[8]。基于射频信号的手势识别技术通过分析信号的变化情况实现手势的分类,其中基于 Wi-Fi 信号的手势识别技术因其易部署、成本低、非接触式、隐私保护性好等优势,逐渐受到研究者的青睐。

在早期的 Wi-Fi 手势识别研究中,研究者通常使用接收信号强度指示(RSSI)作为感知特征载体。Abdelnasser 等提出了一种通过分析 RSSI 变化来感知用户手势系统的 WiGest^[9],在单个无线设备接入点实验条件下实现了 87.5% 的识别准确度。由于多径效应,RSSI 测量的是多路径信号的叠加结果,无线感知的识别易受室内布局影响,因此 RSSI 仅能感知粗粒度的目标行为(如走动、跳跃等)。对于细粒度的目标行为(如手势动作等),则需要通过信道状态信息(CSI)来感知。CSI 可以从一个数据包中同时测量多个子载波的频率响应,而非全部子载波叠加的总体幅度响应,从而更加精细地刻画频率选择性信道。随着 CSI Tool^[10]工具的开源,使得从商用 Wi-Fi 设备中

提取 CSI 成为可能,基于 CSI 的手势识别研究也逐渐占据主导。2016 年,Tan 等提出了利用 CSI 进行细粒度手指动作的系统 WiFinger^[11],实现了 93% 以上的动作识别准确率。但是,由于 WiFinger 只选择了关键子载波的感知数据,造成了大量的数据浪费和信息丢失。Han 等提出的 DANGR^[12]使用生成对抗网络对感知数据进行增强,降低数据集不完整引起的过拟合风险,最终达到了 94.5% 的手势识别平均准确率。Meng 等提出的 FaSee^[13]通过使用动态时间规整算法,使手势识别的平均准确率达到 94.75%。但是,该算法需要和 k -最邻近分类算法相结合,时间复杂度较高。

现有基于 Wi-Fi 感知技术的手势识别研究中,大多数研究者从 CSI 中选择特定子载波,并基于一维时序数据流进行特征识别^[11,14],未能很好地利用不同频率子载波之间的空间特征差异,使得感知特征辨识度较低,且噪声信息不能有效滤除,导致难以实现高准确率的手势识别。为解决此问题,本文在以下方面做出改进。

(1)提出了一种具有较高辨识度的感知特征表征机制。首先,将采集到的一维时序 CSI 数据重构为多维矩阵形式,以提取载波内和载波间所蕴含的手势特征;其次,对重构后的多维感知数据进行二维离散小波变换(2D-DWT),以进行时间和空间层面上的数据降噪和平滑,得到具有较高辨识度的感知特征表征机制。

(2)本文提出了一种基于深度学习的高精度手势识别模型 WiGNet。WiGNet 以 2D-DWT 处理后的多维矩阵为输入,由多个 NGConv 模块构成的卷积神经网络组成。自建数据集和公共数据集上的实验结果均表明,WiGNet 相比同类模型有明显优势。

1 问题定义

1.1 信道状态信息

信道状态信息是在正交频分复用(OFDM)技术中

估计信道状态的信息^[15]。通常来说,每一条无线链路有多个子载波,每一个子载波都会产生对应的 CSI 值。

当信号在无线信道中传播时,会遇到路径中物体的反射、折射等扰动,这些扰动最终体现为接收端的 CSI 数据。假设接收端的和发射端的无线信号分别为 α 和 β , 则有

$$\alpha = H\beta + a \tag{1}$$

式中: a 为无线信号传播过程产生的噪声向量; H 表示所有子载波组成的信道矩阵,如果把噪声表示为 $a \sim cN(O, S)$ 形式的正态分布,那么就可以估计出信道矩阵 H

$$\hat{H} = \alpha/\beta \tag{2}$$

对于 H 的估计就是 CSI 的值,对于单个子载波来说,可以用以下形式表示 CSI

$$h = |h|e^{j\sin\theta} \tag{3}$$

式中 $|h|$ 表示 CSI 的振幅。当人在无线信道中做手动动作时, $|h|$ 就会产生相应的变化,通过分析这些变化就能实现对不同手势的检测。

1.2 问题定义

每一个 OFDM 信道中每一条子载波可以得到一个 CSI 序列。假设有一个具有 N_s 个发射器天线, N_r 个接收器天线的系统,则每个接收到的 CSI 数据包都可以解析出 $N_s \times N_r \times n$ 个 CSI 序列,其中 n 表示每个无线信道中子载波的数量。为了获得 CSI 数据流,需要在接收端持续采集时间间隔为 T 的 CSI 数据,每个子载波可以得到一个长度为 T 的 CSI 序列。目前大部分基于 CSI 的手势识别研究大都仅使用单个子载波的 CSI 数据进行分析实验,其形式如式(4)所示

