

# 基于 LoRa 信号的实时手势识别算法

韩 崇<sup>1,2</sup>, 杜 昊<sup>1</sup>, 郭 剑<sup>1,2</sup>, 孙力娟<sup>1,2</sup>

(1. 南京邮电大学计算机学院, 江苏南京 210023; 2. 江苏省物联网智能感知与计算重点实验室, 江苏南京 210023)

**摘要:** 为了解决在遮挡环境下的实时手势识别问题, 本文提出了一种基于远距离无线电 (Long range Radio, LoRa) 信号的实时手势识别算法. 该算法利用 LoRa 信号频段较低、穿透性较好的特性, 通过两根接收天线计算信号比值, 并结合短时傅里叶变换 (Short-Time Fourier Transform, STFT) 得到包含手部运动特征的时频图. 在此基础上, 设计了 Gesture Encoder 编码器进行手势时频图的特征提取, 从而得到体现手势特征的特征向量, 进而用于手势的分类识别. 本算法不仅有效解决了实际应用中有物体遮挡场景下的识别问题, 还提出了系统状态转换机 (System status Transition Machine, STM) 和数据增强方法, 实现了对手势开始和结束时间的精准控制, 从而完成了手势的自动分割与实时识别. 最终, 在搭载 Android 系统的边缘计算设备上进行了系统部署, 并在遮挡场景下进行测试. 实验结果表明: 所提出的手势识别系统能够在边缘设备上高效、准确地完成手势分类, 具有较强的实用价值和应用前景.

**关键词:** 射频感知; 手势识别; 实时推理; 信号处理; 神经网络

**基金项目:** 国家自然科学基金 (No.62272242)

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112(2025)08-2738-12

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20250096

## Real-Time Gesture Recognition Algorithm Based on LoRa Signals

HAN Chong<sup>1,2</sup>, DU Hao<sup>1</sup>, GUO Jian<sup>1,2</sup>, SUN Li-juan<sup>1,2</sup>

(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China;

2. Jiangsu Key Laboratory of Intelligent Sensing and Computing for Internet of Things, Nanjing, Jiangsu 210023, China)

**Abstract:** To address the real-time gesture recognition problem in occluded environments, this paper proposes a real-time gesture recognition algorithm based on long range radio (LoRa) signals. By utilizing the low frequency band and good penetration of LoRa signals, this algorithm calculates the signal ratio using two receiving antennas, and combines short-time Fourier transform (STFT) to obtain time-frequency maps containing hand motion features. These maps are processed by a neural network encoder called the Gesture Encoder, generating feature vectors that represent the gesture characteristics, which are then used for gesture classification and recognition. This algorithm effectively solves the recognition problem in scenarios with object occlusion, and introduces a system state transition machine (STM) and data augmentation methods to precisely control the start and end times of gestures, thus enabling automatic segmentation and real-time recognition. At last, the system is deployed on an edge computing device running Android, and tested in occlusion scenario. Experimental results show that the proposed gesture recognition system can efficiently and accurately complete gesture classification on edge devices, with strong practical value and application prospects.

**Key words:** radio frequency sensing; gesture recognition; real-time inference; signal processing; neural networks

**Foundation Item(s):** National Natural Science Foundation of China (No.62272242)

## 1 引言

随着智能设备的普及和物联网相关技术的发展, 智能交互系统已成为现代生活中的重要组成部分. 无接触式人机交互 (Human-Computer Interaction, HCI) 技

术不仅能够提升用户体验, 还能够在公共场所、医疗、安防等领域提供更高效、安全的操作方式. 在这些应用场景中, 手势识别作为一种自然直观的交互方式, 具有广泛的应用潜力<sup>[1]</sup>. 目前, 许多工作是基于视觉<sup>[2-4]</sup>和声音<sup>[5]</sup>的识别技术 (如摄像头、红外传感器或超声波).

然而,基于视觉的识别技术易受光照条件和物体遮挡的影响且具有隐私问题.尤其在非视距环境下,基于摄像头视觉的手势识别技术无法工作.此外,超声波技术虽然能够通过声波的反射来识别手势,但由于其传播速度较慢,且受环境噪声的影响较大,因此在实时性和精度方面存在局限,并且很难穿透障碍物<sup>[5]</sup>.因此,近年来,基于射频信号的手势识别技术逐渐成为一种备受关注的替代方案.部分射频信号具有较强的穿透能力,不受环境光照强弱的影响,能够穿透墙壁、木板等障碍物,尤其适合复杂和动态变化的环境.射频技术包括 Wi-Fi、毫米波(millimeter Wave, mmWave)、超宽带(Ultra-WideBand, UWB)、射频识别(Radio-Frequency Identification, RFID)以及调频连续波(Frequency Modulated Continuous Wave, FMCW)雷达等<sup>[6-9]</sup>,它们通过测量信号的时间延迟、频率变化、信号强度等特征,能够获取运动物体的位置信息、速度信息和形状特征,从而实现手势识别.其中, Wi-Fi 和 mmWave 已被广泛应用于手势识别领域,但它们通常依赖大量的多天线阵列和复杂的信号处理算法,计算负担较重,且在精度上仍有一定的提升空间<sup>[6,9]</sup>. UWB 技术由于其高时间分辨率和较短的波长,能够提供精确的定位能力,并且适用于近距离高精度的手势检测,但是 UWB 的抗干扰能力较弱,容易受到环境因素的影响<sup>[7]</sup>.相比之下,基于 FMCW 雷达的手势识别技术具有显著优势, FMCW 雷达能够提供较高的空间分辨率,同时具有较强的穿透能力和抗干扰能力,能够在复杂的环境中稳定工作.尤其在存在遮挡的非视距环境中,其表现尤为出色.然而, FMCW 雷达的造价较为昂贵<sup>[8]</sup>.

尽管 FMCW 雷达在手势识别等感知领域中具有重要的应用潜力,但远距离无线电(Long Range Radio, LoRa)作为一种低功耗、长距离传输的射频通信技术,近年来也在无线感知领域中获得了广泛关注<sup>[10]</sup>. LoRa 可以在不同的频段上运行,包括 433、868 和 915 MHz 等.因其频率较低,所以能够有效穿透塑料板、木墙等障碍物,且不受环境光照等因素的影响<sup>[11]</sup>. LoRa 的低功耗特性使其特别适用于边缘计算设备,能够在有限资源的环境下实现长时间的稳定运行.尽管 LoRa 传统上更多地被应用于远程通信和组网,但近年来其在感知领域的应用也备受关注<sup>[12-17]</sup>.文献[16]将 LoRa 信号用于穿墙粗粒度的行走检测和细粒度的呼吸波形重建.文献[15]提出了 Sen-fence 方法限制仅在特定的范围进行感知,显著提高了传感范围,同时减轻了环境干扰.文献[14]将 LoRa 应用在运动设备下,感知并实现了细粒度呼吸监测和粗粒度行走检测.文献[12]提出的 LoGait 系统实现了人体的步态识别,并且验证了视距、非视距和远距离场景中的感知能力.文献[13]提出了 LoRadar,

可使 FMCW 雷达在传感波中携带 LoRa 信号实现通感一体.文献[17]有效利用了 LoRa 网关上的多天线,结合波束形成技术实现了多目标的感知和同时监测 5 个人的实时呼吸速率.

