

基于连续图像深度学习的 Wi-Fi 人体行为识别方法

周启臻¹, 邢建春¹, 杨启亮¹, 韩德帅²

(1. 陆军工程大学国防工程学院, 江苏 南京 210007; 2. 火箭军工程大学作战保障学院, 陕西 西安 710025)

摘 要: 针对基于深度学习的 Wi-Fi 人体行为识别技术存在抗噪声能力弱、信号尺寸不兼容和特征提取不充分等问题, 提出了一种基于连续图像深度学习的识别方法。首先把时变 Wi-Fi 信号重构为若干个连续图像帧, 确保输入尺寸一致; 进而设计低秩分解算法, 对噪声湮没的关键运动信息进行分离; 同时提出一种时间域和空间域信息融合的深度学习模型, 自动捕捉变长图像序列的时空域特征, 并在 WiAR 数据集和自主采集数据集上对所提方法进行验证。实验结果表明, 所提方法平均识别精度分别为 0.94 和 0.96, 具备普适场景下的高精度和稳健性。

关键词: 行为识别; Wi-Fi 信号; 深度学习; 图像识别; 低秩分解

中图分类号: TN391

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020141

Sequential image deep learning-based Wi-Fi human activity recognition method

ZHOU Qizhen¹, XING Jianchun¹, YANG Qiliang¹, HAN Deshuai²

1. College of Defense Engineering, Army Engineering University of PLA, Nanjing 210007, China

2. College of Combat Support, Rocket Force University of Engineering, Xi'an 710025, China

Abstract: For the problems existing in most of the researches, such as weak anti-noise ability, incompatible signal size and insufficient feature extraction of deep-learning-based Wi-Fi human activity recognition, a kind of sequential image deep learning-based recognition method was proposed. Based on the idea of sequential image deep learning, a series of image frames were reconstructed from time-varied Wi-Fi signal to ensure the consistency of input size. In addition, a low-rank decomposition method was innovatively designed to separate low-rank activity information merged in noises. Finally, a deep model combining temporal stream and spatial stream was proposed to automatically capture the spatio-temporal features from length-varied image sequences. The proposed method was extensively tested in WiAR dataset and self collected dataset. The experimental results show the proposed method could achieve the accuracy of 0.94 and 0.96, which indicate its high-accuracy performance and robustness in pervasive environments.

Key words: activity recognition, Wi-Fi signal, deep learning, image recognition, low-rank decomposition

1 引言

人体行为识别是利用计算机技术实现肢体运动自动检测、分析和理解的技术, 广泛应用于智能家居、安防监控、医疗康复、人机交互等新兴领域^[1-2]。人体行为识别通常可以分为两类^[3]: 接触式和非接触式。

可穿戴设备^[4-5]是接触式行为识别系统的关键载体, 但是存在设备昂贵、用户穿戴不便、注意力侵扰等诸多限制。非接触式的行为识别能提供无设备的感知服务和友好的用户交互, 因此得到了研究人员的广泛关注。目前非接触式的行为识别主要依靠深度摄像头^[6]和无线电信号^[7]。然而, 前者的普及受限于光照条件、

收稿日期: 2020-02-03; 修回日期: 2020-06-08

通信作者: 邢建春, xjc@893.com.cn

基金项目: 国家重点研发计划基金资助项目 (No.2017YFC0704100)

Foundation Item: The National Key Research and Development Program of China (No.2017YFC0704100)

视距路径干扰、监控死角和隐私问题；后者通常需要部署专用的实验设备，难以大规模推广。

作为城市基础设施，商用 Wi-Fi 设备^[8]所提供的无线信号具有普适、易用、高性价比、传输距离远等优势，近年来在人体行为识别领域被广泛关注^[9]，早期工作基于 Wi-Fi 接收信号强度（RSS, received signal strength）实现简单的室内人员运动检测^[10]。为了进一步提升感知粒度和识别精度，研究人员开始从 Wi-Fi 商业网卡中提取物理层的信道状态信息（CSI, channel state information）^[11]。CSI 能够刻画感知目标所引发多径链路变化，从而利用子载波层次的幅值和相位信息实现更复杂精细的行为识别，如室内活动监测^[12]、摔倒检测^[13]、步态认证^[14]和手势识别^[15]。然而大多数方法在建立 Wi-Fi 信号扰动与人体行为之间的映射关系时，需要丰富的专业知识和领域经验来指导模型设计^[16]、信号挖掘^[17]和特征选取^[18]，这使系统的稳定性、实用性和整体精度不高。如何利用最少的人为干预，实现稳健、普适、高精度的人体行为识别成为亟需解决的问题^[19]。

近年来，受图像深度学习^[20]的启发，研究人员尝试将信号矩阵转化为灰度图像并利用卷积神经网络（CNN, convolutional neural network）自动学习尺度不变、位移不变的局部空间特征^[21-22]。然而，相比于专业相机拍摄的主体清晰、固定尺寸连续图像，直接利用 CNN 处理 Wi-Fi 信号序列无疑更具挑战性。不同动作的初始时间-振幅曲线如图 1 所示，其面临的技术挑战可概括如下。

1) 噪声干扰严重。尽管 CSI 能宏观地反映人体运动状态的变化，但是商用 Wi-Fi 设备所提供信号粒度有限，使感知目标所产生的信号振幅容易混杂于背景噪声和多径干扰中^[9,16]。现有工作往往难以提取运动关键信息，导致深度网络无法从混杂扰动中学到目标信号的本质特征。

2) 信号尺寸不兼容。人体活动的发生具有随机性，因此每一次活动发生时刻均不一致，所产生的信号样本在时间尺度上并不统一，因此导致其与 CNN 模型所要求的固定尺寸输入存在矛盾^[22]。现有工作通常截取固定长度的信号片段以适配模型输入，但是这可能遗漏有效信息（过短）或是增加运算成本（过长），从而影响系统部署的灵活性和时间效率。

3) 特征提取不充分，用户执行复合动作时通常存在上下文的逻辑关系，如完成深蹲动作需要

依次执行站立等待、快速蹲下、快速起身、站立等待的动作，因此会出现先平稳、后快速震荡、最终平稳的波形变化。然而，单一的 CNN 模型无法对这种时间尺度上的依赖关系进行有效建模，导致复杂环境下的系统精度受限。

针对上述问题，本文将 Wi-Fi 信号处理问题转化成图像领域的连续目标识别问题^[23-24]，提出一种基于连续图像深度学习的 Wi-Fi 人体行为识别方法——SIL-Fi（sequential image learning with Wi-Fi）。本文的主要创新点如下。

1) 连续帧低秩分解。通过重构固定尺寸的三维信号图像，将时变的 Wi-Fi CSI 信号序列转化为数量不一、尺寸固定的连续图像帧，同时设计不需要复杂调参的低秩矩阵分解算法，提取稳健的低秩信号主成分，提高噪声环境下运动信号的显著性，从而提高模型训练样本的可靠性。

2) 深度模型融合。通过深度融合 CNN 和长短时记忆（LSTM, long short-term memory）模型，实现变长图像序列空间域和时间域特征的充分萃取，进而克服模型输入限制和特征提取不充分的缺点。

3) 真实场景验证。本文通过多组真实场景实验对模型可行性进行验证，并考虑了不同动作、模型组合、测试人员和环境等诸多因素的影响。与其他基于深度神经网络的模型相比，本文方法在不同实验环境均取得较高的识别精度，因此具备普适场景下的高精度和稳健性。