$$\mathbf{X} = [h_{1,k}, h_{2,k}, \dots, h_{t,k}, \dots, h_{T,k}]^T \tag{4}$$

式中 $h_{t,k}$ 表示 t 时刻第 K 个子载波的数值。每个子载波间手势特征变化不总是相同的,为了更好地利用不同频率子载波之间的空间特征差异,本文将每条无线信道间所有子载波的 CSI 序列重构为二维矩阵形式,如式(5)所示

$$\mathbf{X} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,k} & \cdots & h_{1,K} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,k} & \cdots & h_{2,K} \\ \vdots & \vdots & & \vdots & & \vdots \\ h_{t,1} & h_{t,2} & \cdots & h_{t,k} & \cdots & h_{t,K} \\ \vdots & \vdots & & \vdots & & \vdots \\ h_{T,1} & h_{T,2} & \cdots & h_{T,k} & \cdots & h_{T,K} \end{bmatrix} \tag{5}$$

根据 Wi-Fi 感知中 OFDM 和 MIMO(multiple-input multiple-out-put)的技术特性,以及实际采集多发射器、多接收器的场景,对于每个采集时间间隔 T ,可以得到 $N_s \times N_r$ 个如式(5)所示的二维矩阵。由于不同收发天线获取的信号具有一定的差异,为了保留这些信号差异,本文将 $N_s \times N_r$ 个矩阵在深度方向堆叠,增强手势动作的特征,消除不同无线信道间手势特征的差异,提高鲁棒性^[16]。

至此,将带标签的数据集表示为 $Z = \{(\mathbf{X}_j, Y_j)\}_{j=1}^J$,其中数据集 Z 包含 J 个重构数据,每个重构数据 $\mathbf{X}_j$ 有对应的标签 Y_j ,则手势识别问题可以归纳为:构建一个映射模型,根据输入的重构数据 $\mathbf{X}$ 预测对应的手势标签 Y 。

2 模型设计

本文设计的手势识别系统方案总体流程图如图 1 所示。

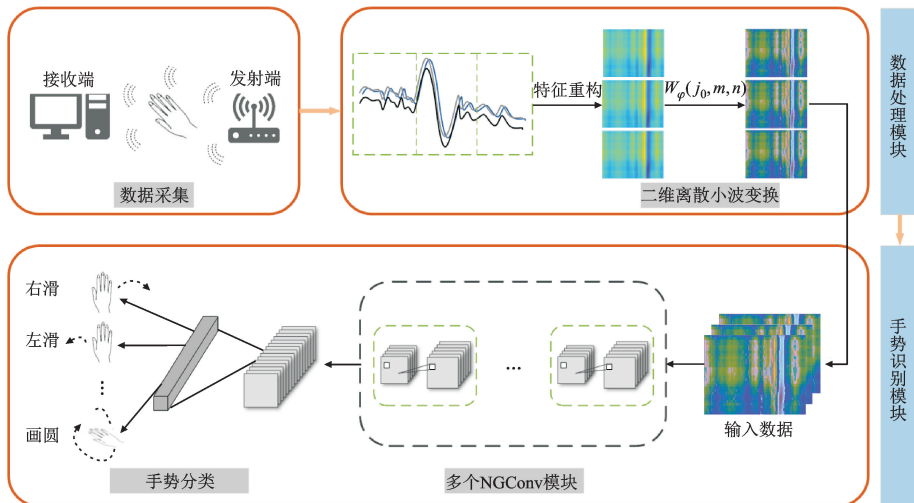


图 1 手势识别总体流程图

Fig. 1 Overall flow chart of gesture recognition

首先,使用收发设备获取包含不同手势特征的 CSI 数据;其次,将采集得到的 CSI 数据重构为 tensor 格式的二维矩阵形式,并对重构数据进行 2D-DWT 变换等达到降噪和平滑的目的;最后,将处理后的数据放入设计好的深度神经网络 WiGNet 中,提取手势特征信息,并对不同手势特征进行分类,实现手势识别功能。

2.1 二维离散小波变换

在采集 CSI 手势数据的过程中,Wi-Fi 无线信号会受到墙壁、桌椅以及身体其他部位微小动作的影响,这些多径效应会让 CSI 信号产生细微波动。时间精度和频率精度是小波变换在高频和低频时具有的特性,这使其可以聚焦到无线信号变化的所有细节,对时频信号分析有很高的契合度。基于这些特点,本文对重构的 CSI 数据采用 2D-DWT 方法进行降噪和降维处理。

在对无线信号的处理中,需要将连续的小波及其小波变换离散化。假设 $f(x)$ 表示无线信号,则其可以在子空间中用小波函数 $\psi(x)$ 和尺度函数 $\varphi(x)$ 的展开形式来表示,如式(6)所示

$$f(x) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_j(k) \psi_{j,k}(x) \quad (6)$$