尽管上述工作中提出了 LoRa 感知的相关理论,但在手势识别方面的研究较为有限.大多数现有的研究主要集中在基于 LoRa 的远程通信、穿墙感知以及呼吸和行走监测等领域<sup>[12-17]</sup>.而 LoRa 在手势识别中的应用尚处于初步阶段,尤其是非视距遮挡下的动态手势检测.因此,本文提出了一种基于 LoRa 信号的实时手势识别系统.与传统的视觉感知方法不同, LoRa 信号结合了 FMCW 射频雷达的优势,不受环境光照变化的影响,适用于较复杂遮挡环境的 HCI. LoRa 信号的频段较低,因此在非视距和复杂遮挡场景中的稳定性和鲁棒性更强,这是 mmWave 等其他商用射频信号难以实现的.此外, LoRa 信号的低功耗特性使其特别适合应用于资源受限的边缘计算设备,并作为一种高效的实时手势识别方案.具体来说,本文主要贡献如下:

(1)通过将 LoRa 信号的相位信息进行差分比值计算,并通过采用短时傅里叶变换(Short-Time Fourier Transform, STFT)生成时频图,进而设计了轻量化 Gesture Encoder 进行特征提取.相比现有需要复杂硬件阵列(如 FMCW 雷达系统)和高算力的手势识别算法(如雷达成像后分类识别),本文算法降低了计算复杂度,保障了实时识别性能,适用于低算力终端.

(2)为解决手势识别中的实时性和准确性问题,提出了一种基于系统状态转换机(System Status Transition Machine, STM)的自动手势分割方法,以精确识别手势的开始和结束时刻,实现手势的实时判断,并将设计的系统部署到搭载 Android 系统的边缘设备上.提出的系统在资源受限的设备上展现出的低功耗运行特性,将进一步推动 LoRa 手势识别在智能交互产品中的应用.

(3)对提出的算法进行了两方面的评估:一方面,与现有其他手势识别算法进行了性能比较;另一方面,在不同场景下进行了验证.这些评估验证了所提算法在手势识别任务中的有效性和准确性,并分别在实时情况和非实时情况下对系统识别能力进行了评估.

## 2 LoRa 感知相关技术

### 2.1 LoRa 感知原理

LoRa 是一种用于长距离、低功耗无线通信的物理层技术. LoRa 采用调频扩频(Chirp Spread Spectrum, CSS)技术<sup>[18]</sup>,该方法通过起始频率变化的 chirp 信号对数据进行编码.线性调频 chirp 波形的复指数  $S(t)$  的表达式为

$$S(t) = e^{j\phi(t)} = e^{j2\pi f_c t + j\pi k t^2} \quad (1)$$

其中,  $\phi(t)$  为瞬时相位;  $f_c$  为信号的中心频率;  $k = B/T$  为频率扫描速率,  $B$  为信号的带宽,  $T$  为 chirp 信号的持续时间. 其瞬时频率  $\frac{1}{2\pi} \cdot \frac{d(\phi(t))}{dt} = kt + f_c$  是一个关于时间的线性函数, 时域信号是一个频率随时间增加的余弦波形, 如图 1 所示. 由于 chirp 信号充分利用了其分配的带宽进行数据编码, 因此在抗噪声、多径效应和多普勒效应等方面具有较强的鲁棒性.

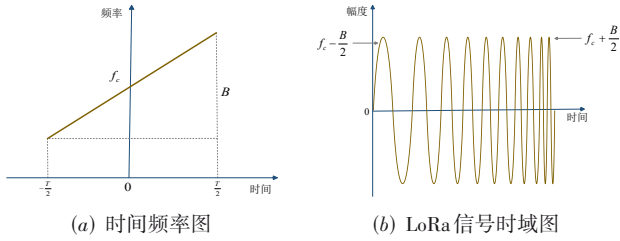


图 1 线性调频 chirp 时域和频域图

LoRa 信号经单目标运动反射产生多径效应被接收机接收<sup>[16]</sup>, 经过 IQ 解调和低通滤波的信号  $R(t)$  可以表示为

$$R(t) = e^{j(\pi k t^2 + \theta_c + \theta_s)} (H_S + H_D(t)) \quad (2)$$

其中,  $\theta_c$  和  $\theta_s$  为由采样频率偏移 (Sampling Frequency Offset, SFO) 和载波频率偏移 (Carrier Frequency Offset, CFO) 引起的项;  $H_S$  和  $H_D(t)$  分别为信号经过不随时间变化的静态路径 (包括直达路径和静态物体反射路径) 以及随时间变化的动态路径 (运动物体反射路径) 的通道响应.  $H_D(t)$  可以表示为

$$H_D(t) = a(t) e^{-j2\pi \frac{d(t)}{\lambda}} \quad (3)$$

其中,  $a(t)$  为动态路径的复衰减因子;  $e^{-j2\pi \frac{d(t)}{\lambda}}$  为动态路径造成的相位变化;  $d(t)$  为动态路径的长度关于时间的函数, 这是由物体运动导致的; 而  $\lambda$  为 LoRa 信号的波长.

如果使用两根天线同步接收 LoRa 信号, 并且两根天线接收装置共享同一个下变频时钟, 那么它们具有相同的 CFO 和 SFO<sup>[15]</sup>. 因此, 本文采用计算两根接收天线信号比值的方法, 有效地消除了共同的 CFO 和 SFO, 从而能够获得稳定的相位差. 具体来说, 信号比  $S_R(t)$  的计算式为

$$\begin{aligned} S_R(t) &= \frac{R_1(t)}{R_2(t)} = \frac{H_{S1} + a_1(t) e^{-j\frac{2\pi d(t)}{\lambda}}}{H_{S2} + a_2(t) e^{-j\frac{2\pi(d(t) + \Delta s)}{\lambda}}} \\ &= H_{S\text{New}} + H_{D\text{New}}(t) = F \left( e^{-j\frac{2\pi d(t)}{\lambda}} \right) \end{aligned} \quad (4)$$

其中,  $R_1(t)$  和  $R_2(t)$  分别为两根不同天线接收到的信

号;  $H_{S1}$  和  $H_{S2}$  分别为来自两根天线接收到信号的静态信道响应; 而  $a_1(t)$  和  $a_2(t)$  分别为两根天线接收到信号在动态路径上的衰减因子;  $d(t)$  为运动物体到天线距离关于时间的函数;  $\Delta s$  为由于两根天线的物理位置不同造成的波程差. 通过对这两根天线接收到的信号进行简单的除法运算以消除 CFO 和 SFO 带来的随机频率偏移, 得到的复信号具有稳定的相位. 文献[15]详细分析了运动目标会导致信号  $S_R(t)$  的相位按一定规律变化, 当目标运动导致  $d(t)$  变化  $\Delta d$  时, 由式(4)可知,  $S_R(t)$  的相位具有约  $2\pi \frac{2\Delta d}{\lambda}$  的明显变化.

## 2.2 特征提取

在手势识别场景中, 所有信号均可以在矢量空间中表示<sup>[15]</sup>. 如图 2(a) 所示,  $S_R(t)$  可以看作是由一个恒定的、与环境相关的静态分量和  $H_{S\text{New}}$  与一个动态的、与手部运动相关的动态分量和  $H_{D\text{New}}(t)$  叠加而成. 其相位定义为矢量与横轴之间的夹角, 其幅度为矢量的长度. 该理论在 Wi-Fi 感知中被广泛用于分析模拟信号变化与人体运动之间的关系<sup>[19,20]</sup>. 在手部运动过程中, 静态路径信号不变, 动态路径信号随手势运动而变化. 对于较小的径向变化, 动态矢量的振幅可以被认为是一个常数, 相反, 动态矢量的相位变化很大. 当手部运动产生 0.5 cm 的变化时, 对于 915 MHz 的 LoRa 信号来说, 相位变化约为  $\frac{2 \times 0.5 \text{ cm}}{33 \text{ cm}} \times 2\pi = 11^\circ$ . 如图 2(b) 所示, 这是因为动态信号矢量相对于矢量空间中的静态信号矢量发生旋转, 从而导致复合信号矢量的相位发生变化.