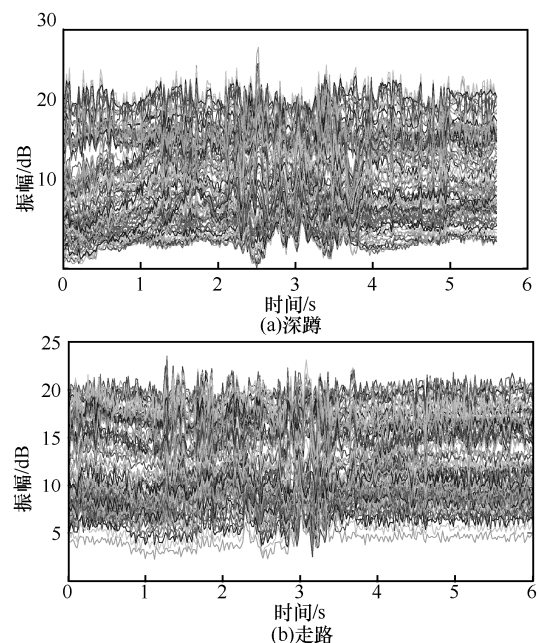


图 1 不同动作的初始时间-振幅曲线

2 SIL-Fi 方法框架

本文提出的 SIL-Fi 的总体框架如图 2 所示。本文的活动数据来源于公开数据集 WiAR^[25]和自主采集的真实数据集。在动作信号预处理模块，首先对所获取的 CSI 原始幅值序列 I 进行分割，即重构成 N 个纵坐标为子载波个数 S' 、横坐标为数据分组长度、通道数为天线数 A 的三维信号矩阵所组成的序列 I' ，以确保每个输入片段的尺寸都保持一致。然后设计稳健主成分分析法 (RPCA, robust principal component analysis)，将分割后的三维信号矩阵视作低秩活动信息、背景信息和测量偏差的叠加，通过约束各个通道低秩矩阵的秩和背景矩阵的稀疏度，提取低秩的关键运动信息，避免运动无关的信号扰动。在时空域特征提取与分类模块，设计了一种融合空间域特征和时间域特征的深度模型，主要包含 3 个卷积层、2 个 LSTM 层和一个全连接层。其中，多层卷积运算通过逐层表征信号矩阵，将低秩矩阵中具有显著识别力的局部空间特征进行抽象并映射到特征图中；LSTM 层通过控制内部门状态，依次实现特征图序列上下文信息的提取和更新，从而对其在时间尺度的全局依赖关系进行建模；全连接层将动作特征映射至样本标记空间，且防止特征过拟合。最后，由 Softmax 层输出预测动作的标签类别。

3 动作信号预处理

为了探究 Wi-Fi CSI 的信号特性，首先介绍 CSI 的背景知识。CSI 信息来源于当前正交频分复用技术中对子载波信道状态信息的估计，反映了信号传

输过程中所产生的时延、能量衰减和相位偏移。令 x 为发射端信号， y 为接收端信号，Noise 为测量噪声， H 为 CSI 信道估计矩阵，则有

$$y = Hx + \text{Noise} \quad (1)$$

设 Wi-Fi 设备持续采集样本， A 根天线接收到的每个数据分组共包含 S 个子载波的信道响应，即

$$H = [H(f_1), \dots, H(f_k), \dots, H(f_S)] \quad (2)$$

其中， $H(f_k)$ 为第 k 个、中心频率为 f_k 的子载波，其幅值为 $\|H(f_k)\|$ ，相位信息为 $\angle H(f_k)$ ，即

$$H(f_k) = \|H(f_k)\| e^{j\angle H(f_k)} \quad (3)$$

当 A 根接收天线以固定采样频率 Q 连续采集 T 个数据分组时，时间长度为 $\frac{T}{Q}$ 。为方便描述，以数据分组长度 T 为时间计量单位。则在 T 范围内，所有 S 个子载波的 CSI 原始幅值序列 I 可以表示为

$$I_{S \times T} = \begin{bmatrix} \|H(f_1)^1\|, & \dots, & \|H(f_1)^T\| \\ \vdots & \ddots & \vdots \\ \|H(f_S)^1\|, & \dots, & \|H(f_S)^T\| \end{bmatrix} \quad (4)$$

由于持续工作的 Wi-Fi 设备所采集的 CSI 幅值序列 I 在时间尺度上不一致，且运动信号往往随机存在于采集样本中，无法通过截短或是填充的方式固定信号尺寸。因此，为方便后续的 CNN 运算，本文首先将完整的 CSI 幅值序列 $I_{S \times T}$ 重构为 $I_{S' \times T' \times A \times N}$ ，即每个序列 I 包括 N 个大小为 $S' \times T' \times A$ 的三维信号矩阵 D 。具体而言，由于本文采用的无线网卡类型为 Intel 5300，接收天线数量 A 为 3，每根接收天线提供子载波个数 S' 为 30，每次截取的数

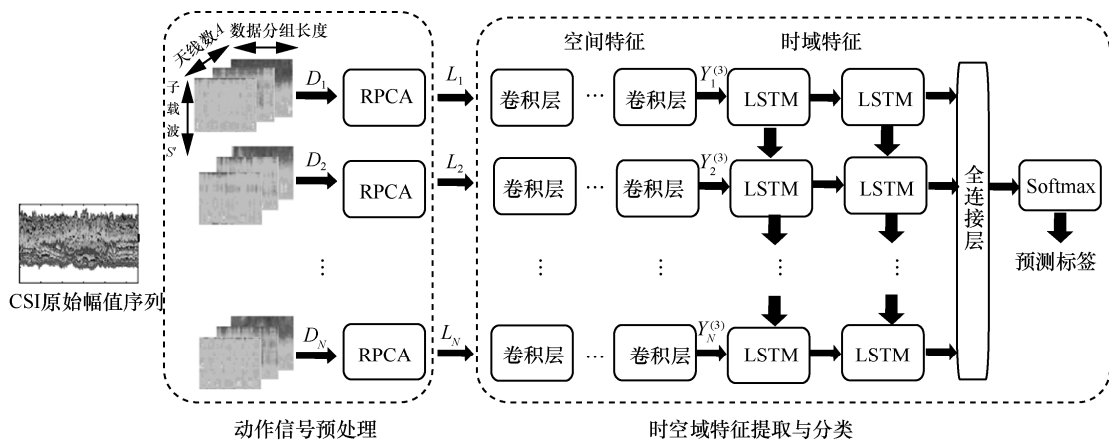


图 2 SIL-Fi 的总体框架

据分组个数 T' 为 30, 截取重叠部分为 15。当采样频率 Q 为 30 Hz 时, 每个输入大小为 $30 \times 30 \times 3$ 的三维信号矩阵 \mathbf{D} 包含了 1 s 内每个通道 (天线) 30 个子载波的全部幅值信息 ($\frac{T'}{Q}=1$ s), 如式(15)所示。

$$\mathbf{D}_{30 \times 30 \times 3} = \begin{bmatrix} \|H(f_1)^1\|, & \cdots, & \|H(f_1)^{30}\| \\ \vdots & \ddots & \vdots \\ \|H(f_{30})^1\|, & \cdots, & \|H(f_{30})^{30}\| \end{bmatrix} \quad (5)$$

因此, 由 N 个三维矩阵 \mathbf{D} 所组成的重构幅值序列 \mathbf{I}' 可以表示为

$$\mathbf{I}'_{30 \times 30 \times 3 \times N} = [\mathbf{D}_{30 \times 30 \times 3}^1, \cdots, \mathbf{D}_{30 \times 30 \times 3}^n, \cdots, \mathbf{D}_{30 \times 30 \times 3}^N] \quad (6)$$

其中, $n \in [1, N]$, 数量 N 由数据分组长度、每次截取的数据分组个数 T' 和滑窗重叠片段共同决定, 表示为 $N = \lfloor \frac{T-30}{15} \rfloor$, 其中 $\lfloor \cdot \rfloor$ 表示数值向下取整。

为了消除无关因素对有效信息的干扰, 大多数相关研究采用主成分分析法进行关联信号的提取。本文受图像前景提取相关工作的启发, 提出利用 RPCA 提取目标动作信息^[23]。假设任意时刻输入 RPCA 的三维信号矩阵 \mathbf{D} 在每个通道的幅值强度矩阵都包含了低秩运动信息、稀疏背景信息以及少量测量噪声。令第 n 个三维信号矩阵 $\mathbf{D}_{30 \times 30 \times 3}^n$ 在通道 v 处的投影为二维矩阵 $\mathbf{E}_{30 \times 30, v}^n$, 通过约束其低秩矩阵 $\mathbf{L}_{30 \times 30, v}^n$ 的秩 r 和背景矩阵 $\mathbf{B}_{30 \times 30, v}^n$ 的稀疏度 b , 实现如式(7)所示目标函数的最小化。

