

分类号: _____
密 级: _____

单位代码: 10300
学 号: 20201221028

南京信息工程大学

硕士 学位 论文



论文题目: 基于 WiFi 的人体动作识别研究

申请人姓名: 廖振龙

指导教师: 苏健

学科名称: 软件工程

研究方向: 人体行为感知

培养学院: 软件学院

提交时间: 2023 年 6 月 1 日

二〇二三年六月

目 录

| | |
|--------------------------------|----|
| 第一章 绪论..... | 1 |
| 1.1 研究背景和意义..... | 1 |
| 1.2 国内外研究现状..... | 2 |
| 1.2.1 不同识别技术的研究现状..... | 2 |
| 1.2.2 不同活动类型的研究现状..... | 3 |
| 1.3 本文研究内容..... | 5 |
| 1.4 论文结构安排..... | 5 |
| 第二章 人体动作识别理论基础 | 7 |
| 2.1 基于 WiFi 的人体动作识别原理 | 7 |
| 2.1.1 正交频分复用技术..... | 7 |
| 2.1.2 多入多出技术..... | 7 |
| 2.1.3 人体动作识别模型..... | 8 |
| 2.1.4 人体动作识别基本框架..... | 9 |
| 2.1.5 信道状态信息..... | 10 |
| 2.2 基于 WiFi 的人体动作识别相关方法 | 12 |
| 2.2.1 传统机器学习方法..... | 12 |
| 2.2.2 深度学习方法..... | 12 |
| 2.3 本章小结..... | 14 |
| 第三章 基于多层双向长短期记忆网络的人体动作识别 | 15 |
| 3.1 引言 | 15 |
| 3.2 系统框架设计 | 15 |
| 3.2.1 MBLSTM 神经网络..... | 16 |
| 3.2.2 注意力模型..... | 17 |
| 3.2.3 网络训练方法..... | 17 |
| 3.3 实验设置 | 18 |
| 3.4 实验结果 | 21 |
| 3.5 本章小结 | 22 |
| 第四章 基于 WiFi 的人体动作识别可视化应用 | 23 |
| 4.1 数字孪生概述 | 23 |
| 4.2 人体动作识别数据集收集 | 25 |
| 4.2.1 原始 WiFi 数据收集..... | 25 |
| 4.2.2 数据预处理..... | 28 |
| 4.3 系统框架设计 | 31 |
| 4.3.1 传感模块..... | 32 |

| | |
|--------------------------|----|
| 4.3.2 元学习框架..... | 32 |
| 4.3.3 Bi-Path 基础网络 | 35 |
| 4.3.4 数字孪生系统搭建..... | 37 |
| 4.4 实验设计与分析..... | 39 |
| 4.4.1 实验对比基线..... | 39 |
| 4.4.2 超参数选择..... | 40 |
| 4.4.3 实验结果..... | 42 |
| 4.5 本章小结..... | 44 |
| 第五章 总结与展望 | 45 |
| 5.1 本文总结..... | 45 |
| 5.2 本文展望..... | 45 |
| 参考文献..... | 47 |

摘要

得益于物联网和传感器技术的发展，人体动作识别（Human Activity Recognition, HAR）在智能家居、医疗保健和摔倒检测等领域取得了广泛的应用。人体动作识别一般分为两种类型：接触式的和非接触式。接触式的方法会给身体带来额外的负担，基于 WiFi 的人体动作识别方法因其部署方便、不受视觉限制和不侵犯人们隐私等优势受到了研究者的关注。同时，在计算机运算速度不断提高的情况下，利用深度神经网络对 WiFi 信号进行特征抽取，以达到更高精度的分类，这也加速人类动作识别的发展。

本文针对目前一些基于 WiFi 的人体动作识别技术中识别准确率低、适应环境能力差、缺少训练数据集、无法可视化等问题，研究了一种利用多层双向长短期记忆网络（Multi-layer Bi-directional Long Short-Term Memory Network, MBLSTM）的深度学习方法提高识别准确率。并且使用元学习（Meta-Learning）训练 WiFi 数据，提高人体动作识别系统适应环境的能力，同时开发人体动作识别的数字孪生（Digital Twins, DT）系统，主要内容包括：

(1) 针对基于 WiFi 的人体动作识别准确率低的问题，本文提出了 MBLSTM 的深度学习的方法用于对人类进行非接触的活动检测。该方法可以从原始的 WiFi 时间序列中学习前向和后向的活动特征，并引入注意力机制用来给学习到的特征分配不同的权重，最后进行活动识别。实验结果表明，本文提出的方法在多轮测试中对六种活动的识别达到了 96% 以上的准确率，超过了其他用于比较的基准方法。

(2) 针对人体动作识别无法可视化的问题，本文创新性的将人体动作识别和数字孪生进行了结合。本文使用元学习进行人体动作识别网络的训练，增加网络适应环境的能力，然后将识别的结果发送给数字孪生系统。在数字孪生系统中，真实世界的物理活动被映射到数字空间人类模型上，并评估活动是否安全，当危险活动发生时可以迅速发出警告。同时，网络的训练需要大量的数据，本文使用了一个公共的 WiFi 收集工具来收集数据，它能够记录从发射器到接收器传输的 WiFi 信号，将收集到的数据制作成一个公开的数据集。

关键词：人体动作识别，无线感知，深度学习，信道状态信息，数字孪生

Abstract

Thanks to the development of IoT and sensor technology, Human Activity Recognition (HAR) has achieved a wide range of applications in smart home, healthcare and fall detection. Human activity recognition is generally divided into two types: contact and non-contact. The contact method puts extra burden on the body. WiFi-based human activity recognition methods have attracted the attention of researchers because of their advantages such as easy deployment, no visual restrictions and no invasion of people's privacy. Meanwhile, with the increasing speed of computer computing, feature extraction of WiFi signals using deep neural networks to achieve higher accuracy classification has accelerated the development of human activity recognition.

In this paper, we address the problems of low recognition accuracy, poor adaptation to environment, lack of training dataset, and inability to visualize in some current WiFi-based human activity recognition techniques, and study a deep learning using Multi-layer Bi-directional Long Short-Term Memory Network (MBLSTM). A deep learning method using Multi-layer Bi-directional Long Short-Term Memory Network (MBLSTM) is investigated to improve the recognition accuracy. We also use Meta-Learning to train WiFi data to improve the ability of human activity recognition system to adapt to the environment, and develop Digital Twins (DT) system for human activity recognition.

(1) To address the problem of low accuracy of human activity recognition based on WiFi, this paper proposes a deep learning method of MBLSTM for non-contact activity detection of humans. The method can learn forward and backward activity features from the original WiFi time series, and introduce an attention mechanism to assign different weights to the learned features, and finally perform activity recognition. The experimental results show that the proposed method achieves an accuracy of more than 96% for the recognition of six activities in multiple rounds of testing, which exceeds other benchmark methods used for comparison.

(2) To address the problem that human activity recognition cannot be visualized, this paper innovatively combines human activity recognition and digital twin. This paper uses meta-learning to train the human activity recognition network and increase the network's ability to adapt to the environment, and then sends the recognition results to the digital twin system. In the digital twin system, real-world physical activities are mapped onto a digital spatial human model, and the activities are evaluated for safety, and warnings can be issued quickly when dangerous activities occur. Meanwhile, the training of the network requires a large amount of data, and this paper uses a public WiFi collection tool to collect the data, which is able to record

the WiFi signals transmitted from the transmitter to the receiver, and make the collected data into a public dataset.

Keywords: Human Activity Recognition, Wireless Perception, Deep Learning, Channel State Information, Digital Twin

第一章 绪论

1.1 研究背景和意义

在信息技术飞速发展的今天，物联网等新兴技术已经深入到人们的日常生活中。物联网终端通过与外界信息的交互，能够对各种物体进行智能探测、识别等。这些技术使得物联网设备可以更加智能地感知和响应周围环境，从而实现更高效的数据采集和处理。人体动作识别一直是物联网技术研究的热点和关注重点，吸引了众多研究者的关注和投入。它被应用于许多领域，例如智能家居^[1-3]、医疗保健^[4, 5]、跌倒检测^[6-8]等。

人体动作识别技术可以根据是否需要用户佩戴特定设备来分为接触式和非接触式。接触式人体动作识别需要用户佩戴传感器设备，通过监测生理信号和环境变化等特征来实现人体识别。这种方法在军事、医疗等领域广泛应用，因其成本低廉、反应速度快等优点而得到广泛关注。但是，由于用户需要佩戴设备，这种方法在实际应用中存在一些问题，例如用户可能会忘记佩戴设备或在洗澡或睡觉时无法佩戴设备，不够方便和用户友好。因此，近年来的人体动作识别研究更偏向于非接触式方法，避免了用户佩戴设备的负担。

非接触式人体动作识别的具体应用可以分为检测入侵^[9]、人群计数^[10, 11]、呼吸监测^[12, 13]和活动识别^[14, 15]等。检测入侵是指通过无线信号采集家中成员的身高体重等个人信息组建指纹库，从而能检测出不属于该环境的人的闯入；人群计数是指研究某一室内环境中人数变化对信号产生的影响得到信号变化情况与具体人数的关系；呼吸监测是现代医学中用于监测人的生命体征的方法，由微小的呼吸时产生的胸部位移引起的相位变化来检测呼吸；活动分类识别是指在同一环境中采集多个动作数据，通过一定的方法将各个动作准确分类。此类方法使用基于视觉的图像数据^[16, 17]、声音^[18, 19]和无线信号^[20, 21]进行人体动作识别。微软的 Kinect^[22] 使用了三维深度摄像头，能够捕捉用户的肢体语言和面部活动。但基于视觉的人体动作识别仍然受到光照条件和视觉盲区的局限，在实际应用时还有入侵用户隐私的隐患。同样，基于声学传感的方法在无噪声环境中表现良好，但通常需要庞大的设备并产生高能耗，因此在实践中很少采用。

随着计算机网络技术的快速发展，无线网络和相关应用在人们的日常生活中变得越来越重要。无论是在繁忙的工作场所还是在舒适的家庭环境，无线网络都在被广泛应用。由于无线网络的广泛普及，人们开始思考是否可以将其应用于动作感知领域，这也引发了对于基于 WiFi 的人体动作识别研究的兴趣。正是因为无线设备的大规模部署，为无线动作感知技术的发展和应用创造了良好的前景。

基于 WiFi 技术的人体动作识别是通过采集无线信道状态信息（Channel State Infor-

mation, CSI), 研究人体动作对 WiFi 传输路径的影响, 从而实现对不同人体动作的识别。与常规的人体动作识别技术相比, 基于 WiFi 的方法有更明显的优点。首先, 它成本低且易于在各种环境中部署。其次, 作为非接触式方法, 无需用户佩戴传感器设备或要求目标配合, 适用范围广泛, 即使在无光的条件下也能完成识别任务。最后, 作为一种非视觉方法, 不易侵犯用户的隐私, 尤其适用于家庭等私密空间。然而, 由于不同的室内环境及人体行为在识别方式上存在差异, 如何在不同的场景下采用不同的识别策略, 从而实现高精度、高可用性、高可靠性的行为识别是一个极具挑战性的课题。本文将在已有工作的基础上, 进一步深入探讨人体动作识别技术。

1.2 国内外研究现状

1.2.1 不同识别技术的研究现状

人工智能技术的飞速发展, 使得模式识别、计算机视觉、无线感知等领域成为研究的热点。在这些领域中, 人体动作识别技术是备受关注的研究方向。近年来, 国内外知名专家教授, 都积极研究和探索人体动作识别技术。同时, 该技术已经应用于智能家居、虚拟现实、医学监测等领域, 有着广泛的发展前景。总体而言, 当前动作识别技术主要是基于传感器技术、视觉技术和无线射频技术。

通过传感器对人体的活动进行识别时, 有的需要参与者佩戴加速度计、惯性传感器、陀螺仪或其他设备收集信号, 有的通过放置在房间各个位置的传感器采集信号。例如 Mohammed 等人^[23] 从可穿戴传感器收集复杂的人类活动, 利用带有注意力机制的多级残差网络提取时间序列特征并执行活动识别, 并和现有人体动作识别系统进行了比较, 取得了显著的性能。Xu 等人^[24] 提出了一种新的可变形卷积网络, 用于从可穿戴传感器收集复杂的感官数据中识别人类活动, 在不同的数据集上, 分别达到 82.91%、80.02%、97.35% 和 99.21% 的准确率。包括加速度计、陀螺仪、惯性传感器、磁力计在内的智能手机已经成为人们生活中不可或缺的一部分, 它也非常适合检测人类活动。最近的研究中智能手机的引入让基于传感器的人类活动识别研究有了新的生机。Jamil 等人^[25] 利用智能手机加速度计和气压计进行活动识别, 加速计数据用于通过观察传感器测量值的大小来检测活动, 气压计传感器通过使用压力和高度数据来检测地板高度。

基于视觉的人体动作识别指的是在实验环境中布置摄像机或者受试者佩戴摄像机采集图像数据, 并对获取到的图像数据进行处理, 以识别目标在特定时刻所执行的动作。摄像头作为传感器的一种, 随着计算机视觉研究的兴起, 在人体动作识别中也逐渐成为热门。Moniruzzaman 等人^[26] 设计了一个简单有效的网络, 以从视频中识别人类活动。该网络包含一个特征库和一个视频片段关注模块。他们的动作识别网络可以用完全

监督的和弱监督的未修剪的视频进行训练。Luo 等人^[27] 提出了一种新的深度架构，称为密集语义辅助卷积神经网络（Dense Semantics-Assisted Convolutional Neural Networks, DSA-CNNs），能够通过自下而上的关注方式在空间流中有效利用视频的密集语义信息进行活动识别。然而基于视频的人体识别技术需要较高的环境光线来保证图像采集的质量，当光线较暗时，该技术的可靠性会受到影响，无法正常工作。

以无线射频为基础的动作识别，在诸如雷达之类的特殊装备中最为普遍。毫米波雷达具有非接触以及能够以高分辨率捕获场景的优势，高分辨率可以对目标进行详细地分析并开发出先进的数据处理算法。Cui 等人^[28] 使用毫米雷达波在室内建立了一个人体姿态检测系统，他们使用两个毫米雷达波进行姿势捕捉，并使用一个神经网络，其中包括一个用于关节位置的分量检测器和一个用于学习关节之间相关性的空间模型来进行姿势识别。然而，这类专业设备通常具有体积较大、能耗高等缺点，加之便携性较差，使得其难以被广泛普及应用。

表 1.1 不同识别技术研究

| 识别技术 | 文献 | 方法 | 识别准确率 |
|-------|------|-------------------|--------|
| 基于传感器 | [23] | 多级注意力残差网络 | 87.82% |
| | [24] | 可变形卷积神经网络 | 95.36% |
| | [25] | 基于自动学习和加权软投票的最佳集合 | 96.31% |
| 基于视觉 | [25] | 判别特征池机制和视频段注意模型 | 98.42% |
| | [27] | 密集语义辅助卷积神经网络 | 96.69% |
| 基于射频 | [28] | 使用卷积层关节之间的相关性进行建模 | 71.30% |

1.2.2 不同活动类型的研究现状

随着现代社会对智慧生活的需求不断提高，无线网络的无线通信技术也在不断发展，以弥补现有技术的缺陷。将无线网络信号与动作感知融合是当前的研究热点，基于无线网络的动作识别技术受到了众多高校、科研院所和企业的高度重视。与传统的动作识别技术相比，WiFi 动作识别技术对光线环境无约束，对设备条件要求较低，在智能家居、体感游戏等方面有着广阔的应用前景。由于无线网络的普及，在社会上大多数的娱乐场所和家庭中都配备了无线设备，这为 WiFi 动作识别的发展提供了最有力的保证，同时也显示出了其成本低、普适性强、操作方便等优点。当前，无线环境下的运动检测技术尚处在初级阶段，但其应用前景十分广阔。根据研究动作幅度的大小可大致分为两个方向：细粒度动作识别和粗粒度动作识别。

对粗粒度动作识别的研究中，Muta 等人^[29] 提出了一种无设备的室内定位方案，在

具有分布式天线的无线网络系统中使用轻量级的机器学习模型和压缩的空间串联 CSI, 用少量的数据集进行低复杂度的实时目标位置检测。天津大学佟鑫宇等人^[30]提出了基于 CSI 的室内人体定位系统 MapFi, 它可以实现 WiFi 的自主测绘, 可以在办公室和开放空间分别实现 1.15 米和 0.74 米以内的 80% 定位误差。Li 等人^[31]提出了一种双流卷积增强的人类活动转化器模型使用双流结构来捕捉时间-通道和通道-时间的特征, 并利用多尺度卷积增强变换器来捕捉基于范围的模式。通过对四个真实实验数据集的广泛实验, 证明了所提模型的有效性和效率, 显示出比最先进的模型更好的性能。郑州大学段鹏松等人^[32]设计了基于 CSI 的多用户活动识别框架 WISDOM, 他们建立一个神经网络 WISDOM-Net, 使用双向门控递归单元网络, 结合了注意力机制和一维卷积神经网络, 以识别相应的用户活动。WISDOM-Net 在 2 个用户和 3 个用户场景下的平均识别准确率分别高达 98.19% 和 90.77%。

表 1.2 不同动作类型研究

| 动作类型 | 文献 | 方法 | 识别准确率 |
|------|------|---------------------------------|--------|
| 粗粒度 | [29] | 串联反馈波束形成权重 | 90% |
| | [30] | 基于不同天线阵列布局的到达角 | 80% |
| | [31] | 双流卷积增强人类活动转化器模型 | 98.60% |
| | [32] | 结合注意力机制和一维卷积神经网络的 双向门控递归单元网络 | 98.19% |
| 细粒度 | [33] | 残差多头模型 | 95.03% |
| | [34] | 前导手势方案估计强相位噪声影响下 动态反射的空间信道 | 95.60% |

细粒度的动作通常是指具有较小的运动幅度, 比如手势、手语、键盘按键识别等。Zhang 等人^[33]设计了一个多策略上下文感知手语识别系统 Wi-Phrase, Wi-Phrase 采用了主成分分析投影来过滤噪音, 并将清理过的 WiFi 信号转换为频谱图, 利用 WiFi 信号将 SL 转换为英语短语。他们综合评估表明, 该系统在包含 40 个手语类别的数据集中实现了 95.03% 的短语识别准确率。北京航空航天大学彭睿等人^[34]为了解决多人同时移动会造成动态信道响应叠加, 并造成相互干扰的问题, 他们提出了一种前导手势方案来估计强相位噪声影响下动态反射的空间信道, 并使用多普勒变化特征来验证信道响应是否属于手部运动的方法进行多用户手势识别。他们深入分析了发射机-用户-接收机几何关系、用户移动速度和方向对用户间干扰的影响。在使用 LTE 信号的真实视距和非视距场景下, 通过原型实验证明了干扰抑制效果。

1.3 本文研究内容

通过以上对于国内外人体动作识别的研究现状分析可知，人体动作识别在智能感知领域取得了巨大进步。然而，现有的人体动作识别方法在检测精度、可视化、环境适应性等方面存在着巨大挑战，而且缺少基于 WiFi 的人体动作识别数据集的使用。本文从以下两个方面对基于 WiFi 的人体动作识别进行研究：