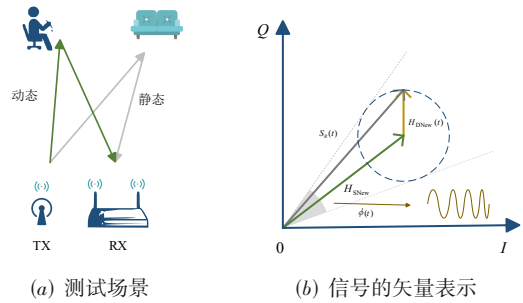


图 2 手势识别场景下的信号路径

$H_{S\text{New}}$  在手势识别场景中几乎不变, 也被视为直流偏移, 且不包含相位信息并且直流分量往往远大于动态分量  $H_{D\text{New}}(t)$ <sup>[21]</sup>. 因此, 本文采用减去信号均值以去除直流偏移. 在天线前方约 1 m 处进行手势操作, 所获得的手势相位模式如图 3 所示.

为了提取相位信息, 本文参考已有文献[17]中使用的方式, 对得到的  $S_R(t)$  进行 STFT, 以提取由运动目标引起的信号相位变化, 从而得到时频图. 在雷达信号处理和手势识别中, 时频图能够帮助分析信号频率随

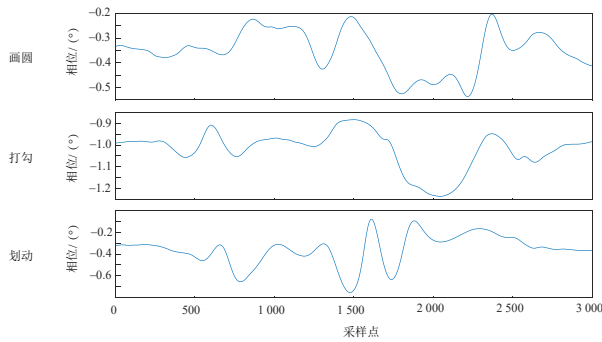


图3 不同手势的相位变化

时间的变化,从而对目标运动信息进行更好地提取. 通过将 STFT 应用于上述的  $S_R(t)$  信号,可得到每个手势对应的时频图.

图 4(a)展示了  $S_R(t)$  在手部前后运动时的相位变化的轨迹,随后,对信号经过巴特沃斯低通滤波器去除高频噪声信号,结果如图 4(b)所示. 信号相位在  $-0.4 \sim 0.4$  弧度往复变化,这是因为被手部反射的信号构成了动态分量  $H_{DNew}(t)$ . 手部的前后运动导致  $d(t)$  往复变化,进而使  $S_R(t)$  呈现周期性变化,这符合手势前后运动的相位特征.

本文根据式(5)得到了如图 4(c)所示的时间频率图  $R(t, f)$ :

$$R(t, f) = \text{abs}\left(\text{STFT}\left(S_R(t) - \text{mean}\left(S_R(t)\right)\right)\right) \quad (5)$$

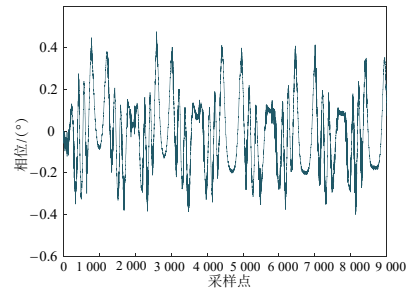
其中,  $\text{abs}(m)$  和  $\text{mean}(n)$  分别为取复数矩阵  $m$  的模和复数序列  $n$  的均值. 在时频图中,正频率和负频率随时间交替出现,这对应于相位的周期性变化,表明相位随着手部的前后运动而往复变化.

### 3 手势识别算法

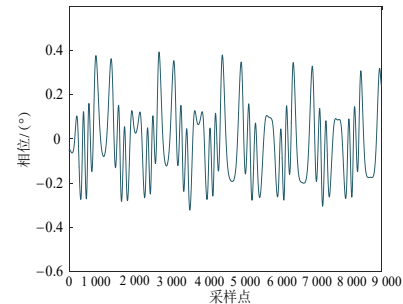
#### 3.1 网络结构

深度学习常用于射频感知的研究中,并且常使用距离角度图<sup>[22]</sup>、距离速度图和角度时间图<sup>[23]</sup>等数据作为神经网络模型的输入数据. 这些数据在一些粗细粒度的感知任务中能够较好地表达特征,并且绝大多数嵌入式神经网络处理单元(Neural Processing Unit, NPU)在处理二维图像数据,如二维卷积神经网络(2 Dimensions Convolutional Neural Networks, 2D-CNN)方面表现出色. 为了提取手部运动特征并在嵌入式设备上保持良好性能,本文提出了一种基于卷积神经网络的轻量化编码器——Gesture Encoder,旨在提取手势特征并生成反映手势特征的特征向量.

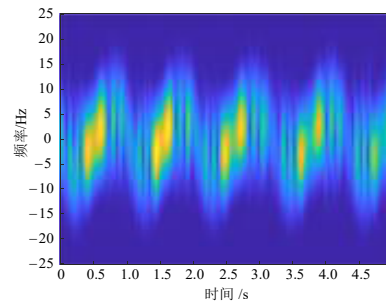
Gesture Encoder 模型由多个类似于 MobileNet<sup>[24]</sup> 的倒残差模块(Residual Block 单元)堆叠而成,如图 5 所示. 模型的输入为时间频率图  $R(t, f)$ ,这是因为在图片形式下,相较于原始的  $S_R(t)$  信号,时间频率图包含更多



(a) 原始信号求比



(b) 求比信号经低通滤波



(c) 时间频率图

图4 信号处理示意图

的特征信息,并且在后序量化模型后依然能够保留大部分信息.

##### 3.1.1 二维卷积层

在卷积神经网络中,卷积层的主要作用是提取图像(或信号)中的局部特征. 在 Gesture Encoder 中,2D-CNN 被用作网络的初始层,用于从输入图像(时频图)中提取基础的特征. 2D-CNN 层包含卷积层、批归一化层和激活函数层. 为了引入非线性特征以增强模型的表达能力,激活函数本文选取 Hard Swish(H-Swish),使用一个分段线性函数取代了计算成本高昂的 Sigmoid 函数,其表达式为

$$\text{H-Swish}(x) = x \cdot \frac{\text{ReLU6}(x+3)}{6} \quad (6)$$

其中,  $\text{ReLU6}(x) = \min(\max(x, 0), 6)$ . 它常用于轻量化网

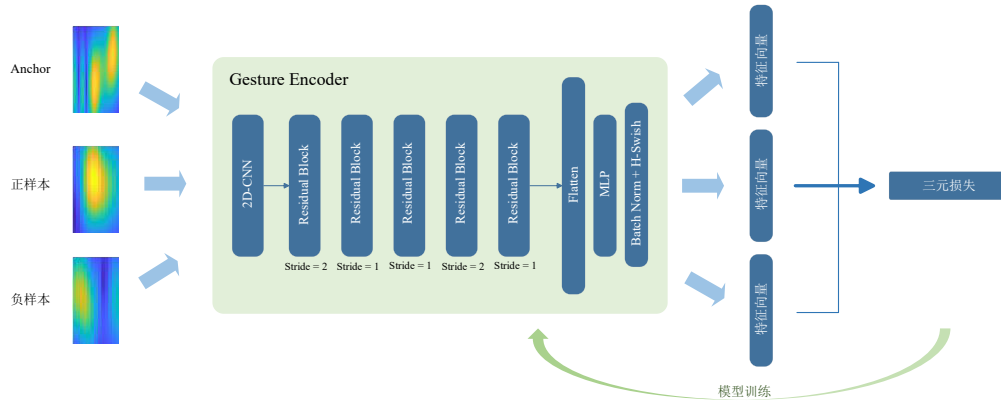


图5 网络架构

络中,且适用于低延迟的移动端和嵌入式设备.

### 3.1.2 残差网络层

在 MobileNet v3<sup>[25]</sup> 中提出的残差网络模块如图 6 所示. 当 Stride=1 时,输入首先经过  $1 \times 1$  的点卷积 (Conv  $1 \times 1$ ) 进行通道数的扩张,其次经过  $3 \times 3$  的深度卷积 (DWise  $3 \times 3$ ),激活函数均为 H-Swish,再次,经过  $1 \times 1$  的点卷积将通道数压缩,激活函数是 ReLU,最后将输入和输出相加构成残差块;而当 Stride=2 时,该模块主要用于特征下采样来减小特征图的空间分辨率,以便在较大感受野下检测整体的图像特征.