式中: x 表示单个子载波的值; j_0 是任意的开始尺度; $c_{j_0,k}$ 为近似系数; $d_j(k)$ 为小波系数。将式(6)展开即可得到正向的离散小波变换(DWT),如式(7)和式(8)所示

$$c_{j_0}(k) = (f(x), \varphi_{j_0,k}(x)) = \int f(x) \varphi_{j_0,k}(x) dx \quad (7)$$

$$d_j(k) = (f(x), \psi_{j,k}(x)) = \int f(x) \psi_{j,k}(x) dx \quad (8)$$

有些背景噪声在一维信号中体现的是随机性,但可能在二维信号中就显示出很强的区域性或较明显的特征性。由于 DWT 只能在无线信号的时间维度上运作,为更好地利用不同频率子载波之间的空间特征差异,本文引入了 2D-DWT。对于重构数据,2D-DWT 不仅能在无线信号的时间维度上运作,还能对不同子载波间的 CSI 信号进行平滑降噪处理。

在二维的情况下,给定尺度函数 $\varphi(x)$ 和小波函数 $\psi(x)$,可以组合出 1 个二维尺度函数和 3 个二维小波函数,分别如式(9)~(12)所示

$$\varphi(x, y) = \varphi(x) \varphi(y) \quad (9)$$

$$\psi^T(x, y) = \psi(x) \varphi(y) \quad (10)$$

$$\psi^V(x, y) = \varphi(x) \psi(y) \quad (11)$$

$$\psi^D(x, y) = \psi(x) \psi(y) \quad (12)$$

式中 T、V、D 分别表示列方向、行方向、对角线方向。

于是,对于大小为 $M \times N$ 的二维矩阵 $f(x, y)$,其离散小波变换如式(13)和式(14)所示

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (13)$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x, y), \quad i = \{T, V, D\} \quad (14)$$

如同一维的情况, j_0 是一个任意开始的尺度,尺度 j_0 处的近似由式(13)定义。图 2 直观地表示了 2D-DWT 如何对矩阵从水平和竖直两个方向进行处理。其中 a 是原始矩阵, b 是经过一阶 2D-DWT 处理后的矩阵, c 是经过 2 阶 2D-DWT 处理后的矩阵,以此类推。

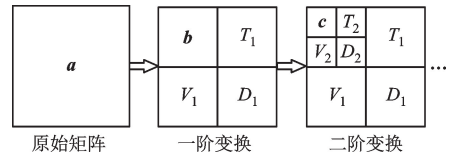


图 2 矩阵的 2D-DWT 处理过程

Fig. 2 2D-DWT process of matrix

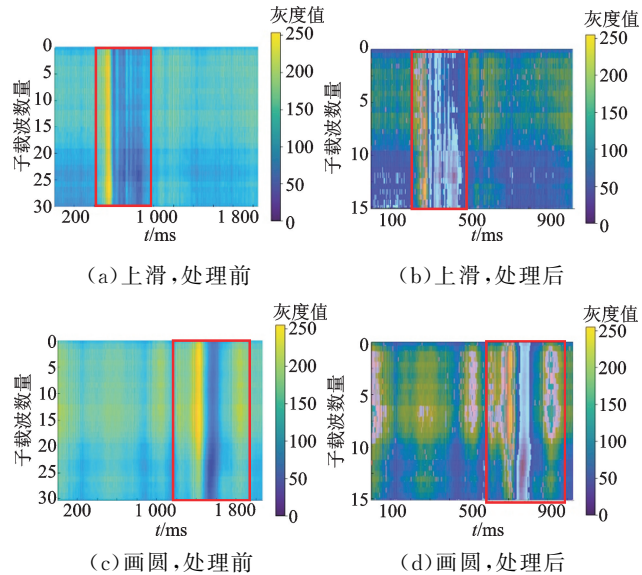


图 3 两种手势动作经 2D-DWT 处理前后对比

Fig. 3 Comparison before and after 2D-DWT treatment

本文所使用的小波函数为多贝西(Daubechies, db)小波函数。在使用 2D-DWT 对数据处理时使用

tensor 数据格式,此格式已被证明能加快数据处理速度^[17]。频率能量图^[16]是 CSI 数据的可视化表示,图 3 所示分别为经过 2D-DWT 处理前和处理后的 2 种手势动作(上滑、画圆)的频率能量图,其中横轴表示采样时间,竖轴表示子载波数量,颜色表示 CSI 数据的大小。可以直观地看出,经过 2D-DWT 处理后,数据的尺寸变小,频率能量图信号的变化情况更加突出。3.2 节的实验也证明了经过 2D-DWT 处理后的数据提高了手势识别的准确率,降低了系统整体运算时间。