(1) 为了提高基于 WiFi 的人体动作识别准确率，本文提出了一个基于深度学习的框架，用于对人体进行非接触实时活动检测，该框架被称为多层双向长短期记忆(Multilayer Bi-directional Long Short-Term Memory, MBLSTM)。收集到的 WIFI CSI 数据被送入提出的网络模型，然后用来从原始的连续 CSI 测量中学习双向的代表性特征。并利用注意力模型为学习到的特征分配不同的权重，最后进行活动识别。实验结果表明，本文提出的方法在多轮测试中对六种活动的识别达到了 96% 以上的准确率，超过了用于比较的其他基准方法。

(2) 针对人体动作识别无法可视化的问题，本文创新性的将人体动作识别和数字孪生进行了结合。本文设计了一个高效的人体动作识别系统，称为 HAR-Net，它使用传感器收集的 WiFi 时间序列数据来训练一个深度学习网络，使用元学习增加网络适应不同环境的能力。同时，本文用 Unity 建立了数字孪生系统，它与人体动作识别系统相结合。在数字孪生系统中，真实世界的物理活动被映射到人类模型上。活动预测的结果可以在数字孪生中得到实时评估，当危险活动发生时可以迅速发出警告。为了解决训练数据集少的问题，本文搭建获取 CSI 数据的实验环境，收集不同动作对应的 CSI 原始数据包，制作基于 WiFi 的人体动作识别数据集。详细的记录收集活动数据的具体步骤，然后给出数据预处理的方法。

1.4 论文结构安排

本文主要由五个章节组成，图1.1为论文组织结构图，具体结构安排如下：

第一章：绪论。对人体行为识别的研究背景和关联性进行了说明，并对国内外有关技术的研究状况进行了介绍。重点介绍了人体动作识别大方向的发展现状，最后介绍主要研究课题和论文框架。

第二章：介绍基于 WiFi 的人体动作识别理论基础。描述了自由空间中信号传播模型和信道状态信息的使用，然后介绍了基于 WiFi 人体动作识别的常用方法，最后对第二章进行了一个总结。

第三章：基于 MBLSTM 的人体动作识别方法。首先描述本文提出的系统框架，其

次介绍 MBLSTM 网络的原理，然后说明注意力机制在系统中的作用和网络训练的方法，最后介绍了实验设置和实验结果。

第四章：人体动作识别的可视化应用。首先对数字孪生技术进行了概述，其次介绍基于 WiFi 的人体动作识别的数字孪生系统实验设计，包括实验环境、实验中进行识别的动作、实验数据集收集、评价指标以及相应的对比实验。然后具体介绍在实验过程中一些函数的参数选择，最后简述实验过实验结果。

第五章：总结与展望。对本文研究内容进行了总结，并且对人体动作识别的未来发展做出了展望。

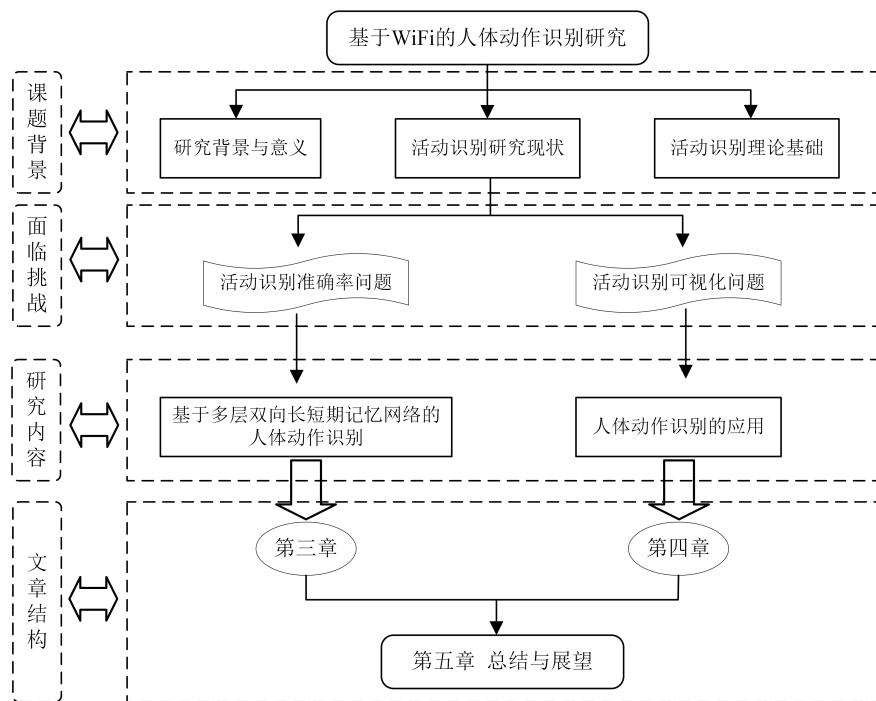


图 1.1 论文组织结构

第二章 人体动作识别理论基础

随着环境感知受到越来越多的关注，人们对于室内环境下活动识别的需求不断增加，WiFi 感知成为了很多科学研究团队的研究焦点。在无线通信中，发射的无线电信号受到物理环境的影响，而这些信号也可以反映物理环境的变化。本章主要阐述无线信号的传播特性、无线信道状态信息的应用以及神经网络的理论基础内容。

2.1 基于 WiFi 的人体动作识别原理

2.1.1 正交频分复用技术

正交频分复用技术 (Orthogonal Frequency Division Multiplexing, OFDM) 是从典型的频分复用 (Frequency-division Multiplexing, FDM) 多载波调制方式的基础之上发展而来。OFDM 技术是一种多载波调制技术，可以将一个高速数据流分成多个低速子流，每个子流都在不同的子载波上进行传输，这样可以降低单个子载波的传输速率，从而提高抗多径衰落和干扰的能力。OFDM 技术的主要特点是子载波之间正交，即每个子载波的正弦和余弦信号都正交于其他子载波的正弦和余弦信号，这种正交性可以避免子载波之间的干扰，OFDM 系统模型如图2.1所示。OFDM 技术被广泛应用于无线通信、数字广播、数字电视等领域。

在人体动作识别中，OFDM 技术可以用于传输和处理传感器采集的人体运动数据。具体来说，OFDM 技术可以将传感器采集的人体运动数据转化为频域信号，并将信号分成多个子载波进行传输，从而提高数据传输效率和准确度。人体动作识别通常使用加速度计、陀螺仪和磁强计等传感器来获取人体运动数据。这些传感器采集的信号可以通过傅里叶变换等信号处理方法转化为频域信号。在使用 OFDM 技术传输人体运动数据时，可以将频域信号分成多个子载波，将每个子载波分别传输，从而避免了多路传输的干扰问题。接收端使用反向傅里叶变换将接收到的信号转换回时域信号，然后使用分类算法对人体活动进行识别。

2.1.2 多入多出技术

多入多出技术 (Multiple Input Multiple Output, MIMO) 是一种利用多个天线进行数据传输的技术，通过利用空间资源，可以在有限的频谱带宽内提高数据传输速率和可靠性。MIMO 技术最早是在无线通信领域被应用，利用多个天线在同一频率带宽内进行数据传输。在 MIMO 系统中，每个天线之间存在一定的信号干扰和信道衰落，但是通过复杂的信号处理和编码技术，可以实现高速和可靠的数据传输。

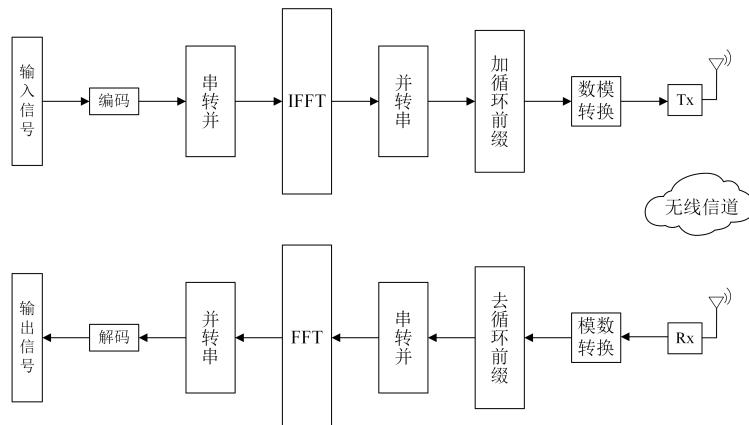


图 2.1 OFDM 系统模型图

MIMO 技术的核心思想是利用空间多样性，也就是通过多个天线接收或发送信号，从而增加传输通路的数量和多样性，提高系统的传输速率和可靠性。在 MIMO 系统中，发射端把数据划分为多个平行的数据流，每一个数据流动经过不同的天线来进行传输，接收端利用多个天线来接收数据，并通过信号处理和解码技术，将多个数据流合并成单个数据流。MIMO 技术可以通过增加天线数量来提高系统性能，同时还可以通过空间编码、调制和信道估计等技术来提高传输的可靠性和抗干扰能力。

在 MIMO 系统中，发射端和接收端都有多个天线，每根发射天线和接收天线之间的组合可以看作一条 CSI 数据流。如图2.2所示，在 3×3 的 MIMO 系统中，发射端有 3 根天线，接收端也有 3 根天线，每个组合表示发射端第 i 根天线和接收端第 j 根天线之间的 CSI 数据流 H_{ij} 。MIMO 系统利用这些数据流提高通信速率和可靠性，从而在无线通信领域得到了广泛应用。

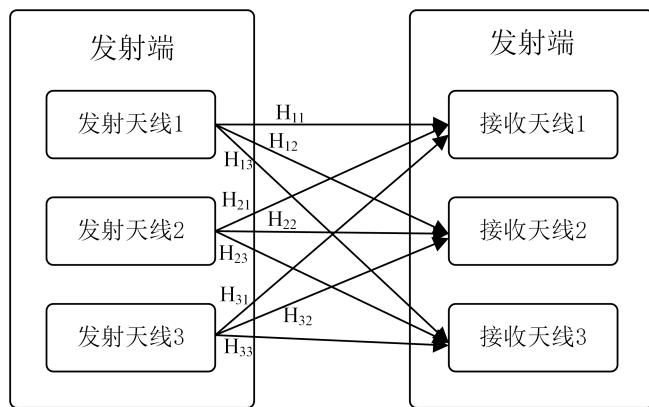


图 2.2 多入多出技术模型图

2.1.3 人体动作识别模型

在自由空间中，无线信号由发射设备产生，在传播的时候受到物理空间和人体动作的影响，产生多条传播路径，多径信号携带了大量的有关当前室内环境特征的信息，并在接收设备端接收。接下来，通过理论角度证明无线信号可以用来识别人体活动。

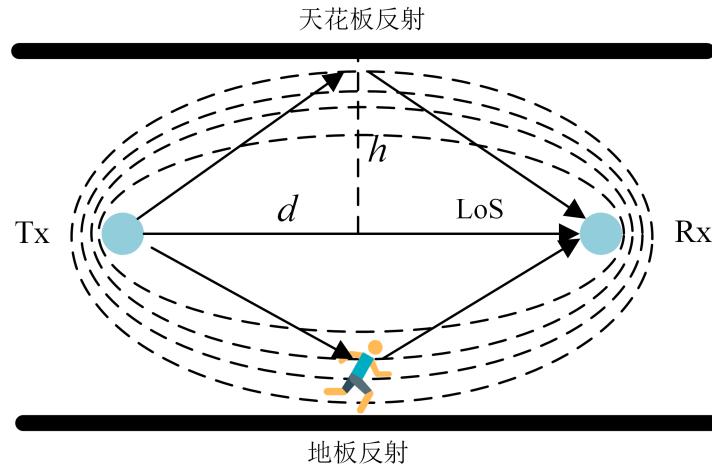


图 2.3 基于 WiFi 的人体动作识别模型

发射器 (AP) 与接收器 (PC) 放置在静态室内环境下, AP 端的无线信号会直接到达发射器端, 也会通过天花板、墙壁等物体的反射到达, 因此 AP-PC 链路存在一条主路径 (Line-Of-Sight, LOS) 和多条反射路径 (None Line-Of-Sight, NLOS), 如图2.3所示。当目标出现在这种静态环境中时会出现由人体造成的多条分散路径。据自由空间传播方程, 室内环境下的接收功率的表达式为:

$$P_{rx}(d) = \frac{P_{tx}G_{tx}G_{rx}\lambda^2}{(4\pi)^2d^2} \quad (2-1)$$

其中, P_{tx} 和 $P_{rx}(d)$ 代表的是发射机的发射信号功率和接收机的接收信号功率, G_{tx} 和 G_{rx} 代表的是发射增益和接收增益, λ 是 WiFi 信号的波长, ε 是因人体散射信号造成的路径长度变化的近似值。从公式2-1中可以看出, 不同的动作会导致接收端接收到的能量存在一定的区别, 而这种区别与动作间的对应关系是 WiFi 人体行为感知的基础。

2.1.4 人体动作识别基本框架

图2.4展示的是人体动作识别系统的基本框架, 主要有四方面组成。

- (1) **数据采集:** 通过布置 WiFi 信号接收器和发射器采集周围环境中的信号强度信息。
- (2) **数据预处理:** 对采集到的 WiFi 信号强度信息进行预处理, 包括信号过滤、信号增强、信号归一化等步骤。预处理后的数据更有利于特征提取和模型训练。
- (3) **特征提取:** 根据采集到的 WiFi 信号强度信息, 通过特征提取算法将数据转化为一组特征向量, 以描述人体活动的不同状态。常用的特征提取算法包括小波变换、离散傅里叶变换、时频分析等。
- (4) **训练和识别:** 将选定的特征输入到机器学习模型中进行训练, 以识别不同的人体活动状态。常用的机器学习模型包括支持向量机、决策树、神经网络等。

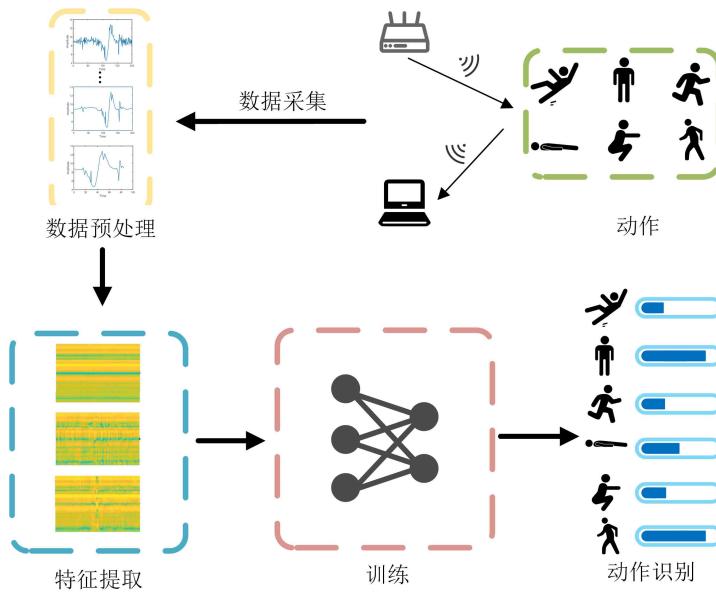


图 2.4 人体动作识别系统的基本框架

2.1.5 信道状态信息

在介绍信道状态信息之前,有必要先提到接收信号强度指示(Received Signal Strength Indicator, RSSI)。RSSI 是射频通信中常用的术语,主要用于测量发射机和接收机之间的距离。它基于接收信号的能量强度进行测量,并需要满足良好的通信信道特性。RSSI 测距技术的原理是基于无线电波或声波在介质中传输的规律,信号功率随着距离的增加而逐渐衰减。节点之间的距离可以通过衰减模型和已知的信标节点发射功率以及节点接收到的信号功率来确定。信号传播会受到距离和障碍物体的阻碍,信号强度减弱,这对准确性也有间接影响。因此,要想达到识别的高精度,这仅在短距离内可实现,所以早期的 RSSI 被广泛应用在定位技术中^[35-38]。图2.5a显示了步行活动的 RSSI 信号图。它表示所有可能信号路径的总能量,包括发射器和接收器之间的直接视线。使用 dBm 表示 RSSI 的接收信号功率。

在无线感知的早期,人们通过无线网络的 RSSI 状态来获取用户的行为。设 P 是信号传输通道的个数, $\|S_i\|$ 是第 i 个通道的振幅,而 α_i 是第 i 个通道的相位,那么在 WiFi 感知信道中的 RSSI 可以被表达为:

在早期 WiFi 感知中,利用 RSSI 的变化情况感知人体活动。设 P 为信号传播路径的数量, $\|S_i\|$ 为第 i 条路径的振幅, α_i 为第 i 条路径的相位,则 WiFi 感知信道中 RSSI 可以表示为:

$$RSSI = 10 \log_2 \left(\left\| \sum_{i=1}^P \|S_i\| e^{-j\alpha_i} \right\|^2 \right) \quad (2-2)$$

理论上,可以利用接收信号强度指示的数值来确定被测物体与收发设备之间的距离,它可以作为被测目标信息的指纹。由于 RSSI 的性质是对接收信号的强依赖性,对

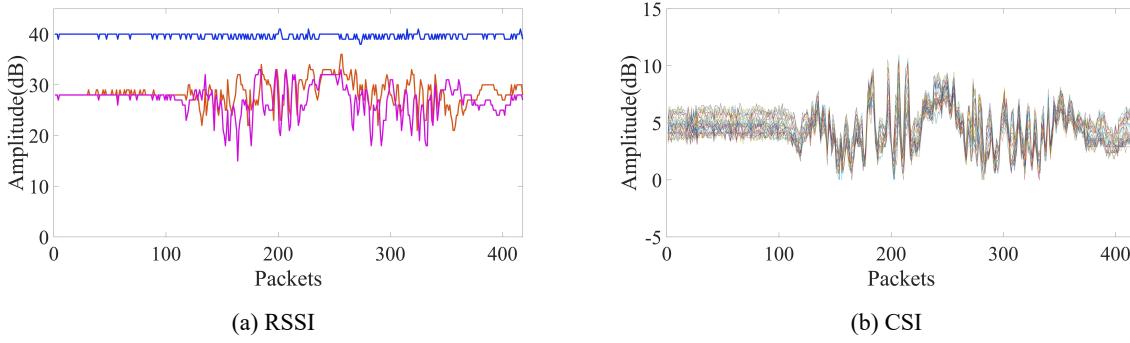


图 2.5 步行活动的 RSSI 和 CSI 图

多路径、窄带等因素的敏感，使得 RSSI 的探测准确率较低。为了克服 RSSI 存在的固有缺陷，一种基于物理层且更细粒度的 WiFi 信道特征 CSI 被发掘。相对于 RSSI，CSI 能利用 OFDM 对各子载波进行信道特性提取，能更好地刻画无线网络中各子载波的衰减特性，从而有效抑制无线网络中多途分量和窄带干扰对无线网络性能的影响。此外，CSI 包含丰富的频域信息，如振幅和相位，能更好的体现出人体对 WiFi 信号的作用，从而达到更高的感知精度。图2.5b显示了受行走活动影响的 30 个子载波，与 RSSI 相比，它包含更多的信息。

具有强稳定性并包含物理层信息的子载波级别是 CSI 的分辨率级别。一些带有开源固件的网卡可以和它一起使用。它指定了频域空间中每个子载波的相应幅度和相位，并结合了许多影响，例如延迟、幅度衰减和相位，以表征信号如何通过信道传播产生的偏移量。CSI 是从对周围环境更敏感的 OFDM 子载波派生的细粒度物理信息，可用于运动检测、手势识别、击键识别、活动跟踪和其他应用。相比于 RSSI 中包含的粗粒度信息，CSI 可以提供更详细、更丰富的用于人体动作识别的信息。