本文在最后一个 Residual Block 后使用全局平均池化 (Global Average Pooling, GAP), 将所有的特征压缩成一个通道向量. 在多层感知机 (MultiLayer Perceptron, MLP) 模块中,使用两个层全连接层并结合 dropout 技术以防止过拟合. 每层均有 256 个隐藏单元. 最后,模型的输出维度设计为与特征向量的维度一致.

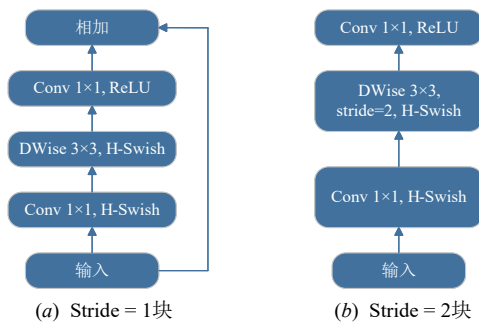


图6 残差模块

### 3.2 损失函数

为了训练本文提出的编码器,使用度量学习的方法,其模型是  $f_{\theta}(x): R^{\text{Image}} \rightarrow R^{\text{Embedding}}$ , 其中  $f_{\theta}(x)$  为需训练的网络函数,参数  $\theta$  为度量学习训练的目标. 它将相似的时频图  $R^{\text{Image}}$  映射到相近的特征向量  $R^{\text{Embedding}}$  上;相反,把不同手势的时频图映射到相距较远的特征向量上. 最终,将训练好的 Encoder 部署在嵌入式设备上,并

保存每个手势的特征向量,其具体实现在第 4 节中进行描述.

定义  $n$  种手势中任意 2 种手势时频图  $R_i(t, f)$  与  $R_j(t, f)$  对应的特征向量的距离为  $D(i, j) = D(f_{\theta}(R_i(t, f)), f_{\theta}(R_j(t, f)))$ , 其中,  $D(x, y) = |x - y|^2$ ,  $i, j \in \{1, 2, \dots, n\}$ . 在本文中,每次训练选择 3 份数据,分别是来自某一手势的  $R_i^a(t, f)$  作为 Anchor,来自同一手势的另一份数据  $R_i^p(t, f)$  作正样本,以及来自其余手势的  $R_j^n(t, f)$  作为负样本<sup>[26]</sup>. 因此,定义三元损失函数  $L$ :

$$L = \left[ D(f_{\theta}(R_i^a(t, f)), f_{\theta}(R_i^p(t, f))) - D(f_{\theta}(R_i^a(t, f)), f_{\theta}(R_j^n(t, f))) + \alpha \right]_+ \quad (7)$$

其中,  $i \neq j, i, j \in \{1, 2, \dots, n\}; y = [x]_+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$ ;  $\alpha$  为超参数表示正负样本对之间的松弛边距.

### 3.3 数据增强

许多深度学习任务依赖于大量的数据集,以确保模型在不同场景下具有良好的通用性. 为了保证模型在实时识别场景下依然具有出色的性能,数据增强在训练模型过程中显得尤为重要. 为了解决这一问题,本文设计了以下数据增强方法.

#### 3.3.1 随机旋转

根据第 2 节中对  $S_R(t)$  的分析以及类似研究<sup>[27]</sup> 中的阐述可知,  $S_R(t)$  由一个静态分量和一个动态分量组合而成. 尽管本文对  $S_R(t)$  序列进行减去均值以消除静态分量,然而,这种方法并不能消除静态分量  $H_{S_{\text{New}}}$  的角度. 当周围场景变化时,静态向量  $H_{S_{\text{New}}}$  的角度也会发生变化. 如图 7 所示,复数序列  $S_R(t)$  被随机旋转,以模拟不同场景中的目标运动,即

$$S_R'(t) = S_R(t) e^{j\theta} \quad (8)$$

其中,  $\theta$  为随机旋转角,且  $\theta \in [-\pi, \pi]$ . 随后,本文对  $S_R'(t)$  按第 2 节中的方式处理得到新的  $R'(t, f)$  参与训练.

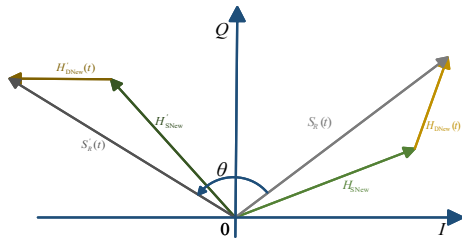
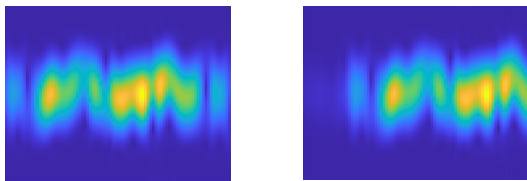


图 7 I/Q 随机旋转示意图

### 3.3.2 随机切分

在基于滑动窗口的实时场景中,算法执行时并不能保证能恰好获得完整的手势时频图  $R(t, f)$ , 具体原因在第 4 节中进行了详细描述. 因此, 本文对  $R(t, f)$  的时间轴进行随机截取, 但截取后的  $R'(t, f)$  保留了原图 70% 以上的部分, 以保留大部分的手势信息, 如图 8 所示.



(a) 数据增强前 (b) 数据增强后

图 8 频谱随机分割示意图

## 4 实时识别系统

本文的编码器仅对已经分割好的手势表现出良好的性能, 如果每个手势都已被完美分割, 那么输入 Gesture Encoder 中就可以得到表示该手势的特征向量. 然而, 在实际使用过程中, 系统在不知道手势何时开始和何时结束的情况下持续接收信号. 为此, 本文提出了一种基于滑动窗口的方法, 以确保任何时刻的手势都能被系统识别, 如图 9 所示. 然而, 如果系统处理每个窗口(其中大多数只是背景噪声, 不含手势数据, 如图 9 中 W1 和 W5), 那么高计算开销和重复识别出低相似度的结果不利于在嵌入式系统中的应用. 并且, 当滑动窗口位于 W2 和 W4 时, Gesture Encoder 仅捕获手势的起始部分和结束部分的信息, 这可能导致错误的手势输出. 只有当窗口位于 W3 时, 即编码器能捕获大部分的手势信息时, 手势分类才能确保正确. 这也是进行数据增强的原因. 因此, 本文根据已有文献[28]中提出的方法, 采用 STM 状态机以减少手势识别的开销并提高识别的准确性.

### 4.1 STM 状态机

为了处理上述问题, 本文在 STM 中设计了 3 种状态: (1) 监测; (2) 判断; (3) 响应. 如图 10 所示.

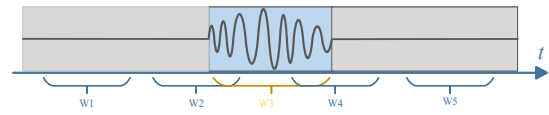


图 9 滑动窗口示意图

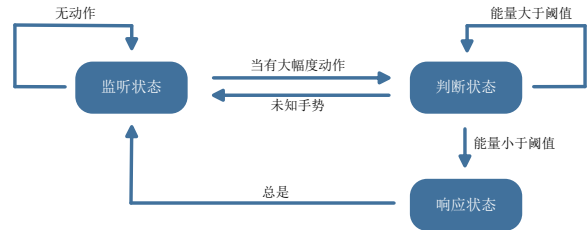


图 10 系统状态转移状态机

#### 4.1.1 监测状态

一个手势包括手势起始部分、做手势和手势结束 3 个部分. 部分已有研究<sup>[29,30]</sup>通过得到每个时刻某个距离 Bin 上频率的最大值, 并计算 STA/LTA (Short-Term Average/Long-Term Average) 来设置阈值, 从而判断手势的起始. 本文采用对某个时间窗口的能量进行长时间指数加权平均的方法来判断手势的起始. 定义某个时间窗口的能量  $e_i$  为

$$e_i = \sum R(t, f) \tag{9}$$

随后进行指数加权平均滤波, 得到

$$E_i = (1 - \beta)E_{i-1} + \beta e_i, i \in N_+ \tag{10}$$

其中,  $E_0 = e_0$ . 当进行大幅度手势时, 窗口的能量  $E_i$  显著高于静止状态, 如图 11 所示. 因此, 本文设置阈值  $E_{i, \text{threshold}}$ , 当  $E_i > E_{i, \text{threshold}}$  时, 视为手势的开始, 进入判断状态.