2.2 WiGNet 模型

对原始 CSI 数据重构后,考虑到卷积神经网络在多维矩阵特征提取方面的巨大优势,本文使用神经架构搜索技术(NAS)^[18-19]搜索网络深度以及卷积核个数的合理化配置,使用搜索得到的网络提取数据特征,进行手势分类。

2.2.1 神经架构搜索技术

NAS 指一系列搜索神经网络架构的算法,其思想为:首先定义一组可能用于网络的“构建块”(如卷积核、池化层等)。在 NAS 算法中,控制器递归神经网络对这些构建块进行采样,将它们放在一起创建某种端到端架构。这种架构的风格通常与经典的网络设计相似,如 ResNet^[20]或 DenseNet^[21],块之间的组合和配置是它们之间的主要差别。然后,小规模训练这种新的网络体系结构,在获得一定的精度后通过生成更好地体系结构更新控制器。

在搜索过程中,增加网络的深度能够得到更加丰富、复杂的特征,但网络深度过深往往会产生过拟合、梯度消失等问题。增加网络的宽度(卷积核的个数)能够获得更高细粒度的特征并且也更容易训练,但对于宽度很大而深度较浅的网络往往很难学习到更深层次的特征。结合重构后 CSI 数据的特点,本文使用 NAS 技术得到了更加适合重构后 CSI 数据的网络结构。

2.2.2 注意力机制结构

本文加入了注意力机制结构以聚焦无线信号的关键信息。当手势做出动作时,无线信号会发生特定的波动,注意力机制的引入能更好地提取波动变化信息。

如图 4 所示,本文设计的注意力机制由一个全局平均池化层和两个卷积层组成,第一个卷积层使用 SiLU 激活函数结构,第二个卷积层使用 Sigmoid 激活函数,其输出和注意力机制结构的输入在深度上堆叠以保留聚焦得到的特征。

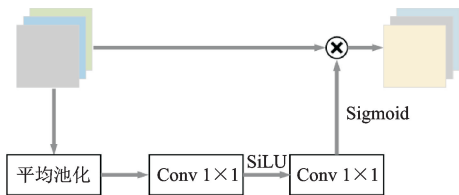


图 4 注意力机制结构

Fig. 4 Attention mechanism structure

2.2.3 NGConv 模块

本文使用的网络结构由 8 个 NGConv 模块组合而成。如图 5 所示,首先,数据经过一个卷积核大小为 1×1 的卷积层达到数据升维作用。之后,进行正则化和激活处理。正则化可以防止模型过拟合和加速模型收敛,激活函数选用 SiLU 激活函数,如式(15)所示, SiLU 激活函数具有平滑、非单调、无上界有下界的特性。

$$\text{SiLU}(x) = x * \text{Sigmoid}(x) \quad (15)$$

然后,数据进入一个卷积核大小为 3×3 的深度可分离卷积层中^[22],同样经过正则化和 SiLU 激活函数处理后输入注意力结构块中。

最后,数据经过注意力机制块处理后经过一个卷积核大小为 1×1 的卷积层以降低矩阵维度,经过正则化层之后使用 Dropout 操作防止模型过拟合,其中 Dropout 率设置为 0.2。

NGConv 模块还使用了残差连接操作^[20],如式(16)所示

$$x_{l+1} = h(x_l) + F(x_l, W_l) \quad (16)$$

在输入下一个模块之前,每次经过 NGConv 模块处理的数据都会和输入数据相加,残差连接能很好地解决梯度消失和网络退化的问题^[20]。

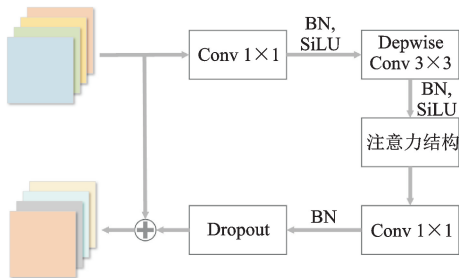


图 5 NGConv 模块示意图

Fig. 5 NGConv module schematic diagram

数据在输入 WiGNet 模型后,先进入一个卷积核大小为 3×3 的卷积层,之后由 8 个以线性结构连接的 NGConv 模块进行特征提取,最后经由卷积核

大小为 1×1 的卷积层展开为各个手势动作的预测值。

3 实验与分析

3.1 实验设置

3.1.1 数据集描述

本文分别使用自建采集的数据集和公共数据集评估模型的性能。

由于无线信号的传播会受到多径效应的显著影响,因此不同多径效应下无线信号的效果有一定区别。考虑模型在不同多径效应环境下的运行效果,弱多径效应环境遮挡物少,环境开阔,信号传播路径相对简单;强多径效应环境物品较多,信号传播路径较为复杂。为保证实验数据的多样性和客观性,在采集数据时选择了弱多径效应的数据采集室和强多径效应的学生实验室两种环境,如图 6 所示。