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{B}} & \|\mathbf{E}_{30 \times 30, v}^n - \mathbf{L}_{30 \times 30, v}^n - \mathbf{B}_{30 \times 30, v}^n\|_F^2, \\ \text{s.t.} & \mathbf{E}_{30 \times 30, v}^n = \mathbf{L}_{30 \times 30, v}^n + \mathbf{B}_{30 \times 30, v}^n \\ & \text{rank}(\mathbf{L}_{30 \times 30, v}^n) \leq r \\ & \text{card}(\mathbf{B}_{30 \times 30, v}^n) \leq b \end{aligned} \quad (7)$$

从而将 N 个三维信号矩阵中的低秩部分视为前景信息进行提取, 稀疏的干扰信息视为背景信息进行剔除。RPCA 的稳健性在于: 1) 不需要人为假设噪声类型和强弱, 因此可以有效应对突发噪声和严重的离群数据, 而传统的 PCA 只能处理密集分布的高斯噪声^[24]; 2) 不需要人工挑选信号主成分, 只需要提取低秩矩阵就可以表征高度相关的 CSI 活动信号。这是因为 CSI 信号相邻子载波之间的运动前景信息具有较强的关联性, 近似位于同一低秩的子空间中。然而传统的 PCA 很难在运动信息丰富的第一主成分^[26]和噪声干扰式微的第二主成分^[16]

之间做出选择。利用 RPCA 提取低秩主成分的流程如图 3 所示, 具体步骤如下。

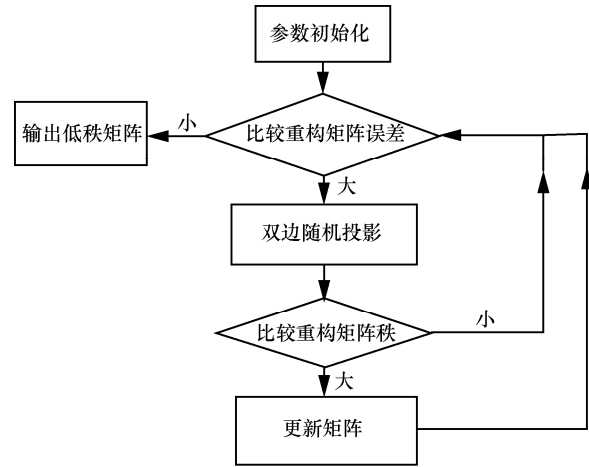


图 3 RPCA 提取低秩信息的流程

步骤 1 参数初始化。需要初始化的参数包括低秩矩阵约束 r 、背景矩阵稀疏度 b 、迭代次数 t 、重构误差 ε 和能量幂值 q 。令初始低秩矩阵 \mathbf{L}_0 为二维矩阵 $\mathbf{E}_{30 \times 30, v}^n$, 背景矩阵 \mathbf{B}_0 为零矩阵。

步骤 2 比较重构矩阵误差。通过增广拉格朗日交替方向乘法, 依次计算第 t 步迭代时, 第 n 个片段在通道 v 处的计算误差 $\frac{\|\mathbf{E}_{30 \times 30, v}^n - \mathbf{L}_{30 \times 30, v}^n(t) - \mathbf{B}_{30 \times 30, v}^n(t)\|_F^2}{\|\mathbf{E}_{30 \times 30, v}^n\|_F^2}$, 并与 ε 比较。若各通道计算误差同时小于 ε , 则执行步骤 6; 否则, 执行步骤 3。

步骤 3 双边随机投影。采取双边随机投影, 对每个通道的低秩矩阵进行逼近。

$$\tilde{\mathbf{L}}_{30 \times 30, v}^n = [(\mathbf{E}_{30 \times 30, v}^n - \mathbf{B}_{30 \times 30, v}^n(t-1)) \cdot (\mathbf{E}_{30 \times 30, v}^n - \mathbf{B}_{30 \times 30, v}^n(t-1))^T]^q \cdot (\mathbf{E}_{30 \times 30, v}^n - \mathbf{B}_{30 \times 30, v}^n(t-1)) \quad (8)$$

$$\mathbf{Y}_1 = \tilde{\mathbf{L}}_{30 \times 30, v}^n \mathbf{G}_1, \mathbf{G}_2 = \mathbf{Y}_1 \quad (9)$$

$$\mathbf{Y}_2 = (\tilde{\mathbf{L}}_{30 \times 30, v}^n)^T \mathbf{G}_2 = \mathbf{Q}_2 \mathbf{R}_2, \mathbf{Y}_1 = (\mathbf{L}_{30 \times 30, v}^n)^T \mathbf{Y}_2 = \mathbf{Q}_1 \mathbf{R}_1 \quad (10)$$

其中, $\mathbf{Y}_1 = \tilde{\mathbf{L}}_{30 \times 30, v}^n \mathbf{G}_1$ 、 $\mathbf{Y}_2 = (\tilde{\mathbf{L}}_{30 \times 30, v}^n)^T \mathbf{G}_2$ 是各通道逼近低秩矩阵的双边随机投影。式(9)表示双边随机投影下的低秩矩阵近似, 式(10)表示 \mathbf{Y}_1 、 \mathbf{Y}_2 的 QR 分解过程, \mathbf{G}_1 、 \mathbf{G}_2 表示随机高斯矩阵。采用 $\tilde{\mathbf{L}}$ 而不是 \mathbf{L} 来进行双边随机投影, 是为了确保奇异值不发生畸变。 q 取值越大, 系统的精度越高, 但是计算量也随之增加, 这里设置 $q=1$ 。

步骤 4 比较重构矩阵秩。若各个通道矩阵秩均小于预设值 r ，则令其等于矩阵 $\mathbf{G}_2^T \mathbf{Y}_1$ 的秩，执行步骤 2；否则，执行步骤 5。

步骤 5 更新各个通道的低秩矩阵。

$$\mathbf{L}_{30 \times 30, v}^n(t) = \mathbf{Q}_1 [\mathbf{R}_1 (\mathbf{G}_2^T \mathbf{Y}_1)^{-1} \mathbf{R}_2^T]^{2q+1} \mathbf{Y}_2^T, \\ \mathbf{B}_{30 \times 30, v}^n(t) = \mathbf{P}_\Omega (\mathbf{E}_{30 \times 30, v}^n - \mathbf{L}_{30 \times 30, v}^n(t)) \quad (11)$$

其中， \mathbf{P}_Ω 表示将一个矩阵投影到元素集合 Ω ，即只保留 $\mathbf{E}_{30 \times 30, v}^n - \mathbf{L}_{30 \times 30, v}^n(t)$ 前 b 个元素，转至步骤 2。

步骤 6 输出各个通道低秩矩阵。如果重构矩阵误差小于 ε ，输出迭代后的各通道低秩矩阵组合 $\mathbf{L}_{30 \times 30 \times 3}^n = [\mathbf{L}_{30 \times 30, 1}^n(t); \mathbf{L}_{30 \times 30, 2}^n(t); \mathbf{L}_{30 \times 30, 3}^n(t)]$ ，并准备第 $n+1$ 轮计算；否则，执行步骤 3。

最终输出的低秩矩阵序列为

$$\mathbf{L}_{30 \times 30 \times 3 \times N}^n = [\mathbf{L}_{30 \times 30 \times 3}^1, \dots, \mathbf{L}_{30 \times 30 \times 3}^n, \dots, \mathbf{L}_{30 \times 30 \times 3}^N] \quad (12)$$

本文方法通过计算矩阵的 F 范数来近似矩阵低秩的约束。秩越低，信号矩阵的线性相关度越高，表明越多的天线子载波表现出相同的变化规律。本文设置最终输出的低秩矩阵约束 $r=1$ ，目的是希望从噪声湮没的 CSI 信号中恢复每根天线 30 个子载波的统一变化规律。同时，由于背景干扰随机影响每根天线的每个子载波，符合误差稀疏性的约束，因此设置背景矩阵稀疏度 $b=100$ ，误差 $\varepsilon=10^{-6}$ ，以控制计算复杂度。因此，最终分离出的低秩矩阵 \mathbf{L} 包含了最丰富的运动信息和最少的噪声干扰。图 4 为经过 RPCA 处理后不同动作的连续信号低秩图，纵轴从上到下分别表示第 1、2 和 3 根天线，横轴从左到右分别表示第 1、3、5、7 和 9 帧。通过观察图 4(a) 的第 7、9 帧和图 4(b) 的第 5、7 帧的实验结果可以发现，RPCA 提取的信号低秩成分能够从不同的空间维度上清晰地揭露运动信号的变化特