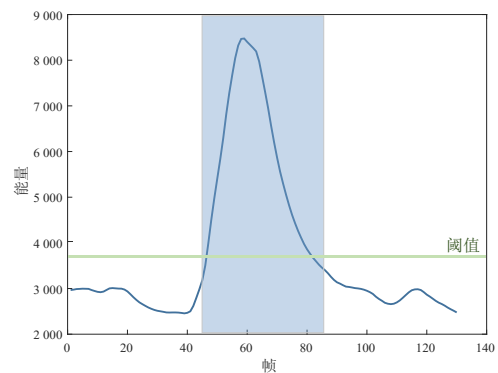


图 11 信号能量随时间变化示意图

#### 4.1.2 判断状态

在此状态下,  $R(t, f)$  被实时输入到神经网络, 并得到实时输出的特征向量. 此外, 每种手势仅保存该手势唯一的特征向量, 具体在第 5 节中进行说明. 已知特征向量数据库第  $i$  种手势的特征向量为  $\text{Feature}(i) = \{v_1^{(i)}, v_2^{(i)}, \dots, v_k^{(i)}\}$ , 其中,  $i = 1, 2, \dots, n, n$  为特征数据库

中保存的手势数量,模型实时输出的特征向量为  $f_{\theta}(R(t, f)) = \{r_1, r_2, \dots, r_k\}$ , 本文定义相似程度

$$\text{Dis}(\text{Feature}(i), f_{\theta}(R(t, f))) = \left( \sum_{m=1}^k (v_m^{(i)} - r_m)^2 \right)^{\frac{1}{2}} \quad (11)$$

其中,  $k$  为特征向量的维度. 此外, 考虑到肢体动作或者正常手部行为导致的误判, 本文设置一个阈值  $R_{\text{threshold}}$ , 并将最大相似度大于  $R_{\text{threshold}}$  作为可信备选结果发送到下一个响应状态, 一旦  $E_i < E_{\text{threshold}}$ , 那么 STM 将进入响应状态.

#### 4.1.3 响应状态

在这种状态下, STM 仅响应一次可靠的识别结果. 当系统得到判断状态下得到一系列相似度大于  $R_{\text{threshold}}$  的可信手势备选及其对应识别窗口的能量后, 系统选取能量值最大的手势作为最终输出的手势, 并执行相应的响应, 以避免因多次输出而导致的误操作. 这是因为当滑动窗口位于 W2 和 W4 时, 虽然有可能得到相似度较高的错误手势, 但这些手势的能量远小于 W3 窗口下的能量. 而在 W3 窗口内, 结合本文的数据增强方法, 编码器在输入包含大部分手势信息的情况下, 能够得到与数据库中对应手势相似度高的特征向量. 图 12 展示了做右划手势 (Right swipe gesture, RT) 时, 滑动窗口下相似度大于  $R_{\text{threshold}}$  的识别结果以及识别窗口内能量值的变化. 采用本文提出的方法可以很好地获得正确的手势结果. 在该状态机的最后, 系统清除缓冲区中记录的可信手势备选, 然后重新回到监测状态.

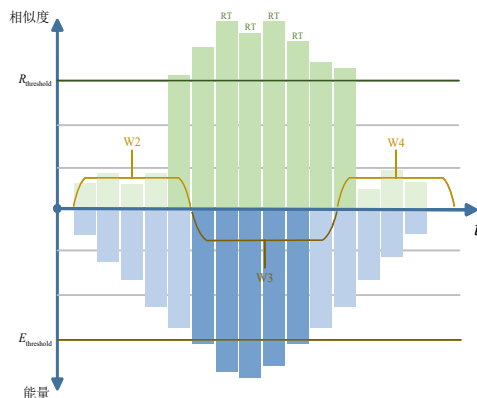


图 12 实时系统中能量值的变化和模型的推理结果

#### 4.2 实时系统设计

本文在 ALIENTEK 的 AtomPi-CA1 开发板<sup>[31]</sup>上部署了提出的轻量化模型. 该开发板搭载了 Rock Chip 的 RK3568 SOC, 它包含 ARM CortexA-55CPU 核心、1 TOPS 能力的 NPU 以及一些常见的外设, 如 USB3.0 和以太网口等<sup>[32]</sup>. 此外, 接收机通用软件无线电 (Universal Software Radio Peripheral, USRP) 通过 USB 连接到开

发板以接收数据, 具体参数将在第 5 节中描述. 开发板运行官方提供的 Android 11 系统, 考虑到本文的应用场景是 HCI, 并且绝大多数交互设备 (如智能电视、自助机等) 均搭载 Android 系统. 本研究在 Android 系统上移植了 USRP 硬件驱动 (USRP Hardware Driver, UHD) 4.7.0<sup>[33]</sup>、Boost 1.85.0<sup>[34]</sup> 和 LibUSB 1.0.27<sup>[35]</sup> (UHD 的依赖库). 本研究使用的交叉编译器套件是 Android NDK R27<sup>[36]</sup>, 编译时指定 ANDROID ABI 为 arm64-v8a, ANDROID PLATFORM 等级为 21.

如图 13 所示, 应用程序包含 3 个线程, 线程 1 通过 UHD 不断读取 USRP 接收机数据并放入环形缓冲区内 (滑动窗口长度为 3 s), 然后进行信号求比, 并在每个时间间隔内 (本文设置为 100 ms) 将数据通过管道队列 (First-in First-out, FIFO) 发送至线程 2, 以进行低通滤波和静态成分减除. 线程 2 负责对时间窗口内的数据进行 STFT, 本文选取的窗函数是汉明 (hamming) 窗, 重叠 (Overlap) 是 100, 快速傅里叶变换 (Fast Fourier Transform, FFT) 输出长度是 2 048, 得到的时频矩阵维度为  $103 \times 103$ . 处理完毕后, 使用 FIFO 将数据传递给线程 3. 线程 3 通过 RKNPU API 1.5.0<sup>[37]</sup> 将预先量化的模型 (本文对训练完保存的模型进行量化, 且量化参数选择为 INT8) 和实时数据移动到板载 NPU 运算, 得到结果输入到 STM 状态机. STM 状态机将读取预先训练的特征数据库中的数据进行比对, 并按照 4.1 节描述的方法得到并输出手势信息.

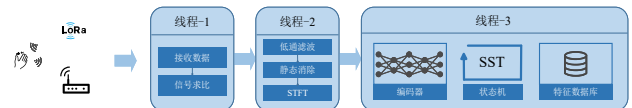


图 13 实时系统设计框图

## 5 实验结果

### 5.1 硬件和数据集介绍

本文设计了如图 14 所示的 8 个标准手势, 这 8 个手势分别如下: (a) 打勾; (b) 交叉; (c) 顺时针画圈; (d) 逆时针画圈; (e) 左划; (f) 右划; (g) 挥手; (h) 下挥.