信道可以被 $Y = HX + N$ 模型表示。 Y, X 分别表示接收和发射的信号向量， N 代表信道噪音向量， H 是信道矩阵，信道矩阵 H 以信道状态信息的格式表示，即按照当前的 WiFi 标准求信道状态信息就是求信道矩阵 H 。在静态场景下，设发射信号的等效基带信号为 $x(t)$ ，其共沿 N_p 条路径到达接收机，忽略噪声的影响，则接收端的等效基带信号为：

$$y(t) = \sum_{n=1}^{N_p} a_n(t) e^{-j2\pi f_c \tau_n(t)} x(t - \tau_n) \quad (2-3)$$

其中 LOS 路径上 $n = 1$ ， $a_n(t)$ 表示时刻 t 时第 n 条路径上的信号的幅值。 $\tau_n(t)$ 表示第 n 条路径的传播时延， f_c 为载波频率， $-j2\pi f_c \tau_n(t)$ 则为时延 $\tau_n(t)$ 造成的相位偏移。由此可见，由于 $a_n(t)$ 、 $\tau_n(t)$ 的不断变化，多径信号的振幅和相位受到相应影响也将不断变化，从而导致多径衰落。

2.2 基于 WiFi 的人体动作识别相关方法

2.2.1 传统机器学习方法

传统的机器学习算法模型中，最原始的分类方式是通过为所有训练样本和测试样本打上标签来表示它们的所属类别。然后，使用单一分类器将测试样本与训练样本进行比对以进行分类。然而，在实际应用中，训练样本和测试样本的属性不可能完全一致。当测试样本的属性与多个训练样本的属性都相似时，需要使用适当的算法来进行分类。因此，如何选择合适的算法来处理这些问题是非常重要的。由此，K 最近邻 (K-Nearest Neighbor, KNN) 应运而生。KNN 方法根据新的样本与已有样本的距离，确定最近的 K 个邻居，并将新样本分类为出现最多的类别。在基于 WiFi 的人体动作识别中，可以利用已知的人体活动数据集，计算出信号强度、信号时延、相位等特征之间的距离，根据距离来确定新的样本所属的人体活动类别。例如，Abidine 等人^[39] 提出了一个新的方案 SV-KNN，该方案基于内核 KNN，使用基于支持向量 (Support Vectors, SV) 的紧凑训练数据来识别正在进行的活动。

决策树算法 (Decision Tree, DT) 是一种基于树结构的分类方法，它通过将数据分成不同的子集来实现分类。在基于 WiFi 的人体动作识别中，可以使用信号强度、时延、相位等特征来构建决策树，通过选择最优的特征和阈值来划分不同的子集，并逐步判断新样本所属的类别。Fan 等人^[40] 对从嵌入在智能手机中的加速度计来对静止、行走、跑步、上楼和下楼这五种活动进行分类。他们使用决策树算法分析来自加速度计的行为数据，提取各种特征，选择高度相关的特征，并构建一个基于位置无关的智能手机的活动识别模型。

支持向量机 (Support Vector Machine, SVM) 分类算法将训练集投影到高维空间，并找到一个超平面，使得两类样本之间的距离最大化。在基于 WiFi 的人体动作识别中，可以使用信号强度、时延、相位等特征来训练 SVM 模型，找到一个最优的分类超平面，使得不同类别之间的距离最大化。

2.2.2 深度学习方法

卷积神经网络 (Convolutional Neural Network, 简称 CNN) 是一种深度学习模型，广泛应用于图像识别、语音识别和自然语言处理等领域。在基于 WiFi 的人体动作识别任务中，使用 CNN 可以将 WiFi 信号转换为图像，即将 WiFi 信号的强度值转换为像素值，然后通过 CNN 学习 WiFi 信号图像的特征，并对不同的人体活动进行分类。这种方法可以实现无需额外传感器或设备，仅通过 WiFi 信号就可以实现人体动作识别，具有实用性和成本效益。

循环神经网络（Recursive Neural Network, RNN）凭借其顺序建模能力，在各种具有挑战性的任务中找到了成功的应用，如语言理解和视频处理。然而，RNN 在学习长序列时面临着梯度消失和爆炸的问题，限制了其性能。为了解决这个问题，Hochreiter 和 Schmidhuber 介绍了一种新的 RNN 架构，名为长短期记忆网络（Long Short-Term Memory, LSTM）。LSTM 采用了具有多个门的记忆单元来捕捉具有长期依赖性的基本信息，从而解决了梯度问题。由于 WiFi CSI 信号的时间依赖性，LSTM 已被广泛用于基于 WiFi CSI 的人类活动识别，并取得了令人印象深刻的结果。

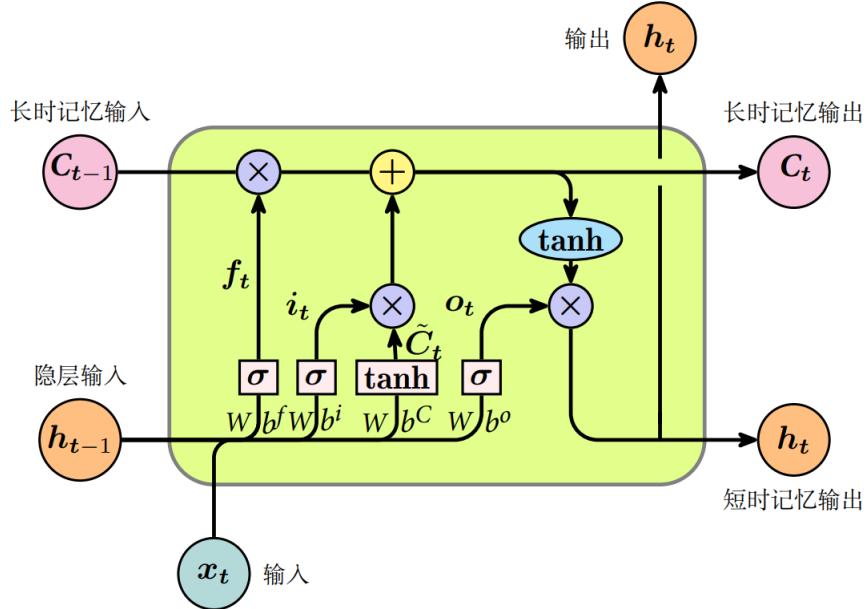


图 2.6 LSTM 网络框架

如图2.6所示，LSTM 在时间 t 基于其输入 x_t 、 h_{t-1} 和 C_{t-1} 的方式更新自己。

$$\begin{aligned}
 f_t &= \sigma(W^f [h_{t-1}, x_t] + b^f) \\
 i_t &= \sigma(W^i [h_{t-1}, x_t] + b^i) \\
 \tilde{C}_t &= \tanh(W^C [h_{t-1}, x_t] + b^C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W^o [h_{t-1}, x_t] + b^o) \\
 h_t &= o_t * \tanh(C_t),
 \end{aligned} \tag{2-4}$$

其中 $\{W^f, W^i, W^C, W^o, b^f, b^i, b^C, b^o\}$ 是权重和偏置。 $\sigma(\cdot)$ 和 $\tanh(\cdot)$ 的函数分别是 sigmoid 和双曲正切激活函数。 $\{h_t, i_t, f_t, o_t, \tilde{C}_t, C_t\}$ 分别是隐藏状态、输入门、遗忘门、输出门、输入调制门和记忆门。记忆单元单元 C_t 由两部分组成，即由 f_t 调制的前一个记忆单元单元 C_{t-1} 和由输入门 i_t 调制的当前输入和前一个隐藏状态所模拟的 \tilde{C}_t 。 i_t 和 f_t 的西格玛性质将自己压制在 $[0,1]$ 的范围内。它们可以被视为 LSTM 学习的旋钮，以选择性地

忘记其先前的记忆或考虑其当前输入。以同样的方式，输出门 o_t 模拟从记忆单元到隐藏状态的转移。基于这些机制，LSTM 学会了在连续的 WiFi CSI 测量中存在的复杂和时间动态，从而在人类活动识别方面有了显著的表现。

传统的 LSTM 网络只能处理单向的 WiFi CSI 测量，这意味着当前的隐藏状态只考虑过去的 CSI 信息。然而，未来的 CSI 信息对人类活动识别也是有意义的。为了实现这种能够同时考虑过去和未来信息的能力，可以利用双向长短期记忆网络（Bidirectional Long Short-Term Memory, BLSTM）。BLSTM 由两层组成，即一个前向层和一个后向层，如图2.7 所示

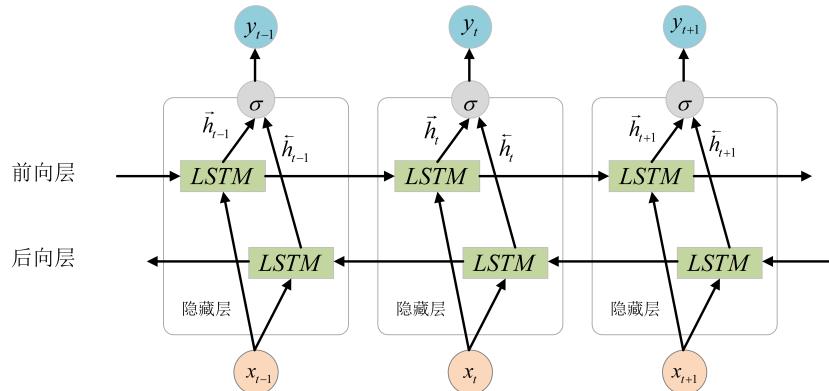


图 2.7 BLSTM 神经网络结构

根据公式2-4，前向层和后向层的隐藏状态可以分别表示为 \vec{h}_t 和 \overleftarrow{h}_t ，其中 \leftarrow 和 \rightarrow 分别表示正向和反向方向。那么，BLSTM 在时间步长 t 的全隐状态，即 h_t 是前向和后向层的隐状态的串联，如下所示：

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (2-5)$$

2.3 本章小结

本章首先介绍 MIMO 和 OFDM 技术的特点，然后阐述了基于 WiFi 的人体动作识别模型。接着又对信道状态信息进行详细的介绍，本章节选择 WiFi 信号作为动作识别信号源，经比较发现 CSI 特性具有更高的鲁棒性，更适合动作识别。在识别分类算法方面，本章简要介绍了 KNN、DT、SVM 等机器学习算法，以及卷积神经网络、长短期记忆网络和双向长短期记忆网络等深度学习方法。

第三章 基于多层双向长短期记忆网络的人体动作识别

3.1 引言

在基于 WiFi 的人体动作识别研究中，准确率低是一个常见的问题。主要原因包括环境干扰、人体姿态和位置变化、选取特征和分类器的影响等。其中，特征选取和训练方法是影响基于 WiFi 的人体动作识别准确率的两个重要因素。目前有一些学者利用 LSTM 网络进行 WiFi 信号的训练^[41]，但是由于 LSTM 网络是单向的，它对输入序列的顺序非常敏感。这意味着它可能对输入序列中前面的信息过于依赖，而无法充分利用后面的信息。在某些任务中，特别是需要全局上下文的任务中，这可能会导致性能下降。

本章节提出了一个基于深度学习的框架，用于使用自供电传感器对人类进行非接触动作检测，该框架被称为多层双向长短期记忆（Multilayer Bi-directional Long Short-Term Memory, MBLSTM）。收集到的 WIFI CSI 数据被送入本章节提出的网络模型，然后用来从原始的连续 CSI 测量中学习双方的代表性特征。并利用注意力模型为学习到的特征分配不同的权重，最后进行活动识别。实验结果表明，本章节提出的方法在多轮测试中对六种活动的识别达到了 96% 以上的准确率，超过了用于比较的其他基准方法。

3.2 系统框架设计

本章节提出的 MBLSTM 框架如图3.1所示。首先，本方案使用一个路由器和一个自供电的传感器来收集来自人类行动的 WiFi 的 CSI 信号。其次，本方案将处理过的 CSI 信号输入 MBLSTM 框架，自动学习前向和后向特征。在本实验中用于特征学习的双向 LSTM 中有 200 个隐藏节点。由于注意力模型没有可用的先验信息，它只能使用从 BLSTM 学到的特征作为输入，得出代表特征和时间步骤重要性的注意力矩阵。

然后，本方案使用元素乘法将学到的特征与注意力矩阵合并，得到带有注意力的修正特征矩阵。之后，将特征矩阵扁平化为特征向量，利用扁平化层进行最终分类。最后，使用 softmax 分类层，用最终的特征向量来识别不同的活动。

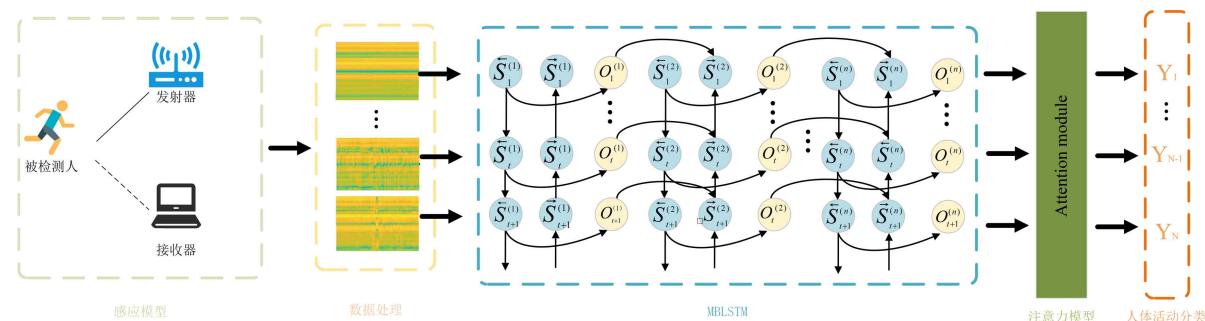


图 3.1 基于 MBLSTM 的人体动作识别框架

3.2.1 MBLSTM 神经网络

在多层堆叠的情况下，BLSTM 神经网络的每一层都由一个前向递归网络和一个后向递归网络组成。前一个层的前向 LSTM 和后向 LSTM 的输出结果的组合被发送到网络的下一个层。图3.2是 MBLSTM 网络框架结构。

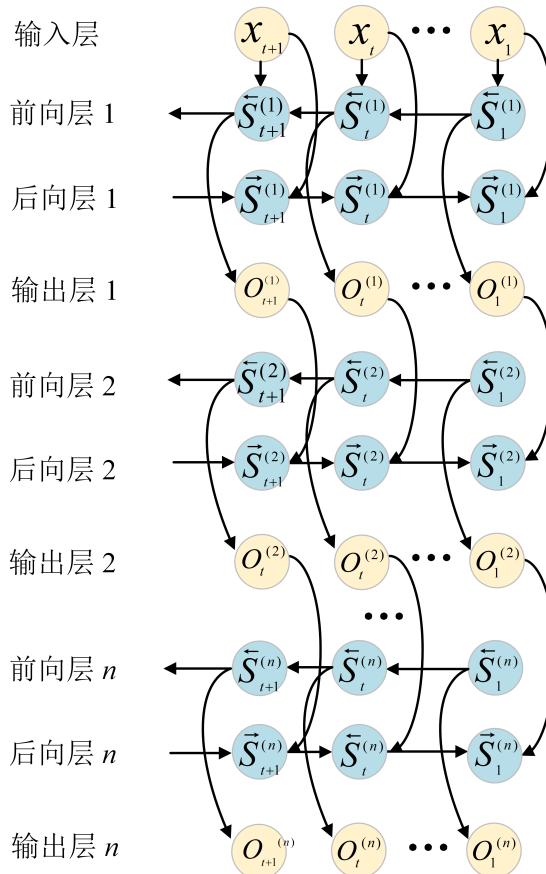


图 3.2 MBLSTM 神经网络结构

输出由各层的正负计算之和决定。其中 $S_{t-1}^{(i)}$ 和 $S_t^{(i)}$ 分别是 i 第 i 个隐藏层在时间 $t-1$ 和 t 的值。正向和反向计算不共享权重， $V^{(i)}$ 、 $U^{(i)}$ 和 $W^{(i)}$ 是 i 第 i 隐藏层到输出层、输入层到隐藏层，以及隐藏层的重量矩阵。 $V^{(i)}$ 、 $U^{(i)}$ 和 $W^{(i)}$ 是用于计算的后向权重矩阵。而 i 是 BLSTM 的层数， $i = 0, 1, 2 \dots \infty$ 是输出层的值。

$$\begin{aligned}
 S_t'^{(1)} &= f(U'^{(1)}x_t + W^{(1)}S_{t-1}') \\
 S_t^{(1)} &= f(U^{(1)}x_t + W^{(1)}S_{t-1}) \\
 &\cdots \\
 S_t'^{(i)} &= f(U'^{(i)}S_t^{(i-1)} + W'^{(i)}S_{t+1}') \\
 S_t^{(i)} &= f(U^{(i)}S_t^{(i-1)} + W^{(i)}S_{t-1}) \\
 O_t &= g(V^{(i)}S_t^{(i)} + V'^{(i)}S_t'^{(i)})
 \end{aligned} \tag{3-1}$$

3.2.2 注意力模型

注意力模型的开发是为了用于图像识别^[42]。这个概念是受人类视觉系统的启发，即在图片识别过程中，人类总是关注图片的某一部分，并随着时间的推移调整他们的注意力。在识别工作中，注意力模型允许计算机在模糊其他区域的同时关注感兴趣的区域。最近，注意力模型被用于语言处理，证明它显然是有效的^[43]。例如，在常用的自然语言处理的编码器-解码器方法中，把输入语句作为一个固定的向量进行编码，在整个翻译过程中进行翻译。这意味着在每个时间步骤中，输入句子中的所有单词对翻译的贡献是相同的。这种处理句子翻译的任务效率很低。当编码器模型与注意力模型一起利用时，翻译将更多地集中在不同时间步骤中与当前翻译过程更相关的词上。由于 MBLSTM 网络学习的是高维序列特征，单个特征和时间-序列对最终识别结果的贡献可能不同。本方案尝试使用注意力模型来智能地学习特征的不同行动的效果，并根据其重要性分配权重。

在识别系统中，没有可用的先验信息用于训练。因此，注意力模型，也被称为自我注意力，将利用 MBLSTM 学到的顺序特征作为输入。这里用一个简单的例子来说明这个注意力模型。给出 n 个特征向量 $h_i, i = 1, 2, \dots, n$ ，可以从特征学习网络获得。本方案建立一个分数函数 $\Phi(\cdot)$ ，通过计算分数 s_i 来评估每个特征向量的重要性，如下所示。

$$s_i = \Phi(W^T h_i + b) \quad (3-2)$$

其中 W^T 和 b 分别为权重向量和偏置。神经网络中的任何激活函数，如 tanh、relu 或 linear，都可以用来建立得分函数。我们可以利用 softmax 函数对每个特征向量的得分进行归一化处理，写成：

$$a_i = softmax(s_i) = \frac{exp(s_i)}{\sum_i(s_i)} \quad (3-3)$$

注意力模型的最终输出特征 O 是向量和其归一化分数的乘积，如下所示。

$$O = \sum_{i=1}^n a_i * h_i \quad (3-4)$$

3.2.3 网络训练方法

为了确定模型参数，提出的 MBLSTM 框架使用带有真实标签的 CSI 数据进行训练。首先，所有的参数都是随机给出的。然后，CSI 数据被送入 MBLSTM，MBLSTM 用它来预测标签。本方案采用类别交叉熵误差作为度量标准，使用基于梯度的优化方法进行反向传播，并根据给定的真实标签调整模型参数。为了提高优化效率，本方案采用 ADAM 算法^[44] 计算每个参数的自适应学习率。