征，具有显著的辨别力；通过观察图 4(a) 的第 1、3、5 帧和图 4(b) 的第 1、3 帧可以发现，RPCA 有效抑制了背景因素和测量噪声的干扰，一定程度上提升了信号帧的分辨能力。

4 连续信号图像时空域特征提取与分类

4.1 空间特征提取

经过预处理后的连续信号序列 \mathbf{I}' 需要按照从 1 到 N 的顺序，依次将其中的三维低秩矩阵 $\mathbf{L}_{30 \times 30 \times 3}^n$ 输入 CNN 进行空间特征提取。CNN 优势在于：1) CNN 可以通过多层的卷积运算逐层自动提取抽象特征，而人体活动天然具有分层结构^[9]；2) CNN 的局部感知特性有利于局部空间域信息的获取，而人体活动往往发生在特定的时间片段内；3) CNN 不需要复杂经验来指导参数设置，就可以提供位移、尺度和旋转不变的空间特征，提高数据处理的容错率。对于一个 j 层的 CNN 模型，第 m 层的卷积运算式可以表示为

$$\mathbf{Y}^{(m)} = \mathbf{g}^{(m)}(\mathbf{M}^{(m)} \mathbf{X}^{(m)} + \boldsymbol{\delta}^{(m)}) \quad (13)$$

其中， $\mathbf{Y}^{(m)}$ 和 $\mathbf{X}^{(m)}$ 分别为第 m 个卷积层的输出和输入，第 m 个卷积层输出即为第 $m+1$ 个卷积层的输入，即 $\mathbf{X}^{(m)} = \mathbf{Y}^{(m-1)}$ 。 $\mathbf{g}^{(m)}$ 、 $\mathbf{M}^{(m)}$ 和 $\boldsymbol{\delta}^{(m)}$ 分别表示第 m 个卷积层的激活函数、权重矩阵和偏置矩阵。特别地，对于第一个卷积层来说，输入为经 RPCA 处理后的第 n 个低秩矩阵，即 $\mathbf{X}^{(1)} = \mathbf{L}_{30 \times 30 \times 3}^n$ ；对于最后一个卷积层，输出 $\mathbf{Y}^{(j)} = \mathbf{Y}$ 。如图 5 所示，本文搭建了一种适用于低分辨率信号帧的空间特征提取模型。借鉴文献[27]的设计思路，仅采用 3 层卷积运算以提取场景无关且具有辨别力的空间特征，防止模型过深所导致的特征过拟合。每个卷积层 (Conv) 后都做批归一化和 ReLu 激活处理，以提高模型的稳健性和训练速度。每次运算均采用较小的 3×3 卷积核压缩模型的

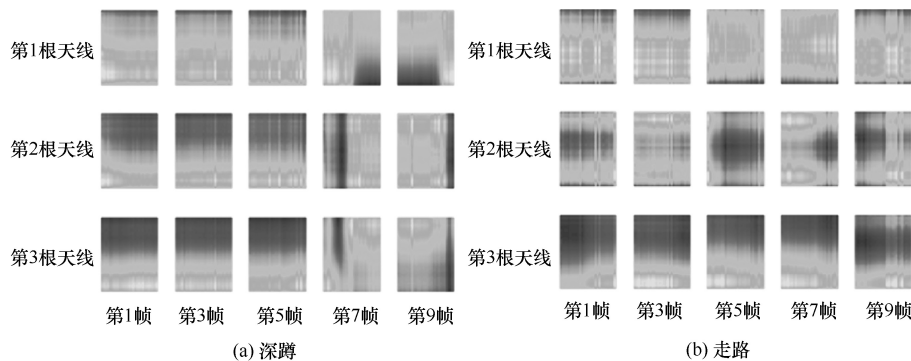


图 4 经过 RPCA 处理后不同动作的连续信号低秩图

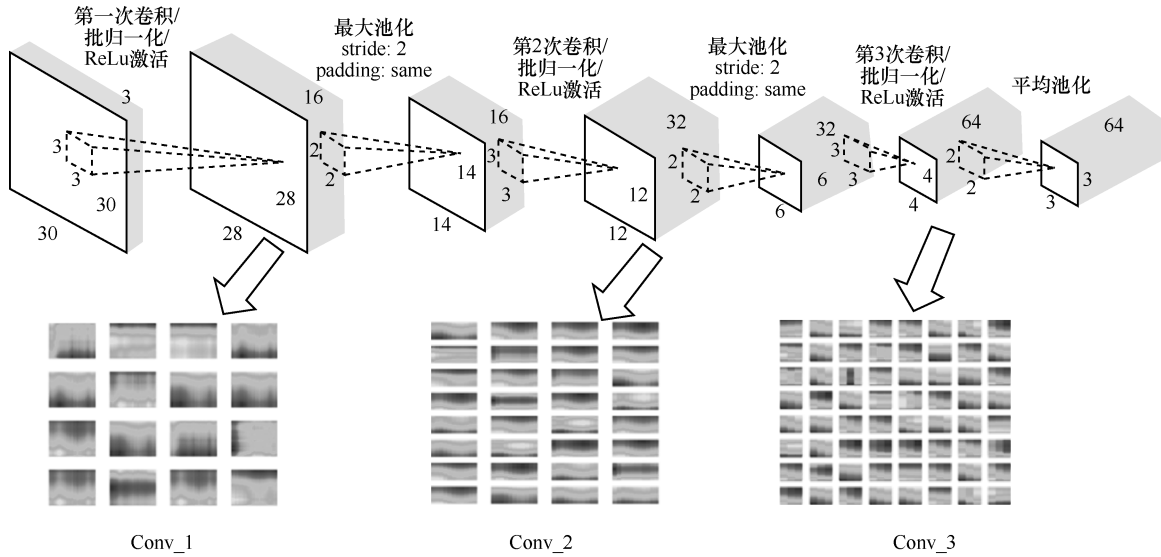


图 5 空间特征提取模型结构

参数量，并在池化操作后翻倍特征通道数，从而保证特征提取的完整性。最大池化层负责对前一轮的卷积特征图进行二次采样，以缩减模型大小；最后一个全局平均层负责提取每个通道全局信息，并输出维度为 $3 \times 3 \times 64$ 的张量至后续的 LSTM 层中。第三个卷积层输出的特征图序列 F 为

$$F = [Y_1^{(3)}, \dots, Y_n^{(3)}, \dots, Y_N^{(3)}] \quad (14)$$

当模型设置默认参数值的情况下，以 WiAR 数据集中走路样本的第 7 帧为例，可以看出所构建的空间特征提取模型可以从多个维度逐层表征具有辨别力的空间特征。

4.2 时域特征提取

由于动作执行存在前后逻辑关系，单一的 CNN 模型无法建模时间尺度上的依赖关系。本文添加 LSTM 层，可以建立连续卷积特征图 F 中的时域关联，提升识别系统的可靠性。原因在于，LSTM 可以通过控制内部的状态，实现记忆单元的存储和更新。具体而言，每个 LSTM 内部主要包括输入门 i_n 、输出门 o_n 和遗忘门 f_n 。输入门 i_n 决定新信息流入当前记忆单元状态 c_n 的数量（式(15)），输出门 o_n 负责控制前一帧记忆单元 c_{n-1} 的信息输出（式(16)），遗忘门 f_n 负责判断 c_{n-1} 中何种信息应当被保留至当前时刻，以防止梯度消失或者梯度爆炸（式(17)）。LSTM 结构如图 6 所示，根据第 n 个信号帧获得输入卷积特征图 $Y_n^{(3)}$ 、前一帧的记忆单元状态 c_{n-1} ，以及上一时刻的信息输出 h_{n-1} ，得到当前时刻的记忆单元状态 c_n （式(18)、式(19)）和输出 h_n （式(20)）。

$$i_n = \sigma(W_i[h_{n-1}, Y_n^{(3)}] + b_i) \quad (15)$$

$$o_n = \sigma(W_o[h_{n-1}, Y_n^{(3)}] + b_o) \quad (16)$$

$$f_n = \sigma(W_f[h_{n-1}, Y_n^{(3)}] + b_f) \quad (17)$$

$$\tilde{c}_n = \tanh(W_c[h_{n-1}, Y_n^{(3)}] + b_c) \quad (18)$$

$$c_n = f_n \circ c_{n-1} + i_n \circ \tilde{c}_n \quad (19)$$

$$h_n = o_n \circ \tanh(c_n) \quad (20)$$

其中， W_i 、 W_o 、 W_f 、 W_c 和 b_i 、 b_o 、 b_f 、 b_c 分别为相应的权重和偏置矩阵， σ 表示激活函数 sigmoid， \circ 表示 2 个矩阵的逐元素相乘运算。