实验设备如图 15 所示, 本文使用的 LoRa 射频芯片是 Micro Chip 的 RF96 芯片<sup>[38]</sup>, 并且使用 Arduino

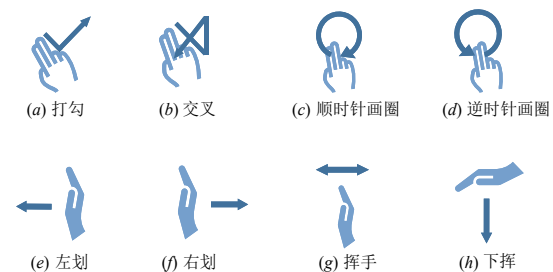


图 14 8 种手势示意图

UNO<sup>[39]</sup>通过 SPI 协议驱动其发射 LoRa 信号,本文设置 LoRa 中心频率为 915 MHz,带宽为 125 K,发射功率是 20 dBm,定向天线的天线增益是 8 dBi;本文选择了 USRP B210<sup>[40]</sup>作为接收机,USRP 采样率是 1 M,使用与发送相同的定向天线接收信号,且 USRP 接收机增益设置为 20 dB. 本文参考现有开发文档<sup>[41]</sup>编写 LoRa 控制程序,设置连续发送 LoRa 普通数据包,使用 USRP UHD 提供的 C++ API 编写数据采集程序. 在数据采集时,USRP 通过 USB 线缆连接到计算机. 在实时场景下,USRP 通过线缆连接到 AtomPi-CA1 开发板.



图 15 实验设备

本研究召集了 6 位志愿者进行手势数据的采集,所有手势在 3 s 内完成,包含手势起始部分(抬起手)、做手势部分(画圈、水平推拉等)以及手势结束(收起手). 数据收集场景如图 16 所示,为办公室实验场景. 在一个办公室中,将发射和接收天线放在检测区域左侧,手势动作在右边区域完成. 中间使用厚度约为 1.5 cm 的木隔板作为障碍物,模拟实际使用中的遮挡情况. 每个手势共采集了 800 个数据,8 个手势共收集了 6 400 个样本. 此外,本文还收集了等量的静止状态下的数据,作为 3.2 节中定义的负样本,以避免模型在静止状态下的输出与现有手势混淆. 采集的信号按照 2.2 节的方法进行预处理后,数据被保存并标记了相应的标签.

## 5.2 算法评估

本节主要将本文提出的算法性能与其他现有的手势识别算法进行比较. 本文采用提出的三元损失函数训练模型,并将数据集按照 7:2:1 划分为训练集、测试集和验证集. 在训练过程中,本文选取特征向量的维度  $k$  为 128. 对于每个手势,本文计算所有样本输出的特征向量的均值,得到唯一的特征向量,并将其作为特征向量数据库的一部分. 手势分类按照第 4 节中描述的方

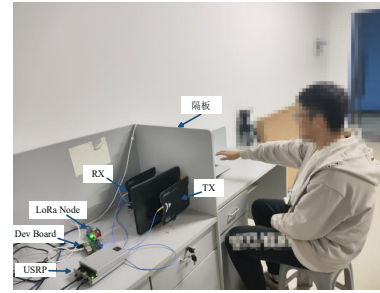


图 16 办公室实验场景

式进行. 本文保存在测试集上准确度最优的特征向量集合,以便在实时系统中使用. 最终,验证集用于评估模型的性能.

### 5.2.1 网络结构

在手势识别算法方面,许多研究工作都采用了类似的神经网络来完成分类任务. 部分现有研究<sup>[29,42,43]</sup>也将信号处理为光谱图像输入深度神经网络. 文献[42]使用了卷积神经网络(Convolutional Neural Networks, CNN)和长短期记忆(Long Short-Term Memory, LSTM)网络来提取空间和时间特征;文献[28]仅使用了简单的 CNN 结构来完成基于 mmWave 的实时手势分类;而在我们之前的工作中,借助了 ResNet 来完成行为的分类<sup>[43]</sup>. 这些工作的共同点在于尝试从光谱图中提取用于分类的特征. 然而,在本文的应用场景中,我们希望用于手势分类的模型尽可能轻量化,以便适用于大多数商业嵌入式系统,这些系统通常具有有限的存储和计算能力.

本文将提出的 Encoder 替换为上述文献中所使用的模型以及已有的图片分类网络<sup>[44]</sup>,并比较了它们在自建数据集上的准确性和模型大小,以评估其特征提取能力,具体结果如表 1 所示. 为了确保公平比较,本文对模型的结构、超参数以及训练参数进行了优化. 本文使用 Adam 优化器训练,设置初始学习率为 0.001, batch size 为 64,并当分类准确率经过多个 epoch 训练不再上升时结束训练. 实验结果表明:单纯的 CNN<sup>[29]</sup>在部分手势上难以区分,而 CNN+LSTM<sup>[42]</sup>、ResNet<sup>[44]</sup>和文献[43]的平均准确率分别为 95.83%、95.54% 和 96.83%,与本文提出的模型精度几乎相同. 然而,本文算法所需要的计算资源和内存要少得多. 结果表明:本文的神经网络以较小的参数量实现了较高的分类精度.

表 1 不同模型架构下的识别准确率

算法	平均/%	(a)/%	(b)/%	(c)/%	(d)/%	(e)/%	(f)/%	(g)/%	(h)/%	参数量大小/M
CNN <sup>[29]</sup>	84.56	79.87	76.58	76.55	71.25	88.36	94.63	92.92	96.35	5.32
CNN+LSTM <sup>[42]</sup>	95.83	93.87	94.63	93.75	89.91	99.53	96.99	98.64	99.35	123.21
ResNet18 <sup>[44]</sup>	95.54	93.23	94.23	93.61	92.13	97.29	98.51	97.42	97.88	56.32
文献[43]	96.83	93.32	95.32	96.61	95.43	97.52	98.10	98.94	99.42	89.74
本文算法	95.98	94.31	94.12	95.32	93.40	96.43	98.17	97.32	98.73	11.32

### 5.2.2 损失函数

已有文献<sup>[29,43,45]</sup>将模型的输出输入到 Softmax 函数,得到概率并选择概率最大的索引作为分类结果,同时使用交叉熵作为损失函数.这种方式在非实时情况下能够获得很高的准确率,因为它将每个输入分配到一个特定类别,并通过最大化该类别的概率来进行训练.然而,在本文的实时场景中,仅对输出概率设置阈值并不能达到很好的效果.这因为在手势识别中存在细粒度的区分问题,尤其是滑动窗口下,不同手势的开始和结束部分具有较高的相似性.图 17 展示了在两种方式下,模型输出向量通过  $t$ -SNE 降维后的特征图.

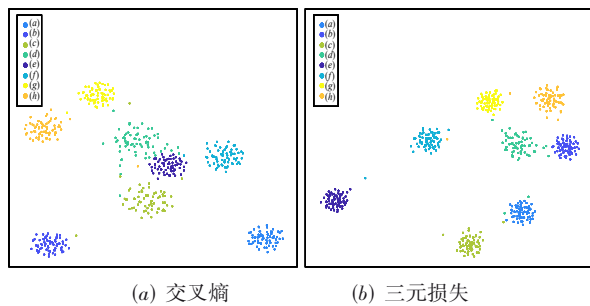


图 17 不同损失函数情况下模型输出的特征图

由图 17 可知,交叉熵函数部分相似的手势出现了混叠,而三元损失函数不仅优化了分类结果,还同时优化了特征空间中不同样本类别之间的距离关系,使其能够更好地捕捉类别之间的差异和相似性.

### 5.2.3 数据包类型

为了探究 LoRa 发射信号对手势识别准确度的影响,本文对 Arduino 编程发送不同类型的数据包,并进行了手势演示.具体来说,对发送全 0 数据包、固定数据包和随机数据包的情况,额外录制了 50 份数据,并使用上述方法训练的编码器及手势分类算法进行了实验.