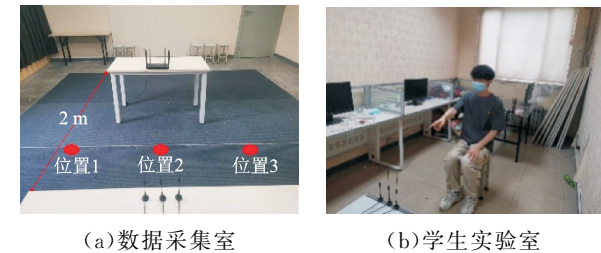


图 6 采集环境
Fig. 6 Collection environment

发射器选择具有一个发射天线的 TP_LINK AC1750 无线路由器,接收器为一台具有 3 个接收天线安装有开源 CSI Tool 工具的计算机。采样频率为 1 kHz,Wi-Fi 工作频率为 5 GHz。如图 6 所示,发射器和接收器相距 2 m,考虑到视距路径和非视距路径对无线信号传播的影响,消除位置误差并结合现实情况,在发射器和接收器中间选择 3 个不同位置进行 CSI 数据的采集,采集过程中受测者始终面向接收器。每次接收含 $90(1 \times 3 \times 30)$ 个子载波的 CSI 数据。本文在每种环境下分别安排 2 个受测者,两种环境下每个受测者分别采集 9 种手势动作,分别为左滑、右滑、上滑、下滑、画 O、画十、画 Z、画 W、画 S,每次采集时间间隔为 2 s,每种动作在每个采集位置处采集 20 条数据,处理后得到 2 160 条数据。

公共数据集^[23]在教室、大厅、办公室 3 个实验地点采集。使用的接收设备是配有 Intel 5300 无线网卡的笔记本计算机。每个采集环境的设备和实验布置相同,有一个发射器(每个发射器 1 条发射天线)、6 个接收器(每个接收器 3 条接收天线),每个

接收器每次接收含 $90(1 \times 3 \times 30)$ 个子载波的 CSI 数据。本文选取数据集集中的 6 种手势进行实验,分别为拍手、横扫、画 O、推拉、摆动、画 Z。每个实验地点分别选取 4 个受测者的手势数据。对于每个实验地点,每种手势动作各整理了 1 500 个数据,每个地点分别整理 9 000 个数据进行实验。

自建数据集和公共数据集主要参数设置如表 1 所示。

表 1 数据集主要参数设置
Table 1 Dataset main parameter settings

参数	自建数据集取值	公共数据集取值
发射天线数	1	1
接收天线数	3	3
接收器数量	1	6
采集位置数	3	5
单天线载波数	30	30
工作频率	5 GHz	—
接收端采样率	1 kHz	—

3.1.2 模型超参数

本文实验基于深度学习框架 PyTorch 实现。优化器使用随机梯度下降算法(SGD),其动量因子设置为 0.9, L_2 正则化参数设置为 0.000 5 以防止过拟合。Batch_size 设置为 16。初始学习率设置为 0.01,并使用余弦学习率衰减策略优化学习速度^[24],最终学习率下降为 0.000 1。将每个环境的数据划分为 80%训练集和 20%测试集,使用测试集评估识别效果。

3.1.3 评估指标

本文采用准确率(A)、精确率(P)、召回率(R)、特异度(S_p)、方差评估模型效果。模型最终需要判断样本的结果是正确(positive)还是错误(negative)。以 F_N 表示真实值是 positive,预测值是 negative 的数量; T_P 表示真实值是 positive,预测值是 positive 的数量; T_N 表示真实值是 negative,预测值是 negative 的数量; F_N 表示真实值是 negative,预测值是 positive 的数量; S^2 为总体方差, Y 表示预测值, N 为数据总数。准确率指正确分类样本数占总体样本数比例;精确率指预测的所有 positive 中,预测正确的比例;召回率指预测正确的 positive 占所有真实 positive 的比例;特异度指预测正确的 negative 占所有真实 negative 的比例。方差描述的是训练模型中每个预测值与预测均值差的平方和的平均数。评估指标的公式如下

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{17}$$

$$P = \frac{T_P}{T_P + F_P} \tag{18}$$

$$R = \frac{T_P}{T_P + F_P} \tag{19}$$

$$S_P = \frac{T_N}{T_N + F_P} \tag{20}$$

$$S^2 = \frac{\sum (Y - \bar{Y})^2}{N} \tag{21}$$

3.2 实验结果

为了评估所提出方法的性能:首先,本文验证了所提出的 2D-DWT 方法的有效性;然后,给出了模型的最终分类结果并和现有模型的性能进行对比;最后,给出所提方法在公共数据集上的表现。