具体来说，假设 θ_t 是待优化的参数， g_t 是相应的梯度，使用 ADAM 更新 θ_{t+1} 的结果如公式3-5所示：

$$\begin{aligned}\alpha_t &= r_1 \alpha_{t-1} + (1 - r_1) g_t \\ \beta_t &= r_2 \beta_{t-1} + (1 - r_2) g_t^2 \\ \alpha_t &= \frac{\alpha_t}{(1 - r_1)} \\ \beta_t &= \frac{\beta_t}{(1 - r_2)} \\ \theta_{t+1} &= \theta_t + \frac{\eta}{\sqrt{\beta_t} + \epsilon} \alpha_t\end{aligned}\tag{3-5}$$

其中 α_t 和 β_t 分别是梯度的一阶和二阶矩， η 是学习率，设置为 1×10^{-4} ，参数 r_1 、 r_2 和 ϵ 分别设置为 0.9、0.999 和 1×10^{-8} 。

3.3 实验设置

本方案将所提出的方法与其他几种基于 CSI 的人体动作识别算法进行了比较，以评估其有效性。根据 Yousefi 等人^[45]的研究，随机森林（Random Forest, RF）模型在基于 WiFi 的人类活动识别中优于支持向量机、Logistic 回归（logistic Regression, LR）和决策树。在 Wang 等人^[46]的研究者中，还发现隐马尔可夫模型（Hidden Markov Model, HMM）对识别人类活动很有用。因此，将本方案的方法与这两种手工方法进行了比较。人工特征提取在^[45] 中有详细描述。本方案还将其与其他基于深度学习的方法进行了比较，这些方法可以自动学习特征，例如稀疏自编码器（Sparse Autoencoder, SAE）^[46, 47] 和传统的 LSTM^[45]。SAE 算法是一种无监督算法，它可以自动从未标记的数据中学习特征，并且可以生成比原始数据更好的特征描述。训练示例中的验证集用于微调所有方法的参数。对于评估，使用了 10 倍交叉验证。本方案将所有数据随机分为 10 倍。然后，本方案选择一倍的数据进行测试，其余的数据进行训练，最后得到 10 倍。所有 10 次运行的平均值决定了最终的识别精度。

用于比较的数据集由作者^[45] 在一个办公室收集。一台路由器被用作发射器，一台带有因特尔 5300 网卡的笔记本电脑被用作接收器。采样频率为 1kHz，具有三个天线和 30 个子载波，原始 CSI 数据的大小为 90。用于数据分割的窗口大小为 2 秒的滑动窗口。在视线条件下，发射器和接收器之间相隔三米。在数据收集期间，每个人执行每个活动 20 秒。值得注意的是，每个人在活动开始和结束时保持静止。六个人参与了数据收集过程，其中包括六项正常的日常活动：躺下、摔倒、跑步、坐下、站起和行走。每个志愿者进行 20 轮每项活动，得到的数据集大小约为 17GB。所有实验都在实验室的工作站上进行，使用 Python 运行代码。该工作站配备了 8 核 16 线程 Intel i9-9900 CPU 和 NVIDIA

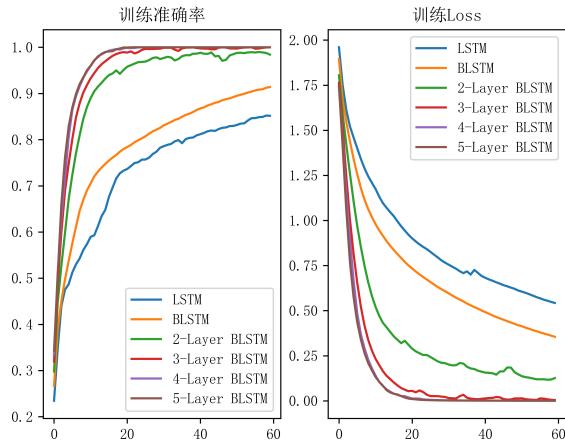


图 3.3 不同网络训练的准确率趋势

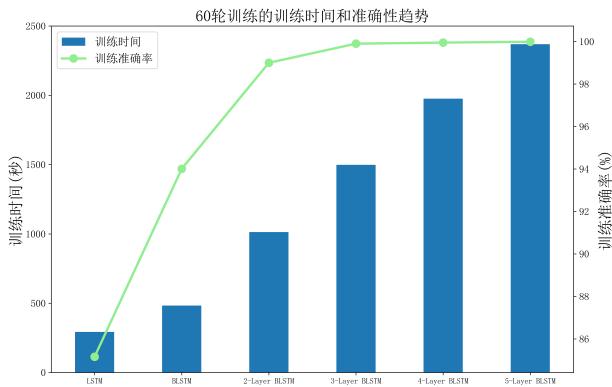


图 3.4 在 60 轮训练中训练时间和精确度的趋势

GeForce RTX 2080 GPU。

本方案比较了训练数据集中不同层数的 BLSTM 网络的准确性和损失的趋势。如图3.3显示，LSTM 和 BLSTM 网络收敛得比较慢，在第 60 轮训练时，准确率几乎没有达到 90%。另一方面，多层次 BLSTM 网络的收敛速度很快，在大约 10 轮训练时准确率超过 90%，在接近 20 轮训练时接近 100%，并在随后的训练中保持了准确率的稳定性。

虽然随着 BLSTM 层数的增加，训练收敛得更快，但并不是层数越多越好。随着层数的增加，网络结构变得越来越大，这意味着将使用越来越多的计算资源，在训练中也将消耗更多的时间。如图3.4所示，本方案用 200 个隐藏节点进行了实验。结果显示，网络结构越复杂，训练的时间就会明显增加。很明显，在训练轮数相同的情况下，BLSTM 网络增加到 3 层后，整体的训练精度没有太大提高，接近 100%，说明已经接近极限了。然而，每个网络所花费的训练时间相差很大。考虑到各种因素，本方案选择 3 层 BLSTM 网络作为本实验的网络模型，以便在保证高准确率的前提下尽量减少计算机资源的消耗。

表 3.1 不同数量隐藏节点的训练和验证时间

| 节点 时间 | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 |
|----------|--------|--------|--------|--------|--------|---------|---------|---------|
| 训练 (s) | 245.85 | 349.14 | 545.11 | 785.69 | 880.83 | 1043.76 | 1238.80 | 1444.24 |
| 验证 (s) | 7.63 | 7.46 | 7.04 | 7.64 | 7.21 | 7.24 | 7.86 | 8.35 |

表 3.2 不同方法的训练和测试时间

| 时间 | RF | HMM | SAE | LSTM | MBLSTM |
|--------|-------|-------|--------|--------|---------|
| 训练 (s) | 5.31 | 0.024 | 158.16 | 493.23 | 1551.14 |
| 测试 (s) | 0.008 | 0.17 | 0.19 | 3.54 | 8.72 |

本方案发现，LSTM 隐藏节点的数量对实验结果有很大影响。因此，本方案进行了第二个实验，看看这个参数对活动识别的准确性有何影响，实验结果如图3.5所示。当使用 50 个隐藏节点时，动作的识别准确率很低，尤其是“坐下”和“站起”这两种活动，本方案猜测它们太相似了。当隐藏节点的数量增加时，每个活动的识别性能都得到了提高，在数量达到 300 个后，准确率趋于稳定。如在表3.1，本方案使用 3 层 BLSTM 网络，在同样的 30 轮训练中，隐藏节点越多，训练时间越长，本方案在 MBLSTM 中选择使用 200 个隐藏节点。

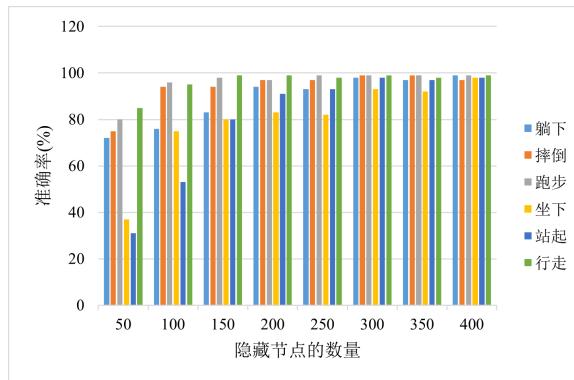


图 3.5 不同数量的隐藏节点对每个活动的识别精度

基于深度学习的方法的时间复杂性是一个常见的问题。本方案使用同一数据集比较了一些方法的训练时间和测试时间。表3.2显示了所有方法的训练时间和测试时间。可以清楚地看到，使用深度学习方法的算法比典型的机器学习算法的训练时间长得多。在所有使用深度学习的方法中，提出的 MBLSTM 消耗的训练时间最长。所有的方法都有很短的测试时间。例如，提出的 MBLSTM 对 420 个测试样本的测试时间为 8.72 秒。这标志着每个样本将被测试 0.0208 秒。每种情况下，数据分割的窗口大小为 4 秒。可以看出，本方案提出的基于 WiFi CSI 的 MBLSTM 方法可用于实时人员活动识别。

3.4 实验结果

图3.6显示了所有基准和本方案提出的 MBLSTM 方法在数据集上的混淆矩阵。需要人工提取特征的活动识别算法，如 RF 和 HMM，表现最差。HMM 算法的表现明显好于 RF 算法。与 RF 和 HMM 人工特征提取不同，使用深度学习方法的 SAE 算法有更好的表现。这证明了使用 SAE 方法进行自动特征学习的有效性。LSTM 网络的表现优于 SAE 方法，因为它将 CSI 序列中的时间因素纳入了特征学习。由于在本方案提出的方法中加入了注意力模型和多层双向 LSTM 的结构，本方案的 MBLSTM 方法在识别六种日常活动时取得了优异的识别效果。六种日常活动的识别准确率都在 96% 以上，这对大多数识别情况来说是足够的。

识别的准确性因活动的不同而有很大差异。较高的体力活动，如“摔倒”、“行走”和“跑步”显示出更高的识别性能。这是由于这些活动对收集的 CSI 数据的特征有很大的影响。同样明显的是，大多数方法对“坐下”这一活动的识别精度相对较低。这可能是因为这一活动对 CSI 特征的影响与“躺下”和“站起”活动相同。值得注意的是，RF 方法对手工制作的特征的识别准确率远远低于 50%。“摔倒”活动是这六种活动中最重要的，特别是对老年人来说^[48]。所提出的 MBLSTM 方法能够以 99% 的准确率识别“摔倒”动作，这将在各种医疗应用中发挥作用。基于深度学习的方法的训练期延长是其缺点之一。然而，这个耗时的程序只需要完成一次。值得注意的是，基于深度学习的方法可以快速在线测试，足以满足大多数实时应用。

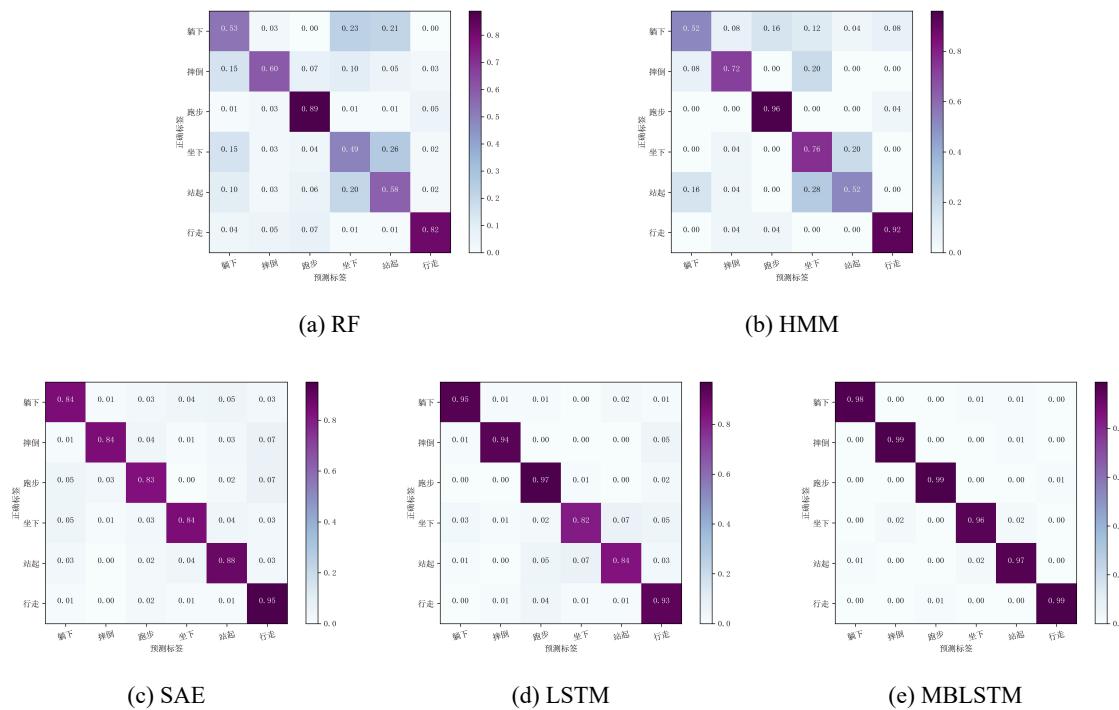


图 3.6 所有方法和提出的 MBLSTM 方法在数据集上的混淆矩阵

3.5 本章小结

在本章中，使用传感器来收集 WiFi 时间序列信息，并通过改进传统的 LSTM 模型，来提取用于人类活动识别的 WiFi 信号特征信息。在两个方向上，BLSTM 网络可以从原始的 WiFi CSI 数据中学习重要的序列特征。多层 BLSTM 网络可以通过加速训练期间的收敛来提高准确性。本章节在实际环境中评估了该方法，并将其与各种基准方法进行了比较，如随机森林、隐马尔可夫模型、稀疏自动编码器和传统的 LSTM。所提出的基于 MBLSTM 的人体动作识别方法在实验中表现出更高的性能。

第四章 基于 WiFi 的人体动作识别可视化应用

传统的人体动作识别系统通常能够对不同的动作进行准确的识别，但其输出结果通常仅是对动作进行分类的结果，难以在视觉上直观展示具体的动作过程。为了解决这一问题，本章节提出了一种创新性的方法，即将人体动作识别和数字孪生系统相结合。这种基于数字孪生系统的可视化方法可以直观地展示人体动作的实际过程，使得对动作的理解更加深入和全面。

基于 WiFi 的人体动作识别需要大量的数据集进行训练，然而公开的数据集数量较少，因此本章节还提出了数据集制作的方法，具体包括数据的采集工具、采集方法和数据处理方式。通过这些步骤，本文得到了一个高质量、具有代表性的数据集，可用于深度学习模型的训练和评估。

4.1 数字孪生概述

制造业数字化的快速发展使数字孪生（Digital Twin, DT）成为可能，它使生产系统更具有创造性。它根据从传感器获得的大数据，建立了一个物理系统的数字副本。这个数字复制品并不是简单地复制物理系统中的对象，它还模拟了这些对象的实时状态，因此它可以像其物理对应物一样实时运行，这就是它的名字的由来^[49]。因此，在制造业中，制造过程的开发和优化可以在构建的数字孪生中进行虚拟研究。这可以大大减少试错实验的数量，最大限度地减少开发时间和成本。在制造、连接和装配的过程优化中已经有了基于数字孪生的方法^[50]。

一个典型的数字孪生模型一般被认为由 5 个部分组成：物理实体空间、虚拟实体空间、孪生数据、连接和服务，如图4.1所示。

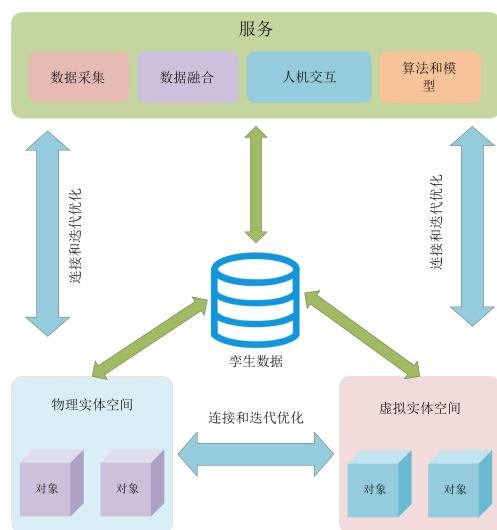


图 4.1 典型的数字孪生模型

物理实体空间代表现实生活中的物体。虚拟实体空间代表数字空间中对现实世界的模拟。孪生数据可以具体表示为物理数据、数字空间数据和服务相关数据。连接表示不同组件之间连接和访问的接口和数据类型的统一标准。服务指的是具体应用中针对不同需求的各种功能。具体来说，它主要包括数据采集、数据融合、人机交互和优化等各种功能。

设计对象的客观功能和用户需求是数字孪生设计的基础。需求经过讨论、分析和设计处理后，转化为实用的文字或图形表示，以帮助生产阶段的规划。通过产品的数字孪生体，使物理和虚拟产品之间的数据共享的双向互动联系成为可能，而产品的数字孪生体则为其提供动力。此外，它提高了两者之间的协同作用，以不断研究前沿的、特殊的和有益的设计方案，以改进产品。因此，它促进了动态创新，超越了客户的期望。

数字孪生体通过允许将真实产品的行为映射到同等的虚拟产品上，增强了设计的准确性。它不仅仅是一个用于产品展示的数字模型。此外，它还能对产品在真实环境中的性能进行建模。在整个生命周期中，数字孪生强调虚拟与物理的结合。在对开发周期进行有力描述的同时，它在产品、生产工具、环境和流程之间建立联系，并将产品生命周期各阶段的数据纳入设计阶段。通过其不同的功能使决策成为可能，并以各种方式提高设计质量和效率，数字孪生可以因此改善和优化商品及其制造系统。

基于数字孪生的监测方法已被应用于多种用途的场景中。Yu 等人^[51]提出了一种基于数字孪生的系统检测方法，使用非参数贝叶斯网络，以使智能制造的系统更加可靠。Leng 等人^[52]提出了一个使用数字孪生的制造网络的物理系统框架。该框架对系统的检测是通过一个在线实时控制模型实现的。Liu 等人^[53]建立了一个用于复杂产品质量检测的数据驱动的数字孪生框架，他们使用深度学习网络来检测产品的微小缺陷。到目前为止，一些研究表明，基于数字孪生的检测方法可以减少产品的开发时间。Perez 等人^[54]提出了一种基于数字孪生的方法，用于设计多个机器人制造单元，可以对制造单元进行监控和模拟。

支持医疗行业深度学习的数字孪生大多是人类数字孪生^[55-57]。很难模拟所有的人类活动，所以人类数字双胞胎只能在某些生物层面上应用。具体来说，Barricelli 等人^[58]使用一个数字孪生系统来监测运动员的健康状况。他们利用可穿戴传感器收集运动员的活动数据，建立运动员的孪生体，并通过算法分析孪生体来预测运动员的健康状况并给出合理的建议。此外，Chakshu 等人^[59]模拟了患者的头部行为来检测颈动脉狭窄的严重程度。他们用主成分分析法对数字模型的数据进行分析，预测患者病情的验证程度。作者提出，可以利用深度学习强大的特征提取能力来提高预测的准确性。