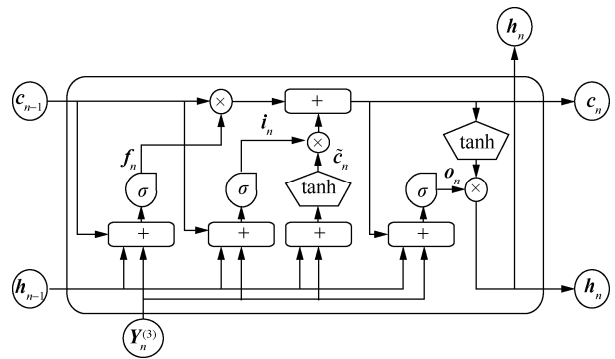


图 6 LSTM 结构

在实际操作中，需要预先在 CNN-LSTM 中分别添加“序列折叠层”和“序列展开层”，以便实现图像序列批次和图像批次之间的转换。当最后一个卷积层输出特征图为 $3 \times 3 \times 64$ 的张量时，需要经过 flatten 层的处理，转换成 576×1 的向量后再输入两层 LSTM 单元中。LSTM 的单元数为 128，两层

LSTM 单元之间 dropout 概率设置为 0.5, 以应对过拟合问题。将最后一次迭代的 LSTM 输出信息送入一个全连接层 (fc), 隐单元数量对应动作的种类。最后, 由 Softmax 层输出动作的最大概率, 并预测对应的标签。本文采用的模型训练优化器为动量梯度下降法 (SGDM, stochastic gradient descent with momentum), 每次随机选择小批量进行梯度下降, 从而实现权值的快速迭代更新和损失函数的最小化。选择 L2 正则化作为损失函数的惩罚项, 目的是控制模型复杂度, 防止模型过拟合。相应的动量超参数默认为 0.9, 轮次上限设置为 50, 每次更新所需的小批量大小为 28。初始学习率设为 0.01, 当迭代至 25 轮以后, 再将学习率调整为 0.001, 以期用更小的步长寻找最优值。本文采用 5 折交叉运算法来验证算法的有效性, 实验训练和测试均部署在 Intel i7-5700HQ 2.70 GHz CPU 的台式机, 并利用 MATLAB 2019a 进行模型实现, 训练过程如图 7 所示。

5 实验性能分析

5.1 实验数据集及实验设置

本文所采用的公开数据集 WiAR^[25]是 Wi-Fi 动作识别领域最新的数据集, 包含三类室内场景、16 个动作类别, 涵盖了 4 800 个采集数据。该数据集

的动作类别包括 10 个上肢运动、2 个下肢运动和 4 个全身运动。每个动作均由 10 名测试人员 (5 名男性、5 名女性) 按照指令执行 30 次, 持续时间超过 7 s, 包括动作准备时间、执行时间和完成后等待时间。配置 Intel 5300 网卡的商用 Wi-Fi 设备以 30 Hz 的采样率持续接收 CSI 数据分组, 接收端的天线数量为 3 根。由于该数据集来源于真实场景, 包含场景噪声、习惯差异、人员多样性等诸多不利因素, 故具有一定的挑战性。

本文还使用了自主采集的真实场景数据进行测试。参考 WiAR^[25]的实验设置, 实验安排在如图 8 所示的 6 m×6 m 相对空旷的会议室和 6 m×12 m 噪声复杂的办公室环境中进行。选择 5 名无相关专业背景的研究生作为测试人员全程参与实验, 其中, 2 名为女性, 3 名为男性。为了增加动作识别的难度, 挑选 4 个相似的全身动作和 4 个上肢运动, 全身动作包括原地踏步、深蹲、跳跃和摔倒, 上肢动作包括喝水、接听电话、挥手和敬礼。在一周内每人每天在不同场景分 2 次完成共计 20 组动作, 以避免劳累所导致的动作变形。数据采集阶段, 提前为两台 ThinkPad 笔记本电脑配置 Intel 5300 网卡和 CSITool^[11], 其中一台装配一根外置天线, 作为发射端 (AP); 另一台装备 3 根外置天线, 作为接收端 (MP)。两台笔记本放置距离为 4 m, 放置高度