3.2.1 2D-DWT 的有效性实验

图 7 为使用 2D-DWT 和不使用 2D-DWT 的准确率,分别在自建数据集中数据采集室和学生实验室进行实验。结果显示:在数据采集室环境下,不使用 2D-DWT 的分类准确率为 97.3%,使用 2D-DWT 的分类准确率为 98.2%;在学生实验室环境下,不使用 2D-DWT 的分类准确率为 97.2%,使用 2D-DWT 的分类准确率为 98.1%。

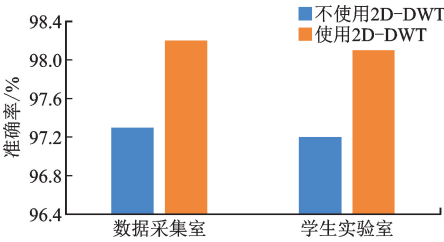


图 7 是否使用 2D-DWT 的准确率对比
Fig. 7 Accuracy comparison of whether to use 2D-DWT or not

图 8 为不使用和使用 2D-DWT 情况下在数据采集室下训练集损失随时间变化图。尽管对数据进

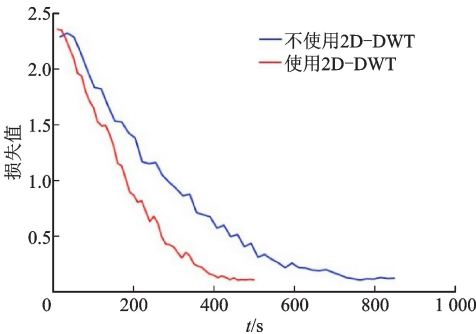


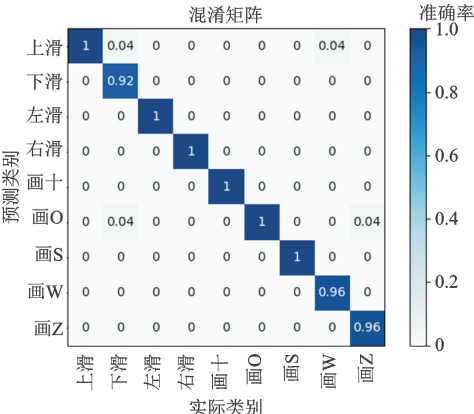
图 8 是否使用 2D-DWT 的 loss 变化曲线

Fig. 8 Loss curve of whether to use 2D-DWT or not

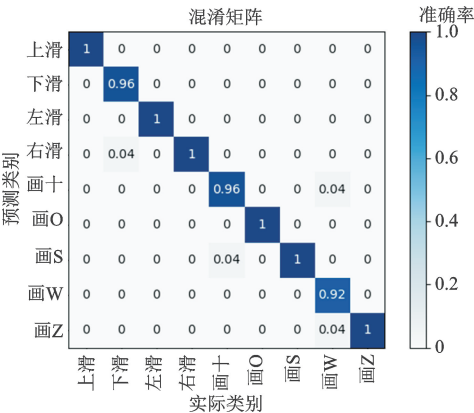
行 2D-DWT 时会消耗一定的运行时间,但先进的处理方式^[17]与处理后更小的数据尺寸使整体的运行时间更短。从图中可以看出,使用 2D-DWT 训练时损失值能够更快地收敛。实验结果充分说明,使用 2D-DWT 对数据进行降噪后,不仅提高了分类结果的准确率,还降低了训练所需时间,证明了 2D-DWT 的有效性。

3.2.2 系统整体性能

混淆矩阵可以观察模型在各个类别上的表现。图 9 为使用本文提出的基于 WiGNet 的方法对 9 种手势动作的最终分类结果的混淆矩阵,其中横轴表示实际的手势动作,竖轴表示预测的手势动作,对角矩阵表示每一种手势动作的识别准确率。可以看出对于所采集的 9 种手势动作,WiGNet 对于每一种动作都有很好的分类效果,对于左滑、画圆、画 S 等区分度高的动作的分类最准确。两种环境下识别准确率分别达到了 98.2%和 98.1%,表明 WiGNet 能实现高精度手势识别,且具有一定的鲁棒性。



(a)标准数据采集实验室



(b)学生实验室

图 9 两种环境下的混淆矩阵

Fig. 9 Confusion matrix in two environments

为了评估 WiGNet 的性能,将 WiGNet 和当前

流行的手势识别方法对比。本文使用在标准数据采集实验室下的数据分别与经典分类模型 ResNet50^[20]、经过神经架构搜索技术改良的适用于本文数据集的最佳深度 ResNet 网络 ResNet36、手势识别最新模型 FaSee^[13]进行对比,结果如表 2 所示。

表 2 不同手势识别方法的结果对比

Table 2 Comparison of different gesture recognition methods					
模型	准确率 /%	精确率 /%	召回率 /%	特异度 /%	方差 /%
ResNet50	93.6	93.8	93.4	94.8	1.092
ResNet36	93.9	94.1	93.4	95.1	1.076
FaSee	95.2	95.3	95.2	96.7	0.822
WiGNet	98.2	98.3	98.1	99.8	0.683