目前，有不同的方法来建立数字孪生，但基本上可以概括为基于物理的建模和基

于数据的建模。Chakshu 等人^[60] 研究了构建心血管系统数字孪生模型的数值建模方法。Dong 等人^[61] 提出了一个 5G 通信网络的数字孪生模型，该模型由实际数据和理论数据组成，以训练深度学习算法。在机器人领域，大多数相关研究都集中在机器人组装上。Malik 等人^[62] 开发了一个用于人机交互的数字孪生系统，其中虚拟空间由 CAD 构建。Kaigom 等人^[63] 构建机器人数字孪生，它反映了用于虚拟测试的物理机器人的特性，并主要对机器人的状态进行建模。Perez 等人^[54] 开发了一个基于虚拟现实的多机器人数字孪生建模框架，该框架允许数字孪生可视化。作为最强大的可视化游戏引擎之一，许多数字孪生使用 Unity3D 进行建模^[64]。对文献的搜索显示，其中很少有基于数字孪生的人类活动识别系统的研究。因此，本章节为人体动作识别研究基于深度学习的数字孪生，希望促进数字孪生在医疗保健中的应用。

4.2 人体动作识别数据集收集

深度学习是近年来的研究热点，由于其从大量信息中提取引人注目的特征的强大能力，被研究人员广泛引用。深度学习依靠众多的数据集进行学习。然而收集和标记无线通信数据需要大量的时间和成本，而且很少有数据集是公开可用的。这就限制了无线通信方向的深度学习研究。因此，建立公开可用的无线通信数据集是非常必要和迫切的。丰富的活动数据可以帮助研究人员更好地研究人体动作识别，可以提高人体动作识别的准确性。除了清华大学分享的基于 WiFi 的 Widar 手势识别数据集^[65]，没有大型的公开可用的基于 WiFi 的人体动作识别数据集。只有一些文章作者分享他们自己收集的 WiFi 数据集^[66]，他们不仅有较少的数据，而且相对较旧。因此，创建一个基于 WiFi 的人体动作识别的公共数据集迫在眉睫。它可以与其他研究人员共享，以探索 WiFi 信号和活动之间的关系，提高人体动作识别的性能。创建基于 WiFi 的人体动作识别数据集可以减少研究人员的时间成本，促进基于 WiFi 的无线感应的发展和应用。

创建一个基于 WiFi 的人体动作识别公共数据集需要考虑许多因素。这些因素包括发射器和接收器之间的位置、信号收集环境、活动幅度以及有多少种活动类型。首先，信号发射器和接收器的位置决定了收集信号的质量。如果距离太远，就不能从收集的信号中提取有用的活动特征，影响活动识别的准确性。其次，环境对信号收集也有很大影响。因为除了人体反射 WiFi 信号外，环境中的其他物体也会反射 WiFi 信号。为了减少其他反射信号对所含活动信号的干扰，选择一个合适的信号收集环境以保证收集到的信号是很重要的。本章节邀请了六名志愿者，在一个会议室和一个办公室收集六种活动的 WiFi 信号。为了探讨活动的幅度，收集的六个动作可以分成上肢、下肢和全身活动。

4.2.1 原始 WiFi 数据收集

A. 实验装置

商业路由器可以提取像 RSSI 这样的粗粒度信息，但 CSI 数据收集需要特定的无线卡和软件。一般来说，大多数研究人员使用一个公开的工具，称为 CSI Tool^[67] 来收集 WiFi 数据。它可以帮助研究人员收集从商用路由器传输到另一台装有英特尔 5300 网卡的计算机的 WiFi 信号。图4.2显示了本方案在实验中使用的路由器和无线网卡。实验将无线网络设置为 2.4GHz 频段，频道号为 10，频道带宽为 20MHz。此外，所使用的接入点和网卡符合 IEEE 802.11n 标准，该标准支持多输入多输出和正交频分复用调制方案，允许信息通过多个天线发送和接收。值得注意的是，CSI Tool 需要在 Ubuntu 系统上运行。此外，CSI Tool 对 Ubuntu 版本有严格限制，要求它基于 3.2 和 4.2 版本的内核，即 Ubuntu 12.04 和 Ubuntu 14.04.4 之间。在本方案的实验中，使用了 Ubuntu 14.04。CSI 工具安装的官方文档可以在<http://dhalperi.github.io/linux-80211n-csitool/installation.html> 找到。

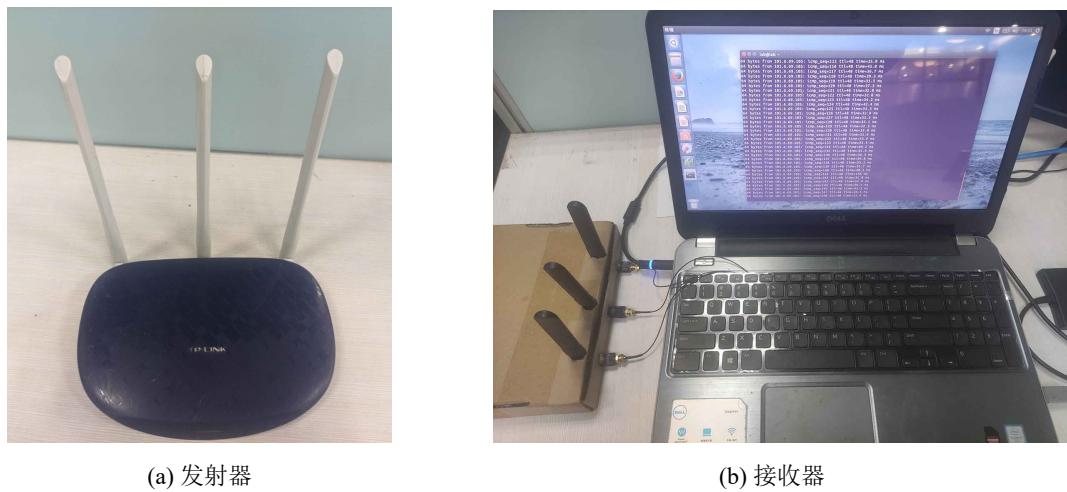


图 4.2 实验设备

B. 实验环境

本方案在两个房间里收集含有活动的 WiFi 信号，以生成人体动作识别数据集。图4.3显示了收集信号的房间的平面图。室内环境的细节如下。

- 会议室:** 大小是 $4 \times 8 \text{ m}^2$ ，包含一张多人会议桌和几把椅子。将发射器和接收器相隔 3 米，让志愿者在他们的中点进行活动。这种设置是为了确保收集到的信号的质量。此外，志愿者们离会议桌有 1 米的距离。
- 办公室:** 大小是 $3 \times 5 \text{ m}^2$ ，包含书柜、沙发、书桌和椅子。接收器和发射器相距 3 米，志愿者在其中点进行活动。

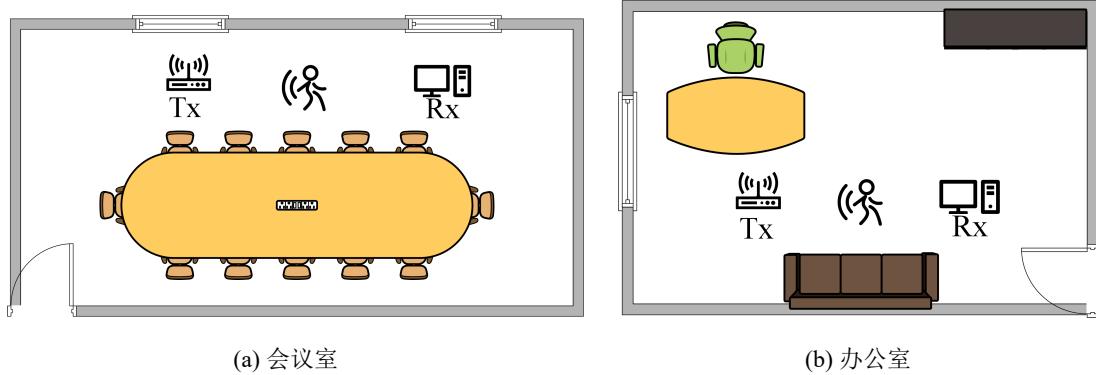


图 4.3 实验环境

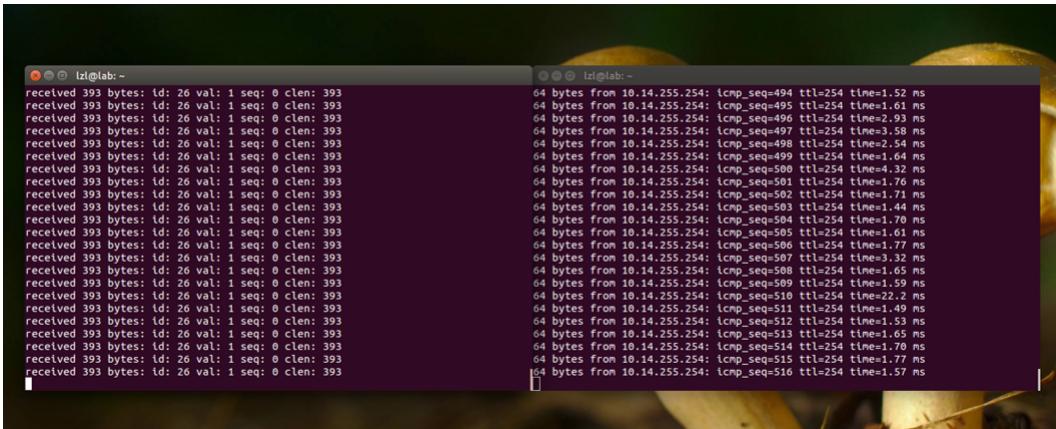


图 4.4 使用 CSI Tool 工具采集数据界面

C. 实验步骤

首先，在需要收集 WiFi 数据的室内设置实验设备，并按照预先规定的规则放置发射器（即路由器）和接收器（即带有英特尔 5300 网卡的笔记本电脑）。然后用接收器连接到由发射器创建的无线网络。需要特别注意的是，要求路由器创建没有密码的无线网络，因为 CSI Tool 的固件没有足够的代码空间来存储波束成形软件路径和加密软件路径（用于 WEP/WPA/WPA2/等网络）。

其次，在笔记本电脑的终端上 ping 发射器创建的 WiFi 的 IP 地址。本方案使用可视化工具，看看接收器是否能成功收集来自发射器的 WiFi 信号。检查没有问题后，开始进行 WiFi 信号收集实验。志愿者们站在发射器和接收器之间，按照预先安排的顺序，依次进行需要收集的活动。运行 CSI Tool 提供的图4.4程序，然后指示志愿者开始活动，活动结束后结束程序，完成活动的收集。为了信号的稳定和减少噪声的影响，在收集信号时要有大约 1 秒的空数据，然后志愿者做 2-3 秒的活动，最后以大约 1 秒的空数据结束收集，整个过程大约是 5-6 秒。用同样的过程来完成其余活动的收集任务，将收集到的 .dat 文件按规定的形式命名。最后，对所有收集到的文件进行排序和分类。图4.5显示了活动数据收集的实验。志愿者站在发射机和接收机的中间，距离两侧约 1.5 米。



图 4.5 活动数据收集

4.2.2 数据预处理

在本节中，本方案首先分析了收集的数据。然后介绍了数据预处理的方法，并使用几种分类算法测试了数据集的性能。在本方案中，实验数据分析和活动分类程序是在实验室计算机上运行的。具体配置是英特尔 i9-9900 CPU，16GB 内存，以及 Nvidia RTX 2080 显卡。使用的编程语言是 Python 和 Matlab。

A. 数据解析及处理

CSI Tool 提供 Matlab 程序解析原始 WiFi 数据，可以使用函数 `read_bf_file` 来解析文件中的原始数据，从而获得一些具有相应值的字段。表4.1详细描述了这些字段的含义。

数据集包含影响 WiFi 的人类活动的原始 WiFi 数据。可以从这些原始数据中获得 RSSI 和 CSI，并且可以从 CSI 中提取细粒度信息以对不同的活动进行分类。发射机每秒发送 30 个包，每个包包含 RSSI 的单个值。RSSI 值表示为一维数据序列 $R = \{RSSI_1, RSSI_2, \dots, RSSI_i, \dots, RSSI_n\}$ ，其中 $RSSI_i$ 表示第 i 个 RSSI 值， n 是序列的长度。如表4.1中所述，CSI 是一个大小为 $N_{tx} \times N_{rx} \times N_{sc}$ 的三维矩阵，其中 N_{tx} 表示发射机的天线数， N_{rx} 代表接收机的天线数； N_{sc} 表示子载波数。在本方案的实验中，发射机天线是 3，接收机天线是 3。因此，CSI 的值可以表示为 $CSI_i = \{3 \times 3 \times 30\}$ 。CSI 样本可以表示为 $C = \{CSI_1, CSI_2, \dots, CSI_i, \dots, CSI_n\}$ 。此外，研究人员可以将数据集中的数据用于研究的其他方面。

表 4.1 数据字段的描述

| 字段 | 字段含义 |
|---------------|---|
| timestamp_low | 时间戳，相连两包此值差单位为微秒。 |
| bfee_count | 驱动记录并发送到用户控件的波束测量值的总数。内核和用户空间中 netlink 频道是有损的，可以用该变量来检测被丢弃的测量值。 |
| Nrx | 接收端使用的天线数量。 |
| Ntx | 发射端使用的天线数量。 |
| rssi_a | 在接收端的第一根天线上收到的 RSSI 值。 |
| rssi_b | 在接收端的第二根天线上收到的 RSSI 值。 |
| rssi_c | 在接收端的第三根天线上收到的 RSSI 值。 |
| noise | 在通道上测量的噪声。 |
| agc | 代表无线网卡的自动增益控制参数，单位为 dB。 |
| perm | 演示无线网卡如何将 3 根接收天线的信号与 3 条射频链对齐 |
| rate | 工作频率。 |
| csi | 信道状态信息（接收天线数量 × 发射天线数量 × 子载波数量）。 |

商用 WiFi 设备中的硬件缺陷、环境中的电磁干扰和不可避免的白噪声导致原始 CSI 数据具有许多噪声成分，这会严重影响执行活动特征提取的能力。因此，在下一步中，必须对数据进行降噪。许多降噪方法使用各种滤波器组合来提取本方案想要的时域信息或频域分量。首先应该去除最明显的异常值，在本方案中，使用了最常用的 Hampel 异常值过滤器。可以直接使用 Matlab 的直接自含 Hampel 数直接从原始数据矩阵中去噪。Hampel 滤波器的效果可以在图4.6中看到。

实验需要压缩数据包中的大量 CSI 数据，希望在 WiFi 数据中保留更多有关环境的信息。因此，仅选择某些载波信号进行判断是不可行的，因为不同的载波包含不同的环境信息特征。将所有载波信号叠加在一起以取其平均值也是不正确的，因为一些载波的叠加会削弱数据中包含的一些明显的环境信息。因此，从 WiFi 数据中提取大量有效信息是必要的。方差在数学中用于测量上述信息量。主成分分析（Principal Component Analysis, PCA）可以整合和压缩数据中的方差，显著减少数据量，同时保持大量信息而不会丢失。

此外，保存低频信号的低通滤波器（Low-Pass Filter, LPF）也经常用于降低数据噪声。LPF 是用于通过低频信号和衰减高频信号的滤波器。然而，阻塞和衰减幅度将根据不同的频率和滤波程序而变化。具体地，可以设置阈值以过滤期望的频率。由于巴特沃

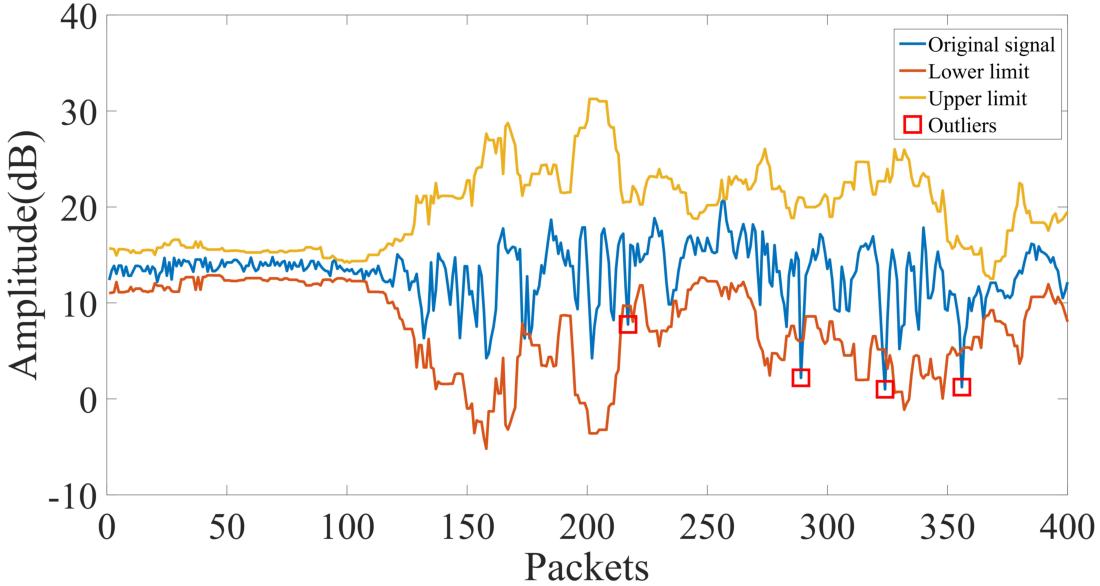


图 4.6 使用 Hampel 滤波器对原始数据进行去噪

斯滤波器 (Butterworth Filter) 在通带中具有最平坦的特性，并且在正频率范围内随着频率的增加而单调递减，因此 3dB 不变性和其他优良的财产使其成为数字信号处理中最常用的滤波器之一。所以，本方案在降噪时也考虑了巴特沃斯低通滤波器。

本方案利用 Matlab 提取收集到活动的 CSI 信息，其中两个活动的 30 个子载波如图4.7所示。鼓掌的活动是周期性的来回拍手，从图4.7a中能看出 WiFi 信号出现周期性的波动。下蹲的活动是一个突然的活动，所以从图4.7f中可以看出，在下蹲前和下蹲后 WiFi 的信号是平稳的，在下蹲活动出现的时候，WiFi 信号出现明显波动。利用算法监测不同活动的 CSI 变化，从而可以实现不同人体活动的识别。