表 2 展示了在 LoRa 通信的场景下,发送不同数据包情况时手势识别的准确率.实验结果表明:不同标准手势在不同数据包下的准确性相似.这是因为信号比值仅保留了到达角度信息,与原始发射信号无关.因此,基于 LoRa 信号的手势识别在一定程度上可以在不影响通信功能的同时进行感知应用,从而为实现通感

一体化功能奠定了基础.与 Wi-Fi 信号和 mmWave 雷达相比,LoRa 信号在手势识别方面具有更好的研究意义和应用前景.

表 2 不同数据包下的识别准确率 单位:%

数据包	平均	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
随机数据	95.96	93.42	94.42	93.32	94.32	96.32	97.32	99.73	98.82
固定数据	97.07	94.32	95.42	96.82	96.23	97.84	97.32	99.84	98.76
全 0 数据	96.17	93.12	93.32	94.12	97.02	97.41	96.48	98.19	99.73

### 5.2.4 不同场景

为进一步验证提出系统的鲁棒性和适应性,本文选择在新的场景下使用不同障碍物测试提出手势识别算法的性能.在一个会议室中,选择厚度约 4 mm 的玻璃作为障碍物进行测试,实验设备放置在会议室外走廊中,穿透玻璃探测屋内人员的手势,作为新场景 1,如图 18(a)所示;选择在相同会议室中使用厚度约 3 mm 的中空双层纸盒作为障碍物,宽度约 6 cm,实验设备穿透双层纸盒探测人员手势,作为新场景 2,如图 18(b)所示.在两个新场景中,对每个手势分别采集了 60 个样本,8 个手势共收集了 480 个样本,使用与办公室场景数据相同方式进行训练.

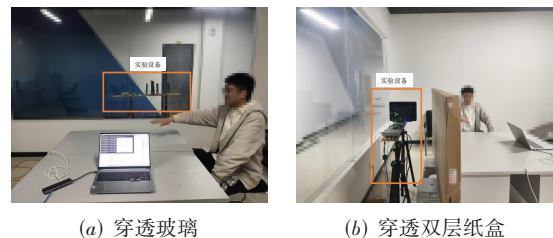


图 18 会议室实验场景

实验结果如表 3 所示,在不同场景和不同遮挡下,均达到 94% 以上的准确率,达到预期效果.这是因为 LoRa 低频信号具有很好的穿透性,同时本文描述的特征提取方法能够很好地捕获运动手部的特征.在会议室场景 2 中,双层硬纸盒除了两层纸板外中间有空隙,导致信号散射,从而使得识别效果更差.在会议室场景 1 中,尽管玻璃信号穿透性能理论上应优于办公室场景中的木制隔板,但由于 LoRa 设备放置在走廊中,用于测量室内的手势动作,距离手势位置较远,因此最终的识别效果与办公室场景相当.

表 3 不同场景下的识别准确率

单位:%

场景	平均	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
办公室场景	95.98	94.31	94.12	95.32	93.40	96.43	98.17	97.32	98.73
会议室场景 1	95.82	94.77	93.21	93.76	96.76	95.47	96.54	98.30	97.74
会议室场景 2	94.17	92.98	93.98	91.31	93.43	95.43	95.43	96.43	94.34

### 5.3 实时性能评估

本节描述了在实时情况下识别算法的准确度以及设备运行的性能. 本文设置指数加权平均滤波器系数 $\beta$ 为0.85,  $E_{\text{threshold}}$ 为7 500,  $R_{\text{threshold}}$ 为0.87.

#### 5.3.1 准确率

在进行实时场景下的性能测试中, 每位数据录制者每个动作连续演示 20 次. 在此过程中, 边缘计算平台直接将得到的原始数据经过上述处理, 并按照所描述的方式输入到 NPU. 系统实时显示并保存 STM 的输出结果. 最后, 统计了整体的实时识别准确率, 结果如表 4 所示.

表 4 实时场景下的识别准确率 单位: %

场景	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
准确率	79.73	82.31	84.32	86.31	89.32	86.43	87.21	92.21

实验结果表明: 本文提出的算法在实时场景下对于各手势分类的平均识别准确率依然能够达到 85.98%. 这一准确率略低于非实时情况下的表现, 原因在于实时场景中存在诸多影响因素, 包括量化模型中的精度损失、身体的额外晃动、手势的准确分割以及场景的细微变化. 这些非理想因素的存在导致了实时场景下的分类准确率低于非实时情况下的准确率.

#### 5.3.2 系统耗时

系统的整体耗时如表 5 所示, 包括数据采集、低通滤波、静态减除、STFT、Gesture Encoder 和 STM 输出. 需要指出的是, 除了 Gesture Encoder, 其余均为使用 C 代码在 CPU 上执行, 而 Encoder 则是在 RK3568 的片上 NPU 中执行的. 在量化后, Encoder 模型的大小仅为 4.8 M. 在未使用任何 FFT 加速引擎的情况下, 编码器仅需花费大约 12.12 ms. 本文提出的手势识别系统考虑到功耗和准确度, 设置整体滑动窗口时间为 100 ms, 可以很好地满足实时性和准确度要求.

表 5 实时场景下的时间消耗 单位: ms

项目	实时推理(NPU)	其他部分(CPU)	总计耗时
平均时间	12.12	8.84	20.96

## 6 结束语

本文提出了一种基于 LoRa 信号的实时手势识别算法. 该算法采用了本文提出的背景消除和数据增强方法, 并在此基础上, 设计了一种轻量化的手势识别编码器, 以及利用 STM 状态机解决实时手势分割问题. 在实时场景中, 8 个标准手势的平均分类准确率可以达到 85.98%. 本文也与最近的雷达手势识别系统进行了比较, 充分证明了该方案在有遮挡物体条件下进行准确识别的可行性. 此外, 本文提出了一个面向边缘计算的手势识别算法, 该系统使用了基于 NPU 的硬件加速器,

能够在大约 20.96 ms 内完成手势分类. 这种手势识别系统有潜力应用于非接触式 HCI 控制系统. 在后续的研究工作中, 将进一步优化系统的实时性和识别精度, 并拓展其在更多复杂场景中的应用.

#### 参考文献

- [1] QIU X R, LIU J H, SONG L L, et al. A survey of gesture recognition using frequency modulated continuous wave radar[J]. Journal of Computer and Communications, 2024, 12(6): 115-134.
- [2] AL FARID F, HASHIM N, ABDULLAH J B, et al. Single shot detector CNN and deep dilated masks for vision-based hand gesture recognition from video sequences[J]. IEEE Access, 2024, 12: 28564-28574.
- [3] QI J, MA L, CUI Z C, et al. Computer vision-based hand gesture recognition for human-robot interaction: A review[J]. Complex & Intelligent Systems, 2024, 10(1): 1581-1606.
- [4] ZERROUKI N, HARROU F, HOUACINE A, et al. Deep learning for hand gesture recognition in virtual museum using wearable vision sensors[J]. IEEE Sensors Journal, 2024, 24(6): 8857-8869.
- [5] ALEMU M Y, LIN Y, SHULL P B. EchoGest: Soft ultrasonic waveguides based sensing skin for subject-independent hand gesture recognition[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024, 32: 2366-2375.
- [6] JIN B, MA X, HU B J, et al. Gesture-mmWAVE: Compact and accurate millimeter-wave radar-based dynamic gesture recognition for embedded devices[J]. IEEE Transactions on Human-Machine Systems, 2024, 54(3): 337-347.
- [7] OH H, BHATTACHARYA S, SEO S. UWB/IMU-assisted gesture recognition using learning approaches for VR/XR applications[C]//ICC 2024 - IEEE International Conference on Communications. Piscataway: IEEE, 2024: 4838-4844.
- [8] ZHANG S C, WANG Q J, GAN M L, et al. RadSee: See your handwriting through walls using FMCW radar[C]//Proceedings 2025 Network and Distributed System Security Symposium. Internet Society, 2025: 112-123.
- [9] ZHAO L Q, XIAO R, LIU J W, et al. One is enough: Enabling one-shot device-free gesture recognition with COTS WiFi[C]//IEEE INFOCOM 2024 - IEEE Conference on Computer Communications. Piscataway: IEEE, 2024: 1231-1240.
- [10] GAO D M, WANG H Y, WANG S, et al. WiLo: Long-range cross-technology communication from Wi-Fi to LoRa[J].