在训练过程中发现,当 ResNet50 达到了一定的效果后,随着轮次的不断增加,训练集和测试集的准确率相差变大,出现过拟合,具体表现为结果的方差较大;尽管 ResNet36 使用神经架构搜索技术得到了适合于本文数据集的网络深度,但其表现和 ResNet50 相差不大,仍出现过拟合问题;本文提出的 WiGNet 采用更加轻量级的网络,网络深度更浅,并且在每个 NGConv 模块中加入了 dropout 层用以防止过拟合,结果显示 WiGNet 训练结果的方差最小,模型稳定性更高,训练效果更好;FaSee 方法的效果虽然好于 ResNet,但动态时间规整算法和 k -最邻近分类算法的时间复杂度高,开销大,且其并没有考虑各个子载波之间的空间特征,最终效果差于本文提出的 WiGNet。从表 2 可以看出,不管是在识别准确率还是其他评估指标下,WiGNet 相较于其他方法的效果都更好。

3.2.3 公共数据集结果

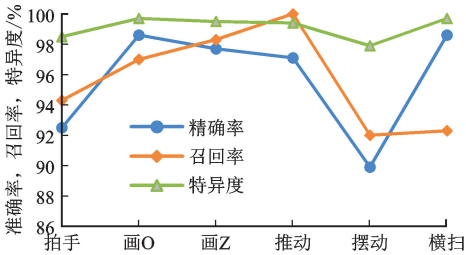
为了突出本文方法的高性能,本节将和使用相同公共数据集的方法对比,分别是 Widar3.0^[23]、CGRU-ELM^[25],对比结果如表 3 所示,可以看出,WiGNet 相较于其他方法具有更好的分类性能。

表 3 公共数据集准确率对比

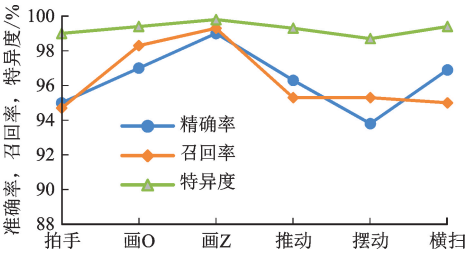
Table 3 Comparison of accuracy in public data sets			
地点	准确率/%		
	WiGNet	Widar3.0	CGRU-ELM
教室	95.7	92.7	93.9
办公室	96.3	—	92.9

图 10 展示了使用 WiGNet 在公共数据集上的

评估结果。通过对两个环境下的每个动作进行评估可以看出,对于画 Z、画 O 等区分度高的动作的分类最准确,对于摆动、横扫等动作的识别准确率相对较低。图 11 所示为经过 2D-DWT 处理过的摆动和横扫动作的频率能量图,可以看出两种动作的特征重合度较高,模型对于不同特征的提取相对困难。总体来说,对于两个环境下的 6 种手势动作,WiGNet 对于每一种手势动作都实现了较好的分类效果,两种环境下的高准确率分类效果表明 WiGNet 具有较好的鲁棒性。



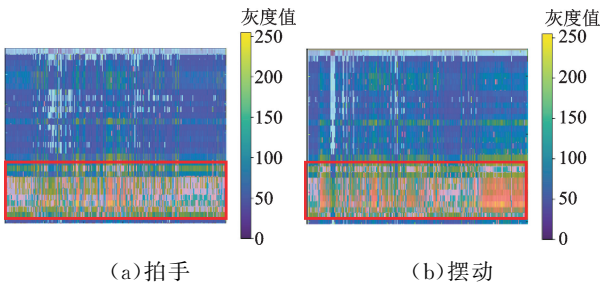
(a)教室环境下评估结果



(b)办公室环境下评估结果

图 10 公共数据集评估结果

Fig. 10 Evaluation results on public datasets



(a)拍手

(b)摆动

图 11 经 2D-DWT 处理后的两种手势动作

Fig. 11 Two gestures processed by 2D-DWT

4 结 论

传统基于 Wi-Fi 的手势动作识别模型,大多数研究者采用从 CSI 信息中选择特定子载波,并基于一维时序数据流进行特征识别,这造成大量的数据浪费,未能很好利用不同频率子载波之间的空间特征差异,存在因感知特征辨识度较低及难以提取而

导致识别准确率不高的问题。本文首次将二维离散小波变换引入感知特征的重构流程中,实现了在时间和空间层面上的数据降噪和平滑。在此基础上,本文还提出了基于深度学习的高精度手势识别模型 WiGNet。WiGNet 以 2D-DWT 处理后的多维矩阵为输入,能准确地提取手势动作的特征。实验结果表明,使用二维离散小波变换对数据处理能提高识别准确率和运算速度,提出的 WiGNet 在自建数据集和公共数据集上均有良好的识别分类性能,并且相比同类模型表现更好。