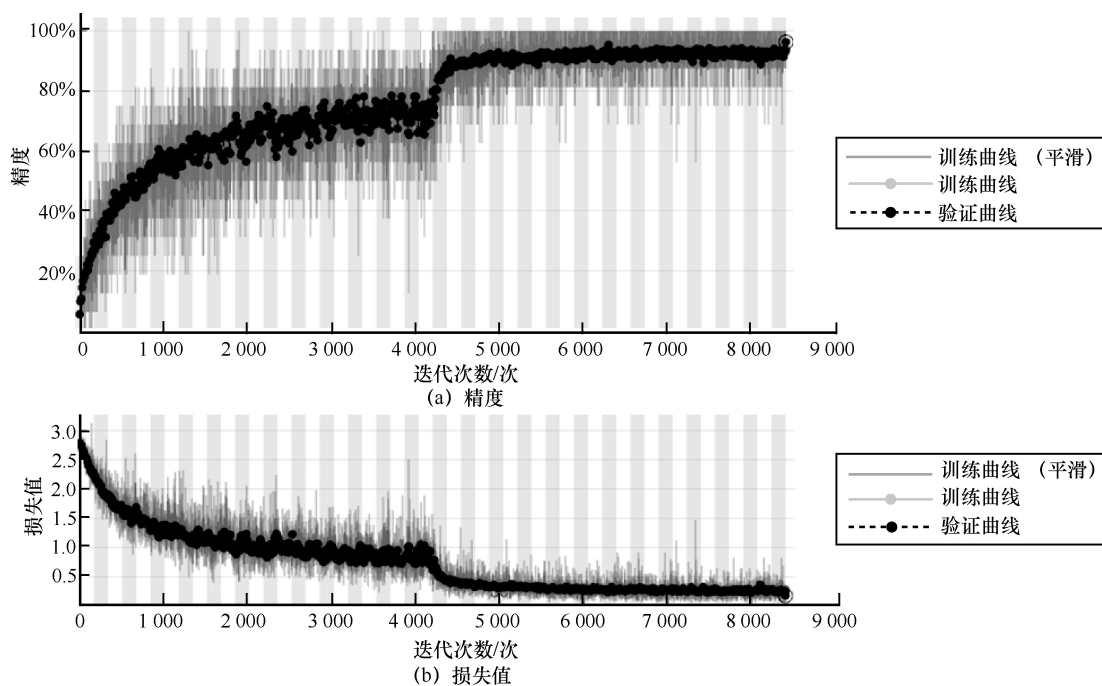


图 7 MATLAB 平台模型训练过程

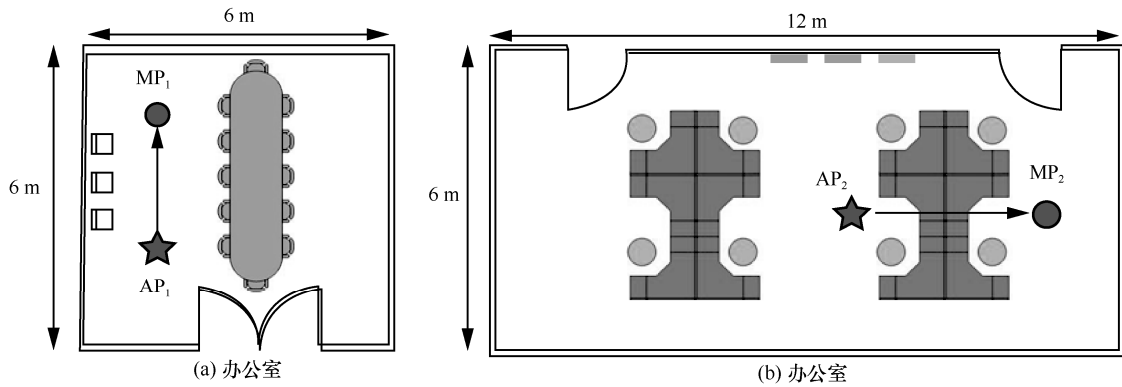


图 8 本文设置的实验环境

为 1.5 m，分别工作在监控模式下，以 30 Hz 的采样率持续发送和接收数据分组。

5.2 基于 CNN-LSTM 的同类工作对比实验

为了准确评估 SIL-Fi 的模型优势，本文选取 2 种基于 CNN-LSTM 模型的同类工作进行对比实验。

如表 1 所示，文献[28]采用了低通滤波、PCA 去噪和有效区域的组合去噪方式，以 0.96 的精度对 8 种活动进行识别；文献[29]前期利用盲源信号分解法 (BSS, blind signal separation) 对原始的幅值和相位信息进行分解，并利用隐马尔可夫模型 (HMM, hidden Markov model) 对识别结果进行修正，最终以 0.85 的精度识别 4 种门窗开关的方式。本文方法在模型主体结构基本类似的情况下，在信号预处理阶段仅利用 RPCA 对运动信号主成分进行提取，避免引入烦琐的信号表征步骤 (文献[28]) 或是额外信号参数及纠正手段 (文献[29])，并在公开数据集

WiAR 和自主采集数据集上均取得理想的识别精度。为了进一步阐述本文方法的优势，复现了表 1 中同类工作的方法，采用 5 折交叉验证法分别对 WiAR 和自主采集数据集的各类活动数据进行比较。其中，表 2 为 WiAR 数据集上 16 类活动的识别正确率，表 3 所示为自主采集数据集上 8 类相似活动的识别正确率。综合表 2 和表 3 的实验结果，文献[29]方法在识别弯腰、坐下、深蹲等全身动作时效果较好，识别喝水、打电话等相似动作时效果较差，这可能来源于 BSS 算法的局限性，即来自人体躯干和门窗等表面积较大的反射信号更有利于分离。文献[28]在 2 个数据集上的平均识别精度最低，分别约为 0.91 和 0.94，这可能是因为该方法要求天线远距离摆放以获取空间多样性，而在本文实验场景中，烦琐的信号预处理方法在处理噪声干扰的同时，也更容易丢弃有益的运动信息。本文方法

表 1 本文方法与基于 CNN-LSTM 的同类工作的比较

方法	信号预处理	模型输入	模型主体结构	活动种类	精度
文献[28]	低通滤波+PCA+有效区域选择	幅值 STFT 图	一层 Conv+2 层 LSTM+一层 fc+Softmax	8 种	0.96
文献[29]	BSS	原始+分解后幅值/相位	3 层 Conv+2 层 LSTM+Softmax+HMM	4 种	0.85
SIL-Fi	RPCA	幅值连续帧	3 层 Conv+2 层 LSTM+一层 fc+Softmax	16 种(WiAR) 8 种(自主采集)	0.94 0.96

表 2 WiAR 数据集上识别正确率比较

方法	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	平均精度
文献[28]	0.92	0.89	0.84	0.94	0.96	0.85	0.92	0.90	0.96	0.93	0.86	0.95	0.90	0.85	0.95	0.97	0.911 9
文献[29]	0.90	0.94	0.93	0.93	0.92	0.83	0.92	0.88	0.97	0.94	0.93	0.92	0.89	0.85	0.98	0.98	0.919 4
SIL-Fi	0.92	0.98	0.93	0.96	0.94	0.85	0.94	0.86	0.99	0.97	0.99	0.94	0.90	0.99	1.00	0.93	0.942 7

注：1-水平挥手，2-单手挥动，3-两手挥动，4-高抛，5-手臂画叉，6-手臂画勾，7-摇晃纸张，8-前向踢腿，9-侧向踢腿，10-弯腰，11-鼓掌，12-走路，13-打电话，14-喝水，15-坐下，16-深蹲。

表 3 自主采集数据集上识别正确率比较

方法	1	2	3	4	5	6	7	8	平均精度
文献[28]	0.92	0.95	0.95	0.96	0.96	0.91	0.93	0.92	0.938 3
文献[29]	0.94	0.95	0.95	0.96	0.92	0.92	0.94	0.94	0.940 6
SIL-Fi	0.95	0.95	0.97	0.97	0.95	0.95	0.96	0.94	0.956 4

注：1-原地踏步，2-深蹲，3-跳跃，4-摔倒，5-喝水，6-接电话，7-挥手，8-敬礼。

在面对不同动作时均能取得较好的识别效果，平均精度分别为 0.94 和 0.96，说明本文模型在仅需要 RPCA 和原始幅值连续帧的情况下，依然能够有效提取运动信号主成分；在不进行深度模型参数复杂调优的情况下，依然能够提取显著性特征。

5.3 不同动作对系统识别准确率的影响

为了进一步探究不同动作对本文方法识别准确率的影响，以混淆矩阵的形式对本文方法在 2 个数据集的验证结果进行呈现，如图 9 所示，纵坐标为真实标签，横坐标为预测标签，每个混淆矩阵的像素 (i, j) 表示第 i 个预测动作被估计成第 j 个真实动作的个数。图 9(a)中动作编号 1~16 所代表的动作如表 2 注所示，图 9(b)中动作编号 1~8 所代表的动作如表 3 注所示。如图 9(a)所示，在 WiAR 数据集上的主要误差来源于手臂画勾和前向踢腿等肢体动作。其中，手臂画勾动作最容易被识别成手臂画叉和摇晃纸张等具有相似逻辑关系的动作，分别有 15 例和 18 例；前向踢腿动作最容易被识别成相似的侧向踢腿和走路，分别有 12 例和 24 例。在图 9(b)中，由于全身运动和肢体运动具有显著差异，因此对应的全身（肢体）动作全部被识别成类似的全身（肢体动作）。在全身动作中，深蹲动作是最容易被误识别的，仅有 0.949 的识别正确率；在肢体动作中，喝水和敬礼动作的识别精度最低，分别为 0.939 和 0.933。通过总结实验结果观察到，连续且具有较大摆动幅度的全身动作最容易被识别，而孤立的、具有相似逻辑关系且摆动幅度较小的肢体动作最难被正确分类。

5.4 不同深度模型组合方式的影响

本节实验分析了不同卷积（Conv）层和 LSTM 层组合方式对训练时间和平均精度的影响，以 WiAR 数据集为例，实验结果如表 4 所示。其中，设置每个 Conv 层后紧跟最大池化层、批归一化层和激活层 ReLu，其一层、2 层和 3 层 Conv 的隐神经元个数分别是(64)、(32, 64)和(16, 32, 64)，LSTM 的单元数统一设置为 128，全连接层的单元数为 16，

其余设置与前文所述一致。观察到，当以序号 1 的方式组合时，耗时最长，平均精度最低。这是由于缺乏 Conv 层的抽象特征提取和最大池化层的降采样操作，导致 LSTM 需要处理多达 14 400 个信号批次。随着 Conv 层数和最大池化层数的增加，训练时间逐渐减小，平均精度也得到了显著的提升，说明网络开始逐层学习具有判别性语义的特征。当固定 Conv 层数，进一步增加 LSTM 层数时，观察到从序号 2 到序号 3，由于模型过度拟合，尽管训练时间增加了 299 s，但是精度却反而下滑。相比序号 4，当采用序号 5 的组合方式时，仅需要多训练 1 464 s，即每个训练样本训练 1.75 s，就可以得到超过序号 4 组合方式 0.12 的精度提升。