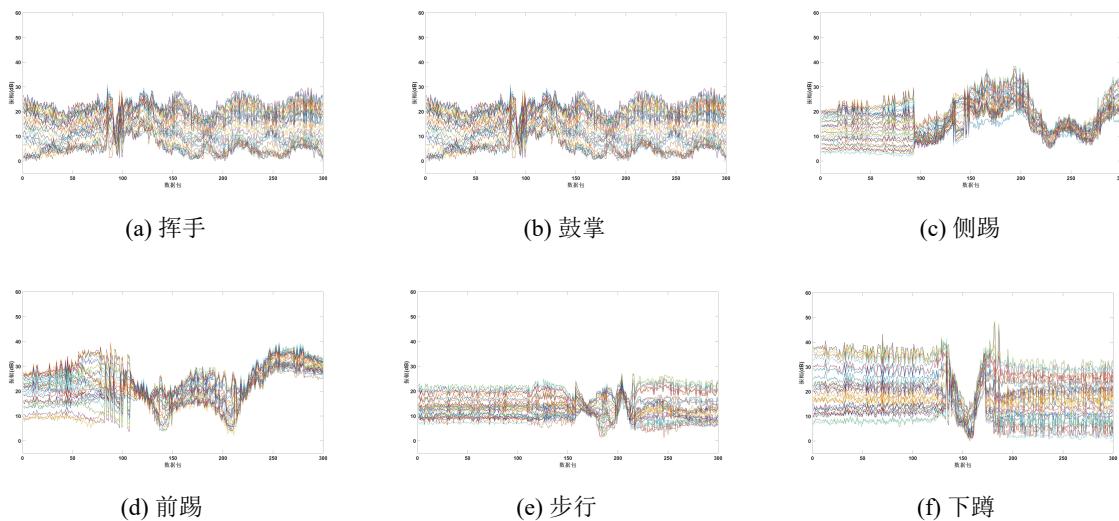


图 4.7 收集到的活动 30 个子载波图

表 4.2 在数据集上不同分类算法动作识别的准确率

| | KNN | DT | SVM | CNN | LSTM | BLSTM |
|----|--------|--------|--------|--------|--------|--------|
| 训练 | 86.37% | 84.25% | 85.13% | 92.52% | 93.85% | 95.78% |
| 测试 | 85.56% | 83.78% | 84.44% | 91.65% | 93.21% | 95.35% |

B. 数据集文件命名格式

本方案还使用 Matlab，将原始数据中的 CSI 数值提取出来，生成 CSV 文件。然后将原始数据和提取生成的 CSV 数据打包分类保存，具体的格式如下。数据集包含两个文件夹，一个命名为 RAW data，另一个名为 CSV data。每个文件夹包含从十名志愿者收集的活动信息。RAW 数据文件夹包含收集的原始 WiFi 数据，其中包含具有命名约定 $ax_ty.dat$ 的活动信息，其中 ax 表示活动类型 x ， ty 代表第 y 次收集此活动，.dat 是文件格式。具体而言，收集到的活动类型包括挥手（1）、鼓掌（2）、侧踢（3）、前踢（4）、步行（5）和下蹲（6）。例如， $a4_t2.dat$ 文件表示为挥动活动第二次收集的原始数据。CSV 数据文件夹包含从原始 WiFi 信号种类中提取的信道状态信息（CSI），命名约定为 $ax_ty.csv$ ，其中 ax 表示活动类型 x 、 ty 代表收集此活动的第一个 y 时间，.csv 是文件格式。

C. 数据集性能评估

为了评估数据集在人体动作识别任务中的具体性能，本方案使用 KNN、DT、SVM、CNN、LSTM 和 BLSTM 算法进行了实验。因为这些方法在各个领域都得到了广泛的应用，所以理论已经非常成熟，我们可以很容易地使用它们。表4.2显示了数据集上活动识别的不同分类算法的准确性。本方案将数据集的一部分用于训练，其余部分用于测试。由于分类算法在每个类别的活动识别精度上有所不同，本方案使用了六个活动的平均精度进行比较。该表显示了所有六种算法的高识别精度。前三种算法是传统的机器学习算法，它们的准确率约为 85%。最后三种是更流行的深度学习算法，其准确率超过 90%。由于 BLSTM 可以学习正向和反向特征，其精度略高于 LSTM。本方案使用了更简单的网络模型，如果进一步的工作可以优化这些网络，准确率可能会更高。通过六种分类算法的实验，可以表明本方案的数据集在基于 WiFi 的人体动作识别任务中是有效的。

4.3 系统框架设计

提出的系统如图所示，可以看出它由三层组成，即数据层、计算层和服务层。可以清楚地看到，这三层之间存在着一步步的联系。该系统被分为若干层，每一层都扩展和加强了前面各层的功能。下面是对这三层功能的描述。

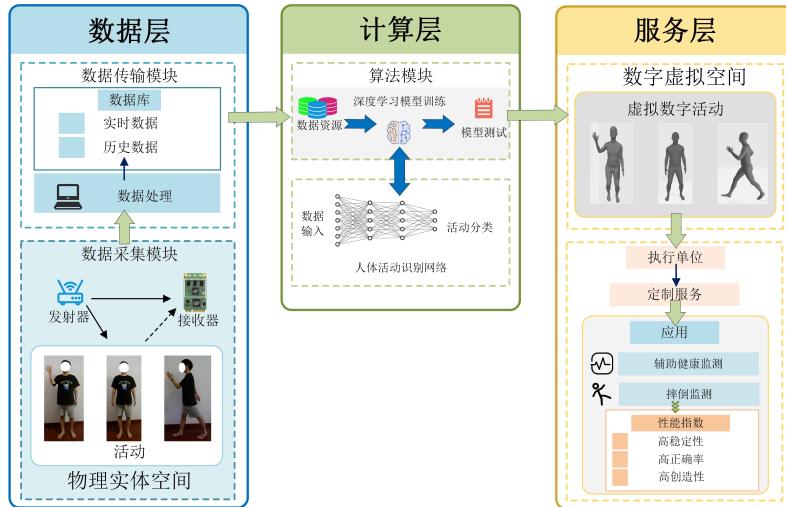


图 4.8 基于数字孪生系统的人类活动识别框架。

- **数据层:** 数字孪生技术从数据层开始，它提供基础和底层支持。其中有两个模块：一个用于数据采集，一个用于数据传输。数据采集模块提供整个数字孪生系统运行所需的数据。它包括实时状态传感数据、历史运行数据和用户数据。数据采集设备从物理空间获取数据。高性能的传感器和分布式传感技术为采集的物理数据的实时性提供保证。历史信息包括用户在特定时间的所有活动信息和危险活动发生的时间。收集到的历史数据被归档和存储，以便在需要时可以查看。所有这些数据通过数据传输模块传输到上层系统，以支持其运行。

- **计算层:** 要实现数字孪生系统的各种复杂功能，离不开强大的计算能力。计算层包括一个算法模块，采用数据驱动的方式，在获得数据层提供的数据后，通过人类活动识别网络对活动进行分类，然后返回分类结果。该层采用深度学习等智能算法，建立数据挖掘和知识学习模型。因此，它可以为数字孪生的构建提供多角度的决策和建议。

- **应用层:** 服务层是一个能够提供定制服务的平台，可以满足不同类型用户的要求。例如，本方案可以为医疗物联网提供健康监测功能，为智能家庭提供老人跌倒检测功能。该服务满足了几个性能指标，包括高可靠性、准确性和高价值的创造力。

4.3.1 传感模块

在4.2.1中已经介绍，本方案的传感模块使用的设备，仍然是一台商用的路由器和一台改造后装有英特尔 5300 网卡的笔记本电脑。数据收集方法和数据处理方式可以在4.2.1中看到。

4.3.2 元学习框架

目前，深度学习使用模型来学习特征，学到的特征和模型构建了一个端到端的训练方法，但这种方法并没有模仿人类的学习能力。本方案希望有一种智能的学习方法，能

够从已知的任务中学习，用有限的标签适应新的环境，简单地说，就是像人类那样学习。近年来，元学习（meta-learning）^[68]的引入为实现上述目标提供了新的可能性。元学习通常被理解为“学习如何学习”，其主要目的是在多个学习任务中改进学习算法的过程。相比之下，传统的机器学习则是在多个数据实例上改进模型预测。在元学习中，底层学习过程由内部学习算法处理，其目的是解决由数据集和目标定义的任务。 D_l 代表第 l 个环境的数据集。每个数据集被分为两部分。一部分用于学习，标记为 D_l^S ，另一部分用于训练和测试，标记为 D_l^Q 。一般来说，基础网络 f_Φ 预测 y 种分类的概率可以表示为 $P_\Theta(y|x, D^S)$ 。其中 D^S 数据集中的 x 被用作输入， Θ 是元参数。

$$\Phi^* = \arg \max_{\Phi} \mathbb{E}_{D^S} \left[\sum_{(x,y) \in D^S} P_\Theta(y|x, D^S) \right] \quad (4-1)$$

然后，通过对源环境中每个查询集的预期值的最大化，找到最佳的元参数 Θ^* 。

$$\Theta^* = \arg \max_{\Theta} \mathbb{E}_{D^Q} [P(\Phi|D^Q)] \quad (4-2)$$

Θ^* 被转移到一个包含 \hat{D} 数据集的新环境。数据集的一部分，即包含 D^S ，有最小的标记数据。数据集中的其余数据是没有标签的，用于最终测试。在这种环境下， \hat{D}^S 被用来完善基础网络。应用于感知环境的两种主要元学习方法在 $P_\Phi(y|x, D^S)$ 的建模方式上有所不同。

对于 WiFi 信号，本方案提出了一个易于训练的基于度量的元学习框架。构成本方案所提出的元框架的两个模块如图4.9a所示。

- **WiFi 度量模块：**在线性映射层的帮助下，它试图训练一个有效的距离度量来评估数据的相似程度。它同时考虑由 Bi-Path 基础网络 f_Φ 产生的各种特征。因此，该模块提供了对 x 的广泛解释，以帮助更正确地进行概括。

- **分类模块：**它的目的是将 f_Φ 的能力进一步整合到识别网络中，而不是将其作为 WiFi 度量模块中的特征提取器。该模块允许基础网络帮助训练距离度量。

A. WiFi 度量模块

WiFi 度量模块首先使用本方案的 Bi-Path 基本人体动作识别网络 $f_\Phi : \mathbb{R}^{D_x} \rightarrow \mathbb{R}^{D_z}$ 提取功能。WiFi 信号矩阵 x 被映射到特征表示 z ，其中 D_x 和 D_z 分别是 WiFi 信号矩阵和相应嵌入的维度。与之前的工作相比，例如 [69, 70]，本方案希望多个特征度量空间可以收集更复杂的特征，而不是简单地依赖于单个特征。WiFi 信号矩阵在时域和频域中都有描述，因此，本方案使用 M 来表示 WiFi 信号矩阵 x ，表示为 $\{H_m^{feat} = f_\Phi(x)\}_{m=1}^M$ 。其中， H_m^{feat} 包括 H^{time} 和 H^{freq} ，分别从时域和频域提取，以及 H^{fuse} ，结合两个域的特征。考虑到这些不同的特征，本方案使用余弦距离 $d(a, b) = -\frac{a \cdot b}{\|a\| \|b\|}$ 来生成由以下

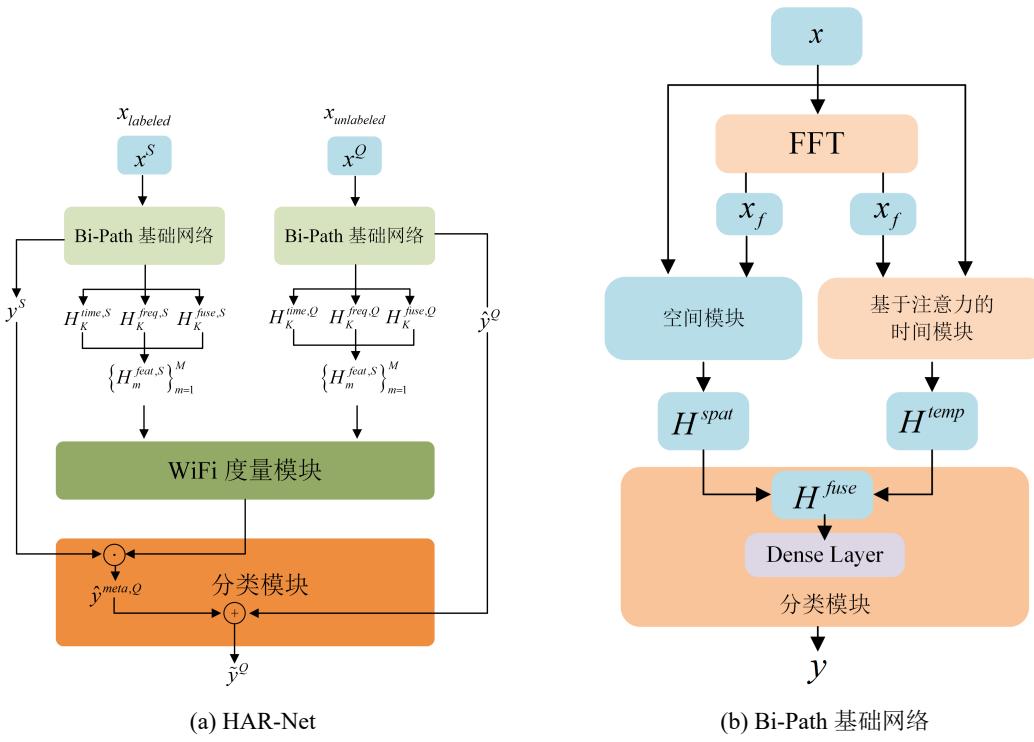


图 4.9 HAR-Net 的整体设计图

等式生成的特征之间的距离集 $\{\lambda_m\}_{m=1}^M$:

$$\{\lambda_m\}_{m=1}^M = \{d(H_m^{feat,S}, H_m^{feat,Q})\}_{m=1}^M \quad (4-3)$$

本方案将 WiFi 度量模块指定为可训练的，而不是简单地依赖余弦距离来进一步扩展 WiFi 度量空间。本方案通过堆叠这些 M 距离 $[\lambda_1, \lambda_2, \dots, \lambda_M]$ 来构建一个标记为 Λ 的向量形式。本方案将 λ 中的每个距离度量 λ_m 通过给定查询观察的线性映射权重 η 进行组合。因此，通过在以下等式中对支持观察 y^S 的标签进行加权，本方案可以计算查询观察的活动概率。

$$p_\beta(y=j|x) = softmax(-\Lambda_j \beta), \beta = \eta y^S \quad (4-4)$$

其中 $\eta \in \mathbb{R}^{M \times N_c}$ 和 $\Lambda_j \in \mathbb{R}^{1 \times M}$ 。为了进一步实现 β ，本方案通过分析证明了第 j -类的类宽交叉熵损失函数为：

$$L_j(\beta) = \sum_{x \in D^Q} [\Lambda_j \beta + \log(\sum_{n=1}^{N_c} e^{-\Lambda_n \beta})] \quad (4-5)$$

上述方程相对于 β 的二阶偏导如下所示：

$$\nabla^2 L_j(\beta) = \frac{1}{1^T \gamma} diag(\gamma) - \frac{1}{(1^T \gamma)^2} \gamma \gamma^T \quad (4-6)$$

其中 $\gamma = [\gamma_1, \dots, \gamma_{N_c}]$ 和 $\gamma_k = e^{-\Lambda_k \beta}$ 。如果 $\nabla^2 L_j(\beta) \geq 0$ ，则公式4-5中的 $L_j(\beta)$ 是凸的。因此，本方案利用下面的等式来确认 $v^T \nabla^2 L_j(\beta) v \geq 0$ 对于每一个 v 。

$$v^T \nabla^2 L_j(\beta) v = \frac{(\sum_{n=1}^{N_c} \gamma_n v_n^2)(\sum_{n=1}^{N_c} \gamma_n) - (\sum_{n=1}^{N_c} v_n \gamma_n)^2}{(\sum_{n=1}^{N_c} \gamma_n)^2} \quad (4-7)$$

由于 Cauchy-Schwarz 不等式，它实际上是非负的，即 $(\sum_{n=1}^{N_c} v_n \gamma_n)^2 \leq (\sum_{n=1}^{N_c} \gamma_n v_n^2)(\sum_{n=1}^{N_c} \gamma_n)$ 。现在 $L_j(\beta)$ 的凸性已经被证明，很明显，本方案的 WiFi 度量模块可以迅速找出理想的参数 β ，以结合各种表征。

B. 分类模块

在这个模块中，本方案试图将基础网络 f_Φ 也用于识别，而不仅仅是作为一个特征提取器，这与标准的基于度量的元学习框架不同。更准确地说，本方案首先计算 $\hat{y}^{meta,Q}$ ，即通过支持观测值 y^S 的标签加权对观测值进行分类，如公式4-4中所述。同时，本方案使用识别网络 f_Φ 直接计算对数 $\hat{y}^Q = f_\Phi(x^Q)$ 。之后，利用残差连接，本方案将 \hat{y}^Q 加入 $\hat{y}^{meta,Q}$ 中。结果 HAR-Net 的最终预测 \tilde{y}^Q 被计算为 $g_c(g_{metric}(f_\Phi(x^S), f_\Phi(x^Q)), y^S, \hat{y}^Q) = g_{metric}(f_\Phi(x^S), f_\Phi(x^Q))y^S + \hat{y}^Q$ 。距离度量的元学习可以通过基础网络来加强。

C. 训练过程

本方案设计了一种有效的算法来训练 HAR-Net。在训练和测试识别网络的过程中，本方案使用不同的数据集，表示为 D 和 \hat{D} 。为了提高识别网络的整体能力，本方案在训练阶段对 f_Φ 和 g_Θ 进行了内部训练和元训练的领导者-追随者不对称训练。首先，本方案使用 D^S 进行内部训练 f_Φ 。通过优化参数 Φ 和 Θ ，使用梯度下降法来减少内部训练损失 $\mathcal{L}(f_\Phi(x^S), y^S)$ 。通过这种方式，元学习参数可以找到一个粗略的初始化点，同时提高 HAR-Net 的收敛速度。之后，利用支持和查询观察，本方案对 g_Θ 进行元训练，生成 q_Ω ，从而改善查询分类。

本方案现在已经通过连续训练 f_Φ 和 g_Θ 实现了一个训练有素的 WiFi-HAR。考虑到测试数据集 \hat{D} ，本方案可以衡量 HAR-Net 在执行一次 WiFi-HAR 时的表现如何。首先，本方案的目标是通过使用标签支持观测 $\hat{D}^S = \{x_j^S, y_j^S\}_{j=1}^{N_c}$ 来改善 q_Ω 。特别是，本方案通过利用相同的内部训练方法修改训练好的 q_Ω ，但本方案只对元学习框架的参数进行微调。最后，本方案将支持观测值标记为 $\tilde{y}^Q = q_\Omega(x^S, x^Q, x^S)$ ，并使用改进的 q_Ω 对查询观测值进行分类。

4.3.3 Bi-Path 基础网络

如图4.9b所示，学习 WiFi 特征的三个主要模块构成了本方案的基本人体动作识别网络。

- **空间模块：**该模块从时域和频域的 WiFi 矩阵中提取复杂的空间特征。

- **基于注意力的时间模块:** 该模块旨在从时域和频域的 WiFi 矩阵中捕捉长期时间特征。它还打算产生跨越两个领域的联合时间表征。

- **分类模块:** 它是最后一个模块，使用前两个模块获得的特征进行最终识别。

对于空间和基于注意力的时间模块，本方案首先通过快速傅里叶变换沿着慢速的 WiFi 矩阵 x 计算出 x_f 。然后，时域和频域都被作为输入使用。

A. 空间模块

该模块被设计用来提取空间特征 H^{spat} 。具体来说，它从时域 x 和频域 x_f 的 WiFi 矩阵学习空间特征。CNN 是一个相对成熟的深度学习网络，所以本方案选择它作为骨干网络。通常情况下，CNN 被用来从图像中提取各种特征。因此，本方案认为 x 和 x_f 是识别网络的图像输入。此外，一个卷积层也被用来使通道尺寸适应骨干网络的预定尺寸。