CSI 含有丰富的无线信号信息,如振幅、相位、多普勒频移、功率分布信息以及人体坐标系下的速度谱等,如何将这些信息结合起来,提高整体识别性能与适应能力可以作为未来的研究方向。

参考文献:

- [1] ZHOU Yimin, JIANG Guolai, LIN Yaorong. A novel finger and hand pose estimation technique for real-time hand gesture recognition [J]. Pattern Recognition, 2016, 49: 102-114.
- [2] VENKATNARAYAN R H, PAGE G, SHAHZAD M. Multi-user gesture recognition using WiFi [C]//Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services. New York, NY, USA: Association for Computing Machinery, 2018: 401-413.
- [3] HE Wenfeng, WU Kaishun, ZOU Yongpan, et al. WiG: WiFi-based gesture recognition system [C]//2015 24th International Conference on Computer Communication and Networks (ICCCN). Piscataway, NJ, USA: IEEE, 2015: 1-7.
- [4] REHG J M, KANADE T. Visual tracking of high DOF articulated structures: An application to human hand tracking [C]//Computer Vision-ECCV'94. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994: 35-46.
- [5] STARNER T, PENTLAND A. Real-time American sign language recognition from video using hidden Markov models [M]//SHAH M, JAIN R. Motion-Based Recognition. Dordrecht, Netherlands: Springer Netherlands, 1997: 227-243.
- [6] NANDAKUMAR R, TAKAKUWA A, KOHNO T, et al. CovertBand: activity information leakage using music [C]//Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. New York, NY, USA: Association for Computing Machinery, 2017: 87.
- [7] KELLOGG B, TALLA V, GOLLAKOTA S. Bringing gesture recognition to all devices [C]//Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation. USA: USENIX Association, 2014: 303-316.
- [8] ZHAO Mingmin, LI Tianhong, ALSHEIKH M A, et al. Through-wall human pose estimation using radio signals [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7356-7365.
- [9] ABDELNASSER H, YOUSSEF M, HARRAS K A. WiGest: a ubiquitous WiFi-based gesture recognition system [C]//2015 IEEE Conference on Computer Communications (INFOCOM). Piscataway, NJ, USA: IEEE, 2015: 1472-1480.
- [10] HALPERIN D, HU Wenjun, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information [J]. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 53.
- [11] TAN Sheng, YANG Jie. WiFinger: leveraging commodity WiFi for fine-grained finger gesture recognition [C]//Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing. New York, NY, USA: Association for Computing Machinery, 2016: 201-210.
- [12] HAN Zijun, GUO Lingchao, LU Zhaoming, et al. Deep adaptation networks based gesture recognition using commodity WiFi [C]//2020 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway, NJ, USA: IEEE, 2020: 1-7.
- [13] MENG Xianjia, FENG Lin, CHHEN Hao, et al. Just-in-Time human gesture recognition using WiFi signals [J]. Chinese Journal of Electronics, 2021, 30(6): 1111-1119.
- [14] VIRMANI A, SHAHZAD M. Position and orientation agnostic gesture recognition using WiFi [C]//Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. New York, NY, USA: Association for Computing Machinery, 2017: 252-264.
- [15] WU Kaishun, XIAO Jiang, YI Youwen, et al. FILA: fine-grained indoor localization [C]//2012 Proceedings IEEE INFOCOM. Piscataway, NJ, USA: IEEE, 2012: 2210-2218.
- [16] 段鹏松, 周志一, 王超, 等. WiNet: 一种适用于无线感知场景的步态识别模型 [J]. 西安交通大学学报, 2020, 54(7): 187-195.

DUAN Pengsong, ZHOU Zhiyi, WANG Chao, et al. WiNet: a gait recognition model suitable for wireless

- sensing scene [J]. Journal of Xi'an Jiaotong University, 2020, 54(7): 187-195.
- [17] COTTER F. Uses of complex wavelets in deep convolutional neural networks [D]. Cambridge, UK: University of Cambridge, 2019.
- [18] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 8697-8710.
- [19] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [21] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2261-2269.
- [22] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2022-07-01]. <https://doi.org/10.48550/arxiv.1704.04861>.
- [23] ZHANG Yi, ZHENG Yue, QIAN Kun, 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.
- [24] HE Tong, ZHANG Zhi, ZHANG Hang, et al. Bag of tricks for image classification with convolutional neural networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 558-567.
- [25] 张鑫, 冯秀芳. WiFi 环境下基于 CGRU-ELM 混合模型的手势识别 [J]. 计算机工程与科学, 2022, 44(2): 298-305.
- ZHANG Xin, FENG Xiufang. Gesture recognition based on CGRU-ELM hybrid model in WiFi environment [J]. Computer Engineering & Science, 2022, 44(2): 298-305.

(编辑 亢列梅)