5.5 不同特征提取方式对识别稳健性的影响

为了直观凸显 SIL-Fi 在特征提取方面的优势，本文分别选取基于手动特征提取^[14,18]、基于 CNN^[22]和基于 LSTM^[30]的相关工作进行复现并进行对比实验。基于手动特征提取的工作一般分为两类：一类运用专家经验，试图枚举时频域显著性特征，从而构建完备的特征集合，如文献[14]；另一类是在前者的基础上，对特征集合的有效性和冗余性进行计算，从而筛选出代表性特征，如文献[18]。为确保对比实验的公平性，在复现文献[18]的方法时，所采用的特征均来源于文献[14]，二者仅在特征筛选环节有所不同。对比文献[22,30]是为了说明单一的空间/时域特征对于系统稳健性的影响。文献[22]利用改进的 CNN 网络提取变尺寸的幅值-相位图像的空间特征，文献[30]则首次将 LSTM 应用于 Wi-Fi 活动信号的时域关系建模。

由于 WiAR 数据集缺乏样本采集时的环境描述，因此在自主采集数据集上对不同特征提取方式的稳健性进行探究。自主采集数据集上不同特征提取方法对系统稳健性的影响如表 5 所示。可以得出以下结论。1) 文献[14]的枚举特征集合无法在自主采集数据集上使用，该特征集合仅能片面地刻画自主采集数据集活动的特征，因此平均精度最低，仅

动作																																															
真实标签	1	276	3					6									92.0%	8.0%																													
	2		294	6													98.0%	2.0%																													
	3		3	279						9	9						93.0%	7.0%																													
	4		3		285		3	6									96.0%	4.0%																													
	5					282			6	6	6						94.0%	6.0%																													
	6				6	15	255		6								85.0%	15.0%																													
	7						9	282	3	6							94.0%	6.0%																													
	8								258	12	6		24				86.0%	14.0%																													
	9								3	297							99.0%	1.0%																													
	10										291						97.0%	3.0%																													
	11				3							297					99.0%	1.0%																													
	12								9	3	6		282				94.0%	6.0%																													
	13						12	15						270	3		90.0%	10.0%																													
	14				2										298		99.3%	0.7%																													
	15															299	1	99.7%	0.3%																												
	16										6					14	280	93.3%	6.7%																												
<table border="1"> <tr> <td>100.0%</td><td>97.0%</td><td>97.9%</td><td>96.3%</td><td>94.9%</td><td>91.4%</td><td>87.9%</td><td>90.5%</td><td>91.7%</td><td>85.8%</td><td>92.5%</td><td>92.5%</td><td>100.0%</td><td>99.0%</td><td>95.5%</td><td>99.6%</td> </tr> <tr> <td></td><td>3.0%</td><td>2.1%</td><td>3.7%</td><td>5.1%</td><td>8.6%</td><td>12.1%</td><td>9.5%</td><td>8.3%</td><td>14.2%</td><td>7.5%</td><td>7.8%</td><td></td><td>1.0%</td><td>4.5%</td><td>0.4%</td> </tr> </table>																100.0%	97.0%	97.9%	96.3%	94.9%	91.4%	87.9%	90.5%	91.7%	85.8%	92.5%	92.5%	100.0%	99.0%	95.5%	99.6%		3.0%	2.1%	3.7%	5.1%	8.6%	12.1%	9.5%	8.3%	14.2%	7.5%	7.8%		1.0%	4.5%	0.4%
100.0%	97.0%	97.9%	96.3%	94.9%	91.4%	87.9%	90.5%	91.7%	85.8%	92.5%	92.5%	100.0%	99.0%	95.5%	99.6%																																
	3.0%	2.1%	3.7%	5.1%	8.6%	12.1%	9.5%	8.3%	14.2%	7.5%	7.8%		1.0%	4.5%	0.4%																																
动作	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16																															
预测标签																																															

(a) WiAR数据集

动作																															
真实标签	1	676	24	10													95.2%	4.8%													
	2	14	664	12	10												94.9%	5.1%													
	3		13	679	8												97.0%	3.0%													
	4	3	5	10	682												97.4%	2.6%													
	5					2		667	13	2	18						95.0%	5.0%													
	6							8	662	16	14						94.6%	5.4%													
	7									13	672	15					96.0%	4.0%													
	8									11	17	14	658				94.0%	6.0%													
<table border="1"> <tr> <td>97.5%</td><td>94.1%</td><td>95.2%</td><td>97.4%</td><td>97.2%</td><td>93.9%</td><td>95.5%</td><td>93.3%</td> </tr> <tr> <td>2.5%</td><td>5.9%</td><td>4.8%</td><td>2.6%</td><td>2.8%</td><td>6.1%</td><td>4.5%</td><td>6.7%</td> </tr> </table>																97.5%	94.1%	95.2%	97.4%	97.2%	93.9%	95.5%	93.3%	2.5%	5.9%	4.8%	2.6%	2.8%	6.1%	4.5%	6.7%
97.5%	94.1%	95.2%	97.4%	97.2%	93.9%	95.5%	93.3%																								
2.5%	5.9%	4.8%	2.6%	2.8%	6.1%	4.5%	6.7%																								
动作	1	2	3	4	5	6	7	8																							
预测标签																															

(b) 自主采集数据集

图 9 2 个数据集上不同动作的混淆矩阵

为 0.77。2) 采用有监督式的特征筛选方法^[18]通过剔除冗余的、无益的时频特征, 可以将系统识别精度提高至 0.85, 这说明在理想情况下, 若能针对领域特点设计出一个考虑足够全面、精炼和显著的特征集合, 则可以确保系统识别的稳健性。然而这种方法极大地依赖设计者的背景知识、调参经验和总结规律能力。3) 即使是单一的 CNN 或者 LSTM 方法, 在面对场景迁移和个体差异时, 都能通过自动提取空间域/时间域上的显著特征, 取得理想的识别精度, 分别为 0.935 和 0.934。4) SIL-Fi

通过深度融合 CNN 和 LSTM 模型, 在不同的实验室环境、不同的受试个体中均取得了最高的实验精度, 验证了本文方法的稳健性。注意到测试人员的个体差异对识别精度的影响, SIL-Fi 在经过适当修改后还适用于用户认证、手势密码和智能推荐等个性化服务。

6 结束语

针对现有基于深度学习的 Wi-Fi 行为识别方法存在的抗噪声能力差、信号输入不兼容以及特征提取不充分的问题, 本文提出一种基于连续图像深度

表 4 WiAR 数据集上不同深度模型组合方式对实验结果的影响

序号	深度模型组合方式	WiAR 数据集	
		训练时间/s	平均精度
1	一层 Conv+一层 LSTM+一层 fc+Softmax	18 846	0.54
2	2 层 Conv+一层 LSTM+一层 fc+Softmax	10 108	0.77
3	2 层 Conv+2 层 LSTM+一层 fc+Softmax	10 407	0.71
4	3 层 Conv+一层 LSTM+一层 fc+Softmax	4 923	0.82
5	3 层 Conv+2 层 LSTM+一层 fc+Softmax	6 387	0.94

表 5 自主采集数据集上不同特征提取方法对系统稳健性的影响

测试人员序号		会议室					办公室				
		女性 1	女性 2	男性 1	男性 2	男性 3	女性 1	女性 2	男性 1	男性 2	男性 3
手动特征提取	完备特征 ^[14]	0.78	0.80	0.81	0.77	0.77	0.72	0.75	0.81	0.76	0.80
	特征挑选 ^[18]	0.85	0.88	0.87	0.84	0.84	0.83	0.86	0.86	0.81	0.83
自动特征提取	CNN ^[22]	0.93	0.97	0.97	0.94	0.93	0.89	0.92	0.95	0.93	0.92
	LSTM ^[30]	0.92	0.95	0.94	0.95	0.94	0.91	0.93	0.93	0.94	0.93
	SIL-Fi	0.95	0.97	0.99	0.95	0.94	0.94	0.96	0.97	0.95	0.93

学习的 Wi-Fi 人体行为识别方法——SIL-Fi。首先将采集到的 CSI 原始幅值序列分割成若干个固定尺寸帧图，克服模型输入限制；进而设计低秩分解算法提取关键的运动主成分信息；并将每一帧低秩矩阵分别输入空间特征模块和时间特征提取模块中。其中，空间特征模块通过逐层表征信号矩阵，将低秩矩阵中具有显著识别力的空间尺度特征抽象并映射到特征图中；时间特征模块接纳变长的特征图序列，并借助其记忆特性对其在时间尺度的上下文依赖关系进行建模。通过整合模型，既能实现低分辨率图像空间域和时间域特征的深度萃取，又可以有效解决 CNN 模型输入和特征提取不足的问题。本文方法分别在 WiAR 公开数据集和自主采集数据集上进行测试，并考虑了不同动作、模型组合、测试人员、特征提取方式等诸多因素的影响。实验结果表明，本文方法相较于同类工作具有良好的性能，具备普适场景下的高精度和稳健性。

为进一步提高方法的实用性，下一步工作还需要解决以下问题：1) 如何在少样本或者零样本的情况下，实现场景独立的人体行为识别；2) 如何在多人或多变量并发的情况，实现目标人体活动信号的分离和提取；3) 如何通过合理修改模型结构，在计算资源有限的情况下进一步提高工作效率。

参考文献：

[1] YANG Q. Activity recognition: linking low-level sensors to high-level

intelligence[C]//Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence. New York: ACM Press, 2009: 20-25.