此外，本方案还对空间分离和融合模块进行了实验。如图4.10a所示，空间分离模块的设计是通过应用骨架生成 H^x 和 H^{x_f} 的 x 和 x_f 特征。然后，通过连接 H^x 和 H^{x_f} 生成最终的空间特征 H^{spat} 。由于主干线的重复使用，空间分离模块具有很高的计算复杂性。因此，为了解决这个问题，本方案进一步设计了空间融合模块，如图4.10b所示。为了使这个过程更容易，本方案首先将 x 和 x_f 串联起来，得到 x_c 。之后，本方案旨在从 x_c 中提取一个复合特征。本方案将 $x_c \in \mathbb{R}^{K \times L \times 2N_r}$ 重塑为 $x_c \in \mathbb{R}^{K \times 2N_r, L}$ 。然后采用由整流线性单元激活的致密层来生成复合特征 H^c 。最后直接使用骨架提取 H^{spat} 。

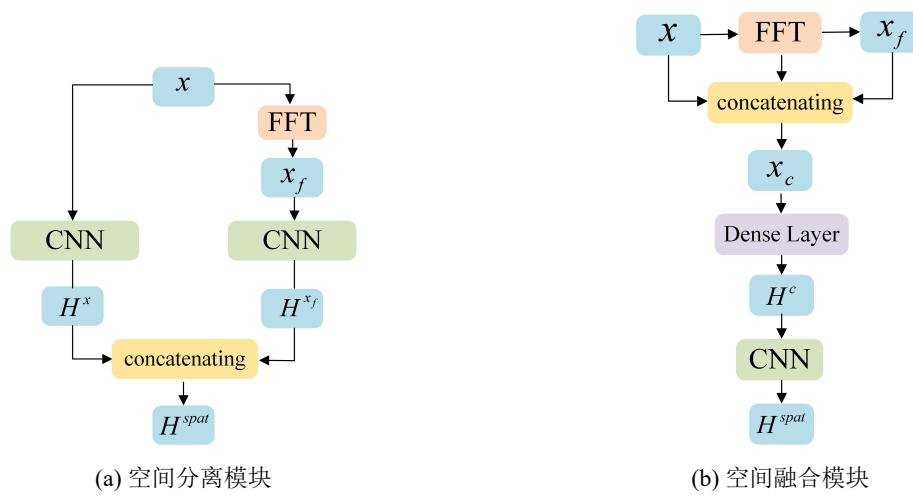


图 4.10 两个可选的空间模块

B. 基于注意力的时间模块

本方案用这个模块学习时间特征 H^{time} 和频率特征 H^{freq} ，从时域和频域的 WiFi 矩阵中提取长期时间特征。如图4.11所示，本方案使用长短期记忆网络来提取时间特征

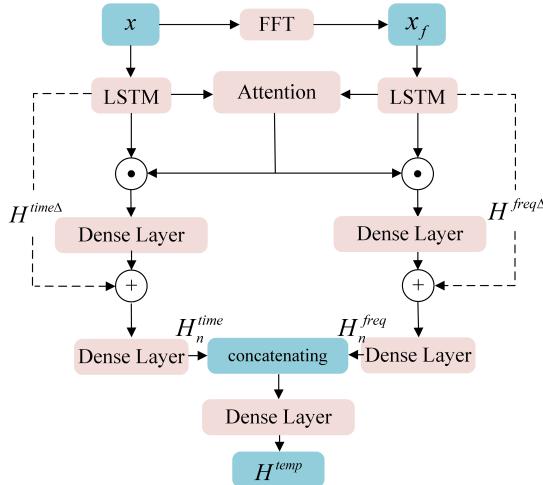


图 4.11 基于注意力的时间模块

$H^{time\Delta}$ 和频率特征 $H^{freq\Delta}$ ，以规避梯度消失问题，重新塑造 x 和 x_f 。然后，LSTM 通过接收输入 x 和 x_f ，分别用于提取时间特征和频率特征。接着，每一步创建一个包含 $H^{time\Delta}$ 和 $H^{freq\Delta}$ 之间联合信息的注意图 \mathbb{A} ，本方案专注于注意机制^[71]以获取初始特征之间有价值组合特征。本方案使用 ReLU 激活的密集层首先将 $H^{time\Delta}$, $H^{freq\Delta}$ 映射到 H^{time*} , H^{freq*} 。最后，注意力图被制作成 $\mathbb{A} = softmax(W \odot H^{time*}(H^{freq*})^T)$ ，其中 \odot 代表 Hadamard 积， T 代表转置， W 代表权重矩阵。

当初始特征与注意力映射 \mathbb{A} 结合时， $((H^{time\Delta})^T \mathbb{A})^T$ 和 $((H^{freq\Delta})^T \mathbb{A})^T$ 的联合特征就会产生，通过密集层增加非线性。为了解决梯度消失问题，构建理想的身份映射，本方案专门将这些映射的联合表示纳入 $H^{time\Delta}$, $H^{freq\Delta}$ ，并通过残差连接^[72]产生 H^{time} 、 H^{freq} 。之后，为了创建一个构建的时间特征 H^{temp} ，本方案选择时间和频率特征， H_n^{time} 和 H_n^{freq} ，参加最后一步（第 n 步）。具体而言，本方案首先使用密集层创建 H_n^{time} 和 H_n^{freq} 。最后，为了计算 H^{temp} ，本方案将这些映射的特征堆叠在密集层之后。

C. 分类模块

为了从时域和频域中提取空间特征 H^{spat} 和时间特征 H^{temp} ，本方案使用了空间模块和基于注意力的时间模块。本方案还通过结合空间和时间表征生成一个综合特征，以准确预测活动标签。本方案适当地设置了隐藏维度，以确保 H^{spat} 和 H^{temp} 的维度相等，以实现这种整合。最后，本方案使用密集层预测活动标签 $y = H^{fuse}W$ ，其中 W 是权重矩阵。

4.3.4 数字孪生系统搭建

A. 人体模型建模

在数字空间中，需要一个熟悉被检测人体的人体模型。这个模型能够映射出物理实体空间中的人体活动。有许多商业上可用的三维建模软件来为人体建模。然而，每个人



图 4.12 使用 ICON 进行人体建模

都有不同的身高和体重，如果对每个人体进行手工建模，这个工作量将是巨大的。更何况本方案对三维建模软件的使用并不熟悉。

虽然有一些方法是通过三维扫描姿势或二维图像与仔细控制的用户姿势进行三维建模，但这些方法需要多余的仪器，并将增加工作成本。幸运的是，文章中的工作^[73]可以在没有冗余仪器的情况下实现人体的快速三维建模，这为本方案的人体建模工作提供了很大的帮助。他们提出的 ICON 能够从无约束姿势的人的二维图像中学习头像，估计出一个详细的三维表面，然后将这些图像组合成一个可动画的头像，还能捕捉到头发和衣服等细节。如图4.12所示，这是使用 ICON 方法对站立时的人体进行的三维建模。这篇文章和代码可以在 <https://icon.is.tue.mpg.de> 找到。

B. 数字场景建模

本方案采用 Unity 游戏引擎作为数字双胞胎的主要模拟平台，它有以下特点：

- **可视化：**Unity 具有强大的图形渲染和可视化能力，所以被广泛用于 3D 游戏开发。
- **平台整合能力：**Unity 的脚本 API 能够与第三方程序集成，增强了仿真平台的能力。

• **素材丰富：**Unity 的官方授权用户和开发者分享他们的 Unity 资源，允许任何项目建立在另一个项目之上，而不需要从头开始。

本方案使用 Unity 素材来构建应用场景，并将建模的人体模型导入项目中。本方案将活动识别系统连接到数字孪生体上，并将识别到的类别转移到 Unity 项目中，该项目将显示相应的人体活动动画。图4.13显示了 Unity 构建的数字孪生中办公室场景中人们鼓掌的活动模型。



图 4.13 用 Unity 创建的数字空间

4.4 实验设计与分析

本方案选择了多个基线场景来比较本方案提出的 WiFi-Net 和基础网络的性能。本方案的实验的实施在以下段落中简单说明。

4.4.1 实验对比基线

本方案用提出的基础网络评估以下基线网络。这是为了证明基础网络在为人体动作识别提取 WiFi 信号矩阵特征方面的有效性。

- **SPACE-time:** 空间模块只包含时域数据。它被用来比较具有时域空间特征的网络。
 - **SPACE-freq:** 空间模块只包含频率域的数据。它被用来比较具有空间特征的频域网络。
 - **SPACE-sep:** 设计的空间分离模块。它被用来研究具有空间特征的网络在时域和频域的能力。
 - **SPACE-fuse:** 设计的空间融合模块。它用于检查具有时域和频域空间特征的网络。
 - **Bi-Path-noATT:** 设计的 Bi-Path 基础网络没有携带注意力模块。它被用来证明注意力机制的有效性。
 - **Bi-Path:** 设计的 Bi-Path 基础网络。
- 本方案探索了 HAR-Net 在适应新环境时的一次性人体动作识别性能。因此，本方案将 HAR-Net 与下面的基线进行比较。
- **Fine-Tuning (FT):** Bi-Path 基础网络首先使用训练集作为源模型进行预训练。然后通过复制源模型的结构和参数来初始化新的网络，并用测试集进行训练。
 - **MAML:** 一个基于优化的元学习基线。使用梯度下降进行独立于模型的元优化是 MAML 的一个主要方面。

• **Prototype Network (PN):** 一个基于度量的元学习基线。由于很少有结果是有标签的，PN 为每个类别生成特征表示。它通过利用欧氏距离度量，寻找最接近的原型作为预测的类别。

- **HAR-Net^{*}:** 没有 WiFi 度量模块的 HAR-Net。

- **HAR-Net:** 带有 WiFi 度量模块的 HAR-Net。

首先，本方案对所有的数据集进行了统一的预处理。之后，本方案在数据集中抽出 80% 的数据进行训练，剩下的数据用于测试，并在实验结束后进行 10 倍交叉验证。本方案在测试数据集上进行了充分的实验，包括 1-Shot、2-Shot 和 3-Shot 场景。为了避免训练过程中的过度拟合，本方案对输入数据进行了批量归一化。本方案的程序是用 PyTorch 编程的，由拥有 24GB 内存的 NVIDIA TESLA P40 运行。

4.4.2 超参数选择

为了以后能够与基线网络进行准确的比较，本方案首先进行了实验，为 Bi-Path 基础网络选择最合适超参数。具体来说，本方案对骨干网络、隐藏维度、优化器和激活函数的选择进行了实验。

A. 骨干网络选择

本方案需要选择一个理想的骨干网络，因为本方案的基础网络使用骨干网络来提取空间特征。因此，本方案选择了几个流行的 CNN 网络，并对它们进行了评估，以选择最适合本方案基础网络的 CNN 网络。这些网络包括两个浅层网络：CNN5 和 Alexnet^[74]，两个中等深度的网络：Resnet18^[72] 和 Vgg19^[75]，以及两个深度网络：Shufflenet44^[76] 和 Resnet152。不同网络的性能显示在图4.14 中。与其他骨干网络相比，本方案发现 Resnet18 具有最大的准确率，特别是对于一次性人体动作识别。本方案还研究了 Resnet 模型开销的复杂性、推理时间和内存要求。一个现有的基线测试显示^[77]，Resnet18 实现了低推理时间、低模型复杂度和突出的实时性能。此外，与其他轻量级模型相比，Resnet18 能够在商业边缘设备上实现相当的性能。考虑到所有方面，本方案选择 ResNet18 作为人体动作识别的骨干网络。

B. 隐藏尺寸选择

在4.3.2中，讨论了本方案的网络有一个隐藏的维度 α 。因此，本方案需要为隐藏维度找到最合适参数。如图所示，本方案为 α 设置了三个参数，分别是 128、256 和 512，在不同的参数下，识别精度有明显的变化。对于平均值来说， $\alpha = 128$ 和 $\alpha = 512$ 设置的准确率差异不大，但对于单次人体动作识别任务来说， $\alpha = 512$ 的表现更好。由于本方案主要关注单次人体动作识别任务，本方案将参数 α 设置为 512。



图 4.14 CNN 骨干网的选择

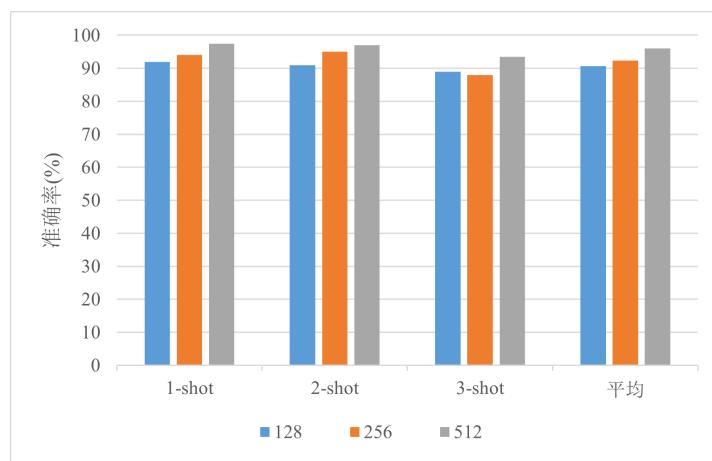


图 4.15 隐藏尺寸选择

C. 优化器的选择

优化器在深度学习中起着至关重要的作用。深度学习模型通常包含大量的参数，而优化器的作用就是根据模型的误差反向传播计算出每个参数的梯度，并利用这些梯度来更新模型参数，以最小化损失函数。通过不断迭代和更新，优化器帮助模型逐渐接近或达到最佳状态。优化器的选择对模型的性能和训练速度都有很大的影响，正确的优化器可以带来更好的训练结果。所以本方案选择 Adam、SDG 和 RMSprop 这三个使用较多的优化器进行比较，为 HAR-Net 选择最适合的优化器。不同优化器的性能可以在图4.16 中观察到。可以看出，SGD 是三个优化器中效果最差的，ADAM 和 RMSPROP 优化器都有较高的精度，但 ADAM 更稳定，所以本方案最终选择了 ADAM 优化器。

D. 激活函数选择

网络中的激活函数对最终的识别精度有着不可忽视的作用，所以本方案试图找到最适合网络的激活函数，本方案用 ReLU、Sigmoid、Softplus 和 Tanh 做了实验。图4.17 显示了不同激活函数对活动识别精度的影响。从图4.17中，本方案可以看到 ReLU 函数优于其他几个函数，所以本方案将网络中的激活函数设置为 ReLU。

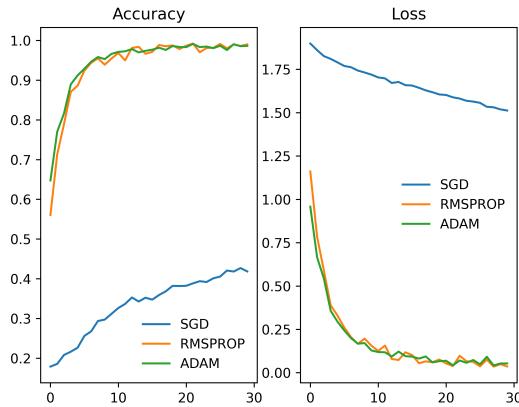


图 4.16 使用不同优化器的训练网络的比较

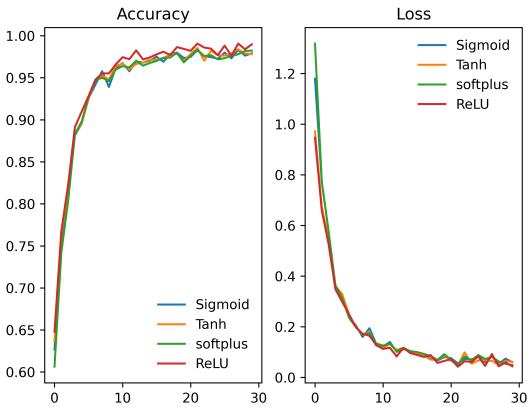


图 4.17 使用不同激活函数的训练网络的对比

4.4.3 实验结果

接下来，本方案在 WiFi 数据集上将提出的网络与基线网络进行比较。本方案使用 Adam 优化器来训练网络，设置如下：批量大小为 10，学习率为 0.0001，epochs 数为 60。

A. Bi-Path 基础网络的性能

图4.18显示了 Bi-Path 基础网络和基线网络的人类活动识别精度。对于所有的场景，本方案的基础网络在活动识别的准确性方面优于其他基线。它证明了 Bi-Path 基础网络在人类活动识别任务中从 WiFi 信号中提取空间和时间特征的能力。本方案观察到，当考虑到准确性和时间复杂性时，SPACE-fuse 比 SPACE-sep 有很大优势。一个潜在的解释可能是，使用两个 CNN 骨干网络增加了参数的数量，从而使 SPACE-sep 的训练和收敛更加困难。

B. HAR-Net 的性能

图4.19显示了 HAR-Net 与其他基线网络的比较结果。可以看出，本方案提出的 HAR-Net 在识别精度方面优于其他基线网络。特别是，带有 WiFi 度量模块的 HAR-Net 比 HAR-Net* 有更高的准确性，并且在不同的场景下都很稳定。这反映了 WiFi 度量模块的功能。

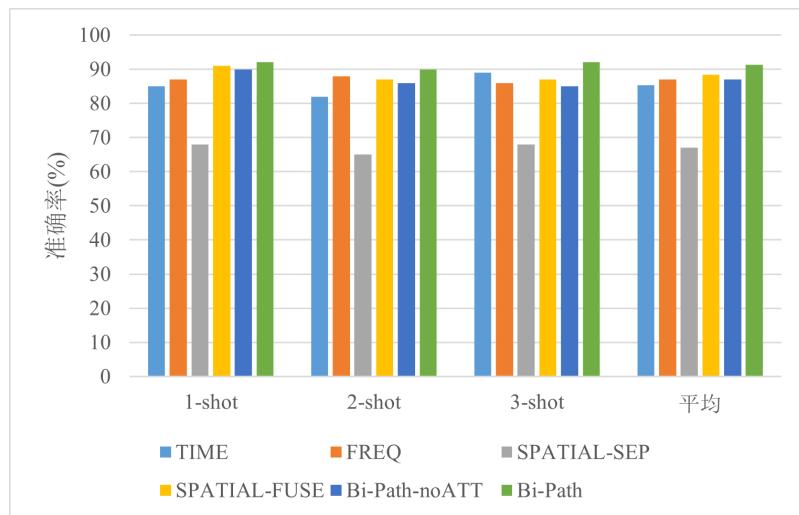


图 4.18 Bi-Path 基础网络的总体比较结果

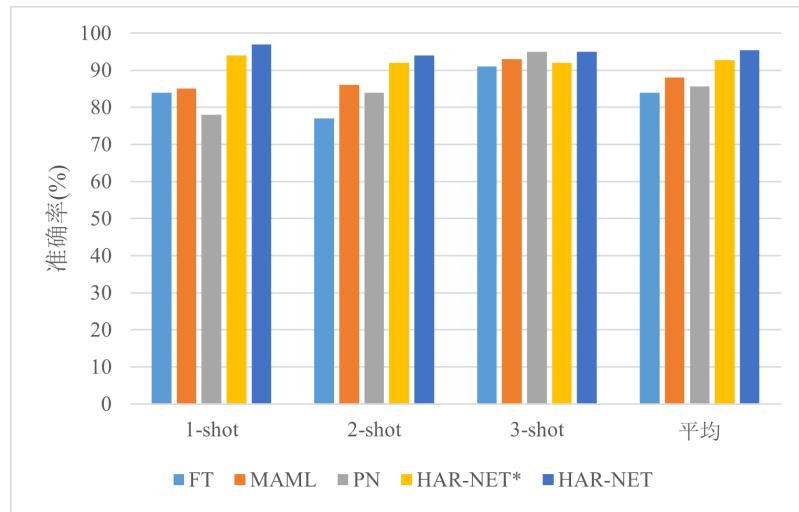


图 4.19 HAR-Net 评估的整体比较结果

表 4.3 不同方法的准确性

| | RF | DT | SVM | CNN | BLSTM | HAR-NET |
|----|--------|--------|--------|--------|--------|---------|
| 训练 | 85.35% | 87.43% | 88.74% | 91.52% | 92.37% | 95.61% |
| 验证 | 83.63% | 86.21% | 88.32% | 90.17% | 91.44% | 93.32% |

C. 与先前方法的比较

为了评估所提出的方法的性能，本方案将其与一些早期基于 CSI 的人体动作识别方法进行了比较。表4.3列出了本方案的方法和其他五种方法的准确率，从中可以看出，本方案的方法具有最高的准确率。为了证明本方案的方法对环境的适应性，本方案还在不同的环境数据集下进行测试。如表4.4所示，本方案的方法对环境的适应性很强。

表 4.4 不同环境下的综合准确性

| | RF | DT | SVM | CNN | BLSTM | HAR-NET |
|----|--------|--------|--------|--------|--------|---------|
| 训练 | 73.61% | 74.36% | 75.14% | 79.33% | 80.42% | 85.72% |
| 验证 | 70.25% | 71.62% | 73.53% | 77.65% | 78.15% | 84.03% |

4.5 本章小结

由于用于 WiFi 人体动作识别的数据集比较少，本章节首先介绍如何制作数据集，详细的说明数据采集需要的软硬件、具体步骤和数据预处理方法。得益于数字孪生技术的快速发展，本章节设计了一个基于深度学习的数字孪生模型用于人类活动识别。基于对 WiFi 感知和元学习方法的探索，本章节提出了一个人体动作识别网络。利用元学习的能力，可以调整模型的超参数，本章节的人体动作识别网络可以适应新环境。元学习框架是本章节人体动作识别网络的一个重要组成部分，它包括 WiFi 度量模块和 Bi-Path 基础网络。它们的功能是训练稳健的距离度量并利用 WiFi 中的高维特征。本章节将人体动作识别与数字孪生相结合，并使用 Unity 建立了一个数字空间。在这个数字空间中，本章节建立了一个类人模型，映射出真实的人类活动。本章节对提出的的人体动作识别网络进行了各种实验，这些实验能够证明人体动作识别网络的性能。

第五章 总结与展望

5.1 本文总结

本文针对现有基于 WiFi 的非接触式人体动作识别方案存在的问题，如检测精度低、数据集少、无法可视化等进行了深入研究和分析，并设计了多种不同的基于 WiFi 的人体动作识别方法，以满足不同场景下的需求。本文的主要工作可以概括如下：

(1) 为了提高基于 WiFi 的人体动作识别准确率，本文提出了一种基于深度学习的框架，通过 WiFi CSI 数据实现非接触式实时活动检测。该框架采用了多层次双向长短期记忆网络模型，并结合注意力模型进行特征学习和活动识别。实验结果表明，本文提出的方法在六种不同的活动识别任务中，准确率达到了 96% 以上，超过了现有的其他方法。本研究为实现高准确度的人体动作识别提供了一种有效的解决方案，并具有广泛的应用前景。

(2) 针对人体动作识别无法可视化的问题，本文提出了一个创新性的人体动作识别系统 HAR-Net，它将人体动作识别和数字孪生相结合，解决了人体动作识别无法可视化的问题。本系统使用传感器收集的 WiFi 时间序列数据来训练深度学习网络，并与 Unity 建立的数字孪生系统相结合。数字孪生系统将真实世界的物理活动映射到人类模型上，并能够实时评估活动预测结果，当危险活动发生时能够迅速发出警报。实验结果表明，本系统对 6 种活动的识别准确率均超过 85%，同时数字孪生系统可以有效地提高预警效率，提高了人体动作识别的实时性和准确性。该研究为实现安全的人体活动监测提供了一个新的思路和解决方案。针对基于 WiFi 的人体动作识别的数据集少的问题，本文尝试搭建获取 CSI 数据的实验环境，收集不同动作对应的 CSI 原始数据包，制作基于 WiFi 的人体动作识别数据集。介绍了采集数据使用的硬件和软件，包括发射器、接收器还有使用的 CSI Tool 软件。详细的记录收集活动数据的具体步骤，然后给出数据预处理的方法。

5.2 本文展望

尽管本文在非接触式人体动作识别方面已经取得了一定的进展，但受限于时间和设备条件，实验并未得到充分的展开和深入探究。因此，可以进一步完善和深入研究当前的工作。下面是对未来研究工作的展望：

(1) 实现多人动作识别。虽然本文所研究的基于 WiFi 的人体动作识别方法在单人场景下取得了很好的效果，但是在实际日常生活中，人们往往是在多人共存的环境下活

动,这就需要将识别单个人的动作扩展到识别多人的动作。因此,未来的研究方向之一是实现多人动作识别。这需要采用多种技术来实现,如多目标跟踪、多传感器融合等。其中,多目标跟踪技术可以用于识别不同人员的动作,而多传感器融合技术可以将多个传感器的信息进行整合,提高动作识别的准确性和鲁棒性。此外,还需要考虑到人员之间的交互,如合作、对抗等,以及对不同人员的个性化动作识别的问题。因此,未来需要进一步深入研究多人动作识别技术,以实现更加准确和可靠的动作识别。

(2) 提高环境鲁棒性。本文中所采集的实验数据是在室内环境中完成的,为了使模型的预测分类效果更准确,本文进行了建模训练,其结果在室内环境下的动作模型分类表现良好。但是在实际应用中,由于不同环境中的信道状态信息差异较大,将训练好的模型直接应用到其他环境中可能会导致表现效果差或者完全失效。例如,信道衰减、干扰和多径效应等因素会对信道状态信息产生影响,导致模型的泛化能力较差。因此,需要针对不同的环境进行数据采集、建模训练和模型调整,以提高模型的适应性和泛化能力。同时,也需要考虑如何更好地利用不同环境中的信道状态信息特征,从而进一步提高动作识别的准确率和鲁棒性。

(3) 增加动作种类。本文对六种动作进行了识别,但是对于人类的日常活动,这些动作的种类和数量都远远不够。未来的研究应该增加更多的动作种类和更全面的动作类型,以使识别系统更加准确和实用。同时,可以进一步研究如何将识别结果与人体动作的语义信息相结合,进一步提升识别系统的实用性和可用性。

参考文献

- [1] Buoncompagni L, Kareem S Y, Mastrogiovanni F. Human Activity Recognition Models in Ontology Networks[J]. IEEE Transactions on Cybernetics, 2022, 52(6):5587–5606.
- [2] Hamad R A, Hidalgo A S, Bouguelia M R, et al. Efficient Activity Recognition in Smart Homes Using Delayed Fuzzy Temporal Windows on Binary Sensors[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(2):387–395.
- [3] Ding J, Wang Y, Si H, et al. Multimodal Fusion-AdaBoost Based Activity Recognition for Smart Home on WiFi Platform[J]. IEEE Sensors Journal, 2022, 22(5):4661–4674.
- [4] Chapron K, Lapointe P, Bouchard K, et al. Highly Accurate Bathroom Activity Recognition Using Infrared Proximity Sensors[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(8):2368–2377.
- [5] Susam B T, Riek N T, Akcakaya M, et al. Automated Pain Assessment in Children Using Electrodermal Activity and Video Data Fusion via Machine Learning[J]. IEEE Transactions on Biomedical Engineering, 2022, 69(1):422–431.
- [6] Sadreazami H, Bolic M, Rajan S. Contactless Fall Detection Using Time-Frequency Analysis and Convolutional Neural Networks[J]. IEEE Transactions on Industrial Informatics, 2021, 17(10):6842–6851.
- [7] Liu J, Tan R, Han G, et al. Privacy-Preserving In-Home Fall Detection Using Visual Shielding Sensing and Private Information-Embedding[J]. IEEE Transactions on Multimedia, 2021, 23:3684–3699.
- [8] Jin F, Sengupta A, Cao S. mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(2):1245–1257.
- [9] Wang J, Tian Z, Zhou M, et al. Leveraging Hypothesis Testing for CSI Based Passive Human Intrusion Direction Detection[J]. IEEE Transactions on Vehicular Technology, 2021, 70(8):7749–7763.
- [10] Wu Z, Zhang X, Tian G, et al. Spatial-Temporal Graph Network for Video Crowd Counting[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(1):228–241.

- [11] Liu Z, Yuan R, Yuan Y, et al. A Sensor-Free Crowd Counting Framework for Indoor Environments Based on Channel State Information[J]. IEEE Sensors Journal, 2022, 22(6):6062–6071.
- [12] Yang Y, Luo Y, Qi A, et al. Huber Kalman Filter for Wi-Fi Based Vehicle Driver's Respiration Detection[J]. IEEE Transactions on Vehicular Technology, 2022, 71(8):8933–8943.
- [13] Yang Y, Cao J, Wang Y. Robust RFID-Based Respiration Monitoring in Dynamic Environments[J]. IEEE Transactions on Mobile Computing, 2023, 22(3):1717–1730.
- [14] Lin G, Jiang W, Xu S, et al. Human Activity Recognition Using Smartphones With WiFi Signals[J]. IEEE Transactions on Human-Machine Systems, 2023, 53(1):142–153.
- [15] Zhang Y, Liu Q, Wang Y, et al. CSI-Based Location-Independent Human Activity Recognition Using Feature Fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1–12.
- [16] Wang T, Ng W W Y, Li J, et al. A Deep Clustering via Automatic Feature Embedded Learning for Human Activity Recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(1):210–223.
- [17] Shu X, Zhang L, Sun Y, et al. Host–Parasite: Graph LSTM-in-LSTM for Group Activity Recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(2):663–674.
- [18] Xu X, Yu J, Chen Y. Leveraging Acoustic Signals for Fine-Grained Breathing Monitoring in Driving Environments[J]. IEEE Transactions on Mobile Computing, 2022, 21(3):1018–1033.
- [19] Do H M, Welch K C, Sheng W. SoHAM: A Sound-Based Human Activity Monitoring Framework for Home Service Robots[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(3):2369–2383.
- [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, 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, 2021. 1–1.

- [22] Zhang Z. Microsoft Kinect Sensor and Its Effect[J]. IEEE MultiMedia, 2012, 19(2):4–10.
- [23] Al-qaness M A A, Dahou A, Elaziz M A, et al. Multi-ResAtt: Multilevel Residual Network With Attention for Human Activity Recognition Using Wearable Sensors[J]. IEEE Transactions on Industrial Informatics, 2023, 19(1):144–152.
- [24] Xu S, Zhang L, Huang W, et al. Deformable Convolutional Networks for Multimodal Human Activity Recognition Using Wearable Sensors[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:1–14.
- [25] Jamil H, Qayyum F, Iqbal N, et al. Optimal Ensemble Scheme for Human Activity Recognition and Floor Detection Based on AutoML and Weighted Soft Voting Using Smartphone Sensors[J]. IEEE Sensors Journal, 2023, 23(3):2878–2890.
- [26] Moniruzzaman M, Yin Z, He Z, et al. Human Action Recognition by Discriminative Feature Pooling and Video Segment Attention Model[J]. IEEE Transactions on Multimedia, 2022, 24:689–701.
- [27] Luo H, Lin G, Yao Y, et al. Dense Semantics-Assisted Networks for Video Action Recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5):3073–3084.
- [28] Cui H, Dahnoun N. Real-Time Short-Range Human Posture Estimation Using mmWave Radars and Neural Networks[J]. IEEE Sensors Journal, 2022, 22(1):535–543.
- [29] Muta O, Takata K, Noguchi K, et al. Device-Free WLAN Based Indoor Localization Scheme With Spatially Concatenated CSI and Distributed Antennas[J]. IEEE Transactions on Vehicular Technology, 2023, 72(1):852–865.
- [30] Tong X, Wang H, Liu X, et al. MapFi: Autonomous Mapping of Wi-Fi Infrastructure for Indoor Localization[J]. IEEE Transactions on Mobile Computing, 2023, 22(3):1566–1580.
- [31] Li B, Cui W, Wang W, et al. Two-stream convolution augmented transformer for human activity recognition[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 286–293.
- [32] Duan P, Li C, Li J, et al. WISDOM: Wi-Fi-Based Contactless Multiuser Activity Recognition[J]. IEEE Internet of Things Journal, 2023, 10(2):1876–1886.

- [33] Zhang N, Zhang J, Ying Y, et al. Wi-Phrase: Deep Residual-Multihead Model for WiFi Sign Language Phrase Recognition[J]. IEEE Internet of Things Journal, 2022, 9(18):18015–18027.
- [34] Peng R, Tian Y, Han S. Multiuser Wireless Hand Gesture Recognition by Spatial Beam-forming[J]. IEEE Transactions on Vehicular Technology, 2023, 72(1):805–820.
- [35] 金浩. 基于 RSSI 的无线传感器网络井下人员定位算法研究 [D]. 江苏大学, 2022.
- [36] 薛卫星. RSSI 在室内定位中的测量特性及应用研究 [J]. 测绘学报, 2022, 51:2558.
- [37] Lin Y, Yu K, Hao L, et al. An Indoor Wi-Fi Localization Algorithm Using Ranging Model Constructed With Transformed RSSI and BP Neural Network[J]. IEEE Transactions on Communications, 2022, 70(3):2163–2177.
- [38] 陈禹. 基于 LSTM 的室内定位系统设计与实现 [D]. 南京邮电大学, 2022.
- [39] Abidine M B, Fergani B. An Efficient Kernel KNN classifier for Activity Recognition on Smartphone[C]. Proceedings of 2022 IEEE 21st International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA). 2022, 73–78.
- [40] Fan L, Wang Z, Wang H. Human Activity Recognition Model Based on Decision Tree[C]. Proceedings of 2013 International Conference on Advanced Cloud and Big Data. 2013, 64–68.
- [41] 郑志远, 朱文章. 双域融合和 LSTM 下的 FMCW 雷达人体活动识别方法 [J]. 厦门理工学院学报, 2022, 30:15–21.
- [42] Ji J, Xu C, Zhang X, et al. Spatio-Temporal Memory Attention for Image Captioning[J]. IEEE Transactions on Image Processing, 2020, 29:7615–7628.
- [43] Galassi A, Lippi M, Torroni P. Attention in Natural Language Processing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(10):4291–4308.
- [44] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. International Conference on Learning Representations, 2014.
- [45] 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.
- [46] Wang W, Liu A X, Shahzad M, et al. Device-Free Human Activity Recognition Using Commercial WiFi Devices[J]. IEEE Journal on Selected Areas in Communications, 2017,

- 35(5):1118–1131.
- [47] Gao Q, Wang J, Ma X, 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.
- [48] Wang C, Lu W, Narayanan M R, et al. Low-Power Fall Detector Using Triaxial Accelerometry and Barometric Pressure Sensing[J]. IEEE Transactions on Industrial Informatics, 2016, 12(6):2302–2311.
- [49] Tao F, Zhang H, Liu A, et al. Digital Twin in Industry: State-of-the-Art[J]. IEEE Transactions on Industrial Informatics, 2019, 15(4):2405–2415.
- [50] Hung M H, Lin Y C, Hsiao H C, et al. A Novel Implementation Framework of Digital Twins for Intelligent Manufacturing Based on Container Technology and Cloud Manufacturing Services[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(3):1614–1630.
- [51] Yu J, Song Y, Tang D, et al. A Digital Twin approach based on nonparametric Bayesian network for complex system health monitoring[J]. Journal of Manufacturing Systems, 2021, 58:293–304. Digital Twin towards Smart Manufacturing and Industry 4.0.
- [52] Leng J, Zhang H, Yan D, et al. Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10:1155–1166.
- [53] Liu D, Du Y, Chai W, et al. Digital Twin and Data-Driven Quality Prediction of Complex Die-Casting Manufacturing[J]. IEEE Transactions on Industrial Informatics, 2022, 18(11):8119–8128.
- [54] Perez L, Rodríguez-Jiménez S, Rodríguez N, et al. Digital Twin and Virtual Reality Based Methodology for Multi-Robot Manufacturing Cell Commissioning[J]. Applied Sciences, 2020, 10:3633.
- [55] Bjornsson B, Borrebaeck C, Elander N, et al. Digital twins to personalize medicine[J]. Genome Medicine, 2019.
- [56] Zheng Y, Lu R, Guan Y, et al. Towards Private Similarity Query based Healthcare Monitoring over Digital Twin Cloud Platform[C]. Proceedings of 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS). 2021, 1–10.

- [57] Mazumder O, Roy D, Bhattacharya S, et al. Synthetic PPG generation from haemodynamic model with baroreflex autoregulation: a Digital twin of cardiovascular system[C]. Proceedings of 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019, 5024–5029.
- [58] Barricelli B R, Casiraghi E, Gliozzo J, et al. Human Digital Twin for Fitness Management[J]. IEEE Access, 2020, 8:26637–26664.
- [59] Chakshu N K, Carson J, Sazonov I, et al. A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration—A coupled computational mechanics and computer vision method[J]. International Journal for Numerical Methods in Biomedical Engineering, 2019, 35(5):e3180. E3180 cmn.3180.
- [60] Chakshu N K, Sazonov I, Nithiarasu P. Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis[J]. Biomechanics and Modeling in Mechanobiology, 2021.
- [61] Dong R, She C, Hardjawana W, et al. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn From a Digital Twin[J]. IEEE Transactions on Wireless Communications, 2019, 18(10):4692–4707.
- [62] Malik A A, Bilberg A. Digital twins of human robot collaboration in a production setting[J]. Procedia Manufacturing, 2018, 17:278–285. 28th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018), June 11-14, 2018, Columbus, OH, USAGlobal Integration of Intelligent Manufacturing and Smart Industry for Good of Humanity.
- [63] Kaigom E G, Roßmann J. Value-Driven Robotic Digital Twins in Cyber-Physical Applications[J]. IEEE Transactions on Industrial Informatics, 2021, 17(5):3609–3619.
- [64] Wang K J, Lee Y H, Angelica S. Digital twin design for real-time monitoring –a case study of die cutting machine[J]. International Journal of Production Research, 2020, 59:1–15.
- [65] 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.
- [66] 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.

- [67] Halperin D, Hu W, Sheth A, et al. Tool release[J]. ACM SIGCOMM Computer Communication Review, 2011, 41:53–53.
- [68] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[J]. Advances in neural information processing systems, 2016.
- [69] Chen Z, Zhang L, Jiang C, et al. WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM[J]. IEEE Transactions on Mobile Computing, 2019, 18(11):2714–2724.
- [70] Jiang W, Miao C, Ma F, et al. Towards environment independent device free human activity recognition[C]. Proceedings of the 24th annual international conference on mobile computing and networking. 2018, 289–304.
- [71] Kim J H, Jun J, Zhang B T. Bilinear Attention Networks[C]. Proceedings of NeurIPS. 2018.
- [72] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, 770–778.
- [73] Xiu Y, Yang J, Tzionas D, et al. ICON: Implicit Clothed humans Obtained from Normals[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022, 13296–13306.
- [74] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Commun. ACM, 2017, 60(6):84–90.
- [75] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. CoRR, 2015.
- [76] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
- [77] Bianco S, Cadene R, Celona L, et al. Benchmark Analysis of Representative Deep Neural Network Architectures[J]. IEEE Access, 2018, 6:64270–64277.