[2] 罗会兰, 王婵娟, 卢飞. 视频行为识别综述[J]. 通信学报, 2018, 39(6): 169-180.

LUO H L, WANG C J, LU F. Survey of video behavior recognition [J]. Journal on Communications, 2018, 39(6): 169-180.

[3] 王钰翔, 李晟洁, 王皓, 等. 基于 Wi-Fi 的非接触式行为识别研究综述[J]. 浙江大学学报(工学版), 2017, 51(4): 648-654.

WANG Y X, LI S J, WANG H, et al. Survey on Wi-Fi based contactless activity recognition[J]. Journal of Zhejiang University (Engineering Science), 2017, 51(4): 648-654.

[4] LARA O D, LABRADOR M A. A survey on human activity recognition using wearable sensors[J]. IEEE Communications Surveys & Tutorials, 2012, 15(3): 1192-1209.

[5] 殷晓玲, 陈晓江, 夏启寿, 等. 基于智能手机内置传感器的人体运动状态识别[J]. 通信学报, 2019, 40(3): 157-169.

YIN X L, CHEN X J, XIA Q S, et al. Human motion state recognition based on smart phone built-in sensor[J]. Journal on Communications, 2019, 40(3): 157-169.

[6] HERATH S, HARANDI M, PORIKLI F. Going deeper into action recognition: a survey[J]. Image and Vision Computing, 2017, 60: 4-21.

[7] BAHL P, PADMANABHAN V N, BAHL V, et al. RADAR: an in-building RF-based user location and tracking system[C]//The 19th Annual Joint Conferences of the IEEE Computer and Communications Societies. Piscataway: IEEE Press, 2000: 775-784.

[8] CISCO. Global mobile data traffic forecast update, 2016-2021 white paper[R]. Cisco, (2017-02-07)[2020-02-26].

[9] 鲁勇, 吕绍和, 王晓东, 等. 基于 Wi-Fi 信号的人体行为感知技术研究综述[J]. 计算机学报, 2019, 42(2): 3-23.

LU Y, LYU S H, WANG X D, et al. A survey on Wi-Fi based human behavior analysis technology[J]. Chinese Journal of Computers, 2019, 42(2): 3-23.

- [10] SEIFELDIN M, SAEED A, KOSBA A E, et al. Nuzzer: a large-scale device-free passive localization system for wireless environments[J]. IEEE Transactions on Mobile Computing, 2012, 12(7): 1321-1334.
- [11] HALPERIN D, HU W, SHETH A, et al. Tool release: gathering 802.11 n traces with channel state information[J]. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 53-53.
- [12] 党小超, 黄亚宁, 郝占军, 等. 基于信道状态信息的无源室内人员日常行为检测方法[J]. 通信学报, 2019, 40(4): 160-170.
DANG X C, HUANG Y N, HAO Z J, et al. Passive indoor human daily behavior detection method based on channel state information[J]. Journal on Communications, 2019, 40(4): 160-170.
- [13] PALIPANA S, ROJAS D, AGRAWAL P, et al. FallDeFi: ubiquitous fall detection using commodity Wi-Fi devices[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 1(4): 1-25.
- [14] ZENG Y, PATHAK P H, MOHAPATRA P. WiWho: Wi-Fi-based person identification in smart spaces[C]//2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks. Piscataway: IEEE Press, 2016: 1-12.
- [15] TAN S, YANG J. WiFinger: leveraging commodity Wi-Fi for fine-grained finger gesture recognition[C]//Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computin. New York: ACM Press, 2016: 201-210.
- [16] WANG W, LIU A X, SHAHZAD M, et al. Understanding and modeling of Wi-Fi signal based human activity recognition[C]//Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2015: 65-76.
- [17] YUE S, HE H, WANG H, et al. Extracting multi-person respiration from entangled RF signals[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(2): 86.
- [18] ZOU H, ZHOU Y, YANG J, et al. Device-free occupancy detection and crowd counting in smart buildings with Wi-Fi-enabled IoT[J]. Energy and Buildings, 2018, 174: 309-322.
- [19] MA Y, ZHOU G, WANG S. Wi-Fi sensing with channel state information: a survey[J]. ACM Computing Surveys (CSUR), 2019, 52(3): 1-36.
- [20] 周鑫, 何晓新, 郑昌文. 基于图像深度学习的无线电信号识别[J]. 通信学报, 2019, 40(7): 114-125.
ZHOU X, HE X X, ZHENG C W. Radio signal recognition based on image deep learning[J]. Journal on Communications, 2019, 40(7): 114-125.
- [21] MA Y, ZHOU G, WANG S, et al. SignFi: sign language recognition using Wi-Fi[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 1-21.
- [22] ZHOU Q, XING J, CHEN W, et al. From signal to image: enabling fine-grained gesture recognition with commercial Wi-Fi devices[J]. Sensors, 2018, 18(9): 3142-3163.
- [23] ZHOU T, TAO D. Godec: randomized low-rank & sparse matrix decomposition in noisy case[C]//Proceedings of the 28th International Conference on Machine Learning. [S.n.: s.l.], 2011: 33-40.
- [24] WANG X, GAO L, SONG J, et al. Beyond frame-level CNN: saliency-aware 3-D CNN with LSTM for video action recognition[J]. IEEE Signal Processing Letters, 2016, 24(4): 510-514.
- [25] GUO L, WANG L, LIN C, et al. WiAR: a public dataset for Wi-Fi-based activity recognition[J]. IEEE Access, 2019, 7(1): 154935-154945.
- [26] QIAN K, WU C, ZHANG Y, et al. Widar2. 0: passive human tracking with a single Wi-Fi link[C]//Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services. New York: ACM Press, 2018: 350-361.
- [27] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks?[J]. In Advances in Neural Information Processing Systems, 2014, 1(1): 3320-3328.
- [28] WANG F, GONG W, LIU J. On spatial diversity in Wi-Fi-based human activity recognition: a deep learning-based approach[J]. IEEE Internet of Things Journal, 2018, 6(2): 2035-2047.
- [29] OHARA K, MAEKAWA T, MATSUSHITA Y. Detecting state changes of indoor everyday objects using Wi-Fi channel state information[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2017, 1(3): 1-28.
- [30] YOUSEFI S, NARUI H, DAYAL S, et al. A survey on behavior recognition using Wi-Fi channel state information[J]. IEEE Communications Magazine, 2017, 55(10): 98-104.

[作者简介]



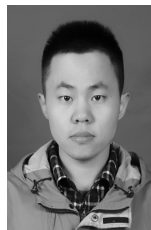
周启臻(1993-), 男, 浙江温州人, 陆军工程大学博士生, 主要研究方向为无线感知和智能建筑。



邢建春(1964-), 男, 河北正定人, 博士, 陆军工程大学教授、博士生导师, 主要研究方向为国防工程智能化和智能建筑。



杨启亮(1975-), 男, 河南信阳人, 博士, 陆军工程大学副教授、硕士生导师, 主要研究方向为软件工程、信息物理融合系统和智能建筑。



韩德帅(1990-), 男, 山东聊城人, 博士, 火箭军工程大学讲师, 主要研究方向为软件工程和国防工程智能化。