- IEEE Transactions on Communications, 2025, 73(3): 1677-1691.
- [11] ZHANG F S, CHANG Z X, XIONG J, et al. Exploring LoRa for sensing[J]. *GetMobile: Mobile Computing and Communications*, 2021, 25(2): 33-37.
- [12] GE Y, LI W D, FAROOQ M, et al. LoGait: LoRa sensing system of human gait recognition using dynamic time warping[J]. *IEEE Sensors Journal*, 2023, 23(18): 21687-21697.
- [13] HUANG Q Y, LUO Z Q, ZHANG J, et al. LoRadar: Enabling concurrent radar sensing and LoRa communication[J]. *IEEE Transactions on Mobile Computing*, 2022, 21(6): 2045-2057.
- [14] XIE B B, GANESAN D, XIONG J. Embracing LoRa sensing with device mobility[C]//*Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. New York: ACM, 2022: 349-361.
- [15] XIE B B, XIONG J. Combating interference for long range LoRa sensing[C]//*Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. New York: ACM, 2020: 69-81.
- [16] ZHANG F S, CHANG Z X, NIU K, et al. Exploring LoRa for long-range through-wall sensing[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020, 4(2): 1-27.
- [17] ZHANG F S, CHANG Z X, XIONG J, et al. Unlocking the beamforming potential of LoRa for long-range multi-target respiration sensing[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021, 5(2): 1-25.
- [18] BERNI A, GREGG W. On the utility of chirp modulation for digital signaling[J]. *IEEE Transactions on Communications*, 1973, 21(6): 748-751.
- [19] WANG W, LIU A X, SHAHZAD M, et al. Understanding and modeling of WiFi signal based human activity recognition[C]//*Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. New York: ACM, 2015: 65-76.
- [20] WANG H, ZHANG D Q, MA J Y, et al. Human respiration detection with commodity WiFi devices: Do user location and body orientation matter[C]//*Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York: ACM, 2016: 25-36.
- [21] GE J F, ZHU D S, SUN L J, et al. A long-range signal-based target localization algorithm[J]. *Electronics*, 2024, 13(6): 1069.
- [22] 韩崇, 韩磊, 孙力娟, 等. 基于时空压缩特征表示学习的毫米波雷达手势识别算法[J]. *电子与信息学报*, 2022, 44(4): 1274-1283.
- HAN C, HAN L, SUN L J, et al. Millimeter wave radar gesture recognition algorithm based on spatio-temporal compression feature representation learning[J]. *Journal of Electronics & Information Technology*, 2022, 44(4): 1274-1283. (in Chinese)
- [23] 王勇, 王沙沙, 田增山, 等. 基于FMCW雷达的双流融合神经网络手势识别方法[J]. *电子学报*, 2019, 47(7): 1408-1415.
- WANG Y, WANG S S, TIAN Z S, et al. Two-stream fusion neural network approach for hand gesture recognition based on FMCW radar[J]. *Acta Electronica Sinica*, 2019, 47(7): 1408-1415. (in Chinese)
- [24] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 4510-4520.
- [25] QIAN S Y, NING C R, HU Y P. MobileNetV3 for image classification[C]//*2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. Piscataway: IEEE, 2021: 490-497.
- [26] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2015: 815-823.
- [27] ZHU D S, HAN C, GUO J, et al. TWLip: Exploring through-wall word-level lip reading based on coherent SISO radar[J]. *IEEE Internet of Things Journal*, 2024, 11(19): 32310-32323.
- [28] LIU H P, ZHOU A F, DONG Z H, et al. M-gesture: Person-independent real-time in-air gesture recognition using commodity millimeter wave radar[J]. *IEEE Internet of Things Journal*, 2022, 9(5): 3397-3415.
- [29] SUN Y L, FEI T, LI X B, et al. Real-time radar-based gesture detection and recognition built in an edge-computing platform[J]. *IEEE Sensors Journal*, 2020, 20(18): 10706-10716.
- [30] VAEZI Y, VAN DER BAAN M. Comparison of the STA/LTA and power spectral density methods for microseismic event detection[J]. *Geophysical Journal International*, 2015, 203(3): 1896-1908.
- [31] Aliantek Technology. RK3568 Dev board[EB/OL]. (2023-

- 01-01)[2024-12-22]. <https://www.alientek.com>.
- [32] Rockchip Technology. RK3568 brief datasheet[EB/OL]. (2022-01-01)[2024-12-22]. <https://www.rock-chips.com/uploads/pdf/2022.8.26/192/RK3568%20Brief%20Datasheet.pdf>.
- [33] Research Ettus. UHD (Universal Hardware Driver)[EB/OL]. (2023-01-01)[2024-12-22]. <https://github.com/Ettus-Research/uhd>.
- [34] MORITZ-WUNDKE. Boost for android[EB/OL]. (2023-01-01)[2024-12-22]. <https://github.com/moritz-wundke/Boost-for-Android>.
- [35] LIBUSB. Libusb for android[EB/OL]. (2023-01-01)[2024-12-22]. <https://github.com/libusb/libusb>.
- [36] Android, Inc. Android NDK[EB/OL]. (2022-01-01)[2024-12-22]. <https://developer.android.com/ndk>.
- [37] Rockchip Technology. KNN-toolkit2[EB/OL]. (2024-01-01)[2024-12-22]. <https://github.com/airockchip/rknn-toolkit2>.
- [38] Microchip Technology. LoRa transceiver[EB/OL]. (2006-01-01)[2024-12-22]. [https://microchip.ua/wireless/RF96\\_97\\_98.pdf](https://microchip.ua/wireless/RF96_97_98.pdf).
- [39] Arduino. Arduino UNO rev3[EB/OL]. (2015-01-01)[2024-12-22]. <https://store.arduino.cc/usa/arduino-uno-rev3>.
- [40] Research Ettus. USRP B210[EB/OL]. (2013-01-01)[2024-12-22]. <https://www.ettus.com/all-products/UB210-KIT>.
- [41] CHANDRAWI. LoRaRF-arduino[EB/OL]. (2021-01-01)[2024-12-22]. <https://github.com/chandrawi/LoRaRF-Arduino>.
- [42] 陈君毅, 蒋德琛, 王智铭, 等. 一种基于双维度滤波和自适应定长化的FMCW雷达手势识别算法研究[J]. 电子学报, 2023, 51(8): 2179-2187.  
CHEN J Y, JIANG D C, WANG Z M, et al. Research on a gesture recognition algorithm for FMCW radar based on bidimensional filtering and adaptive fixed length[J]. Acta Electronica Sinica, 2023, 51(8): 2179-2187. (in Chinese)
- [43] 韩崇, 樊卫北, 郭澳. 基于特征融合的毫米波雷达行为识别算法[J]. 计算机科学, 2024, 51(12): 181-189.  
HAN C, FAN W B, GUO A. Millimeter wave radar human activity recognition algorithm based on feature fusion[J]. Computer Science, 2024, 51(12): 181-189. (in Chinese)
- [44] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [45] GAN L Y, LIU Y, LI Y Z, et al. Gesture recognition system using 24 GHz FMCW radar sensor realized on real-time edge computing platform[J]. IEEE Sensors Journal, 2022, 22(9): 8904-8914.

### 作者简介



韩 崇 男, 1985年6月生, 河南南阳人。现为南京邮电大学计算机学院副教授、硕士生导师。主要研究方向为射频感知和普适计算。  
E-mail: hc@njupt.edu.cn



郭 剑 男, 1978年7月生, 江苏南通人。现为南京邮电大学计算机学院副教授、硕士生导师。主要研究方向为无线传感器网络和无线感知。  
E-mail: guoj@njupt.edu.cn



杜 昊 男, 2001年4月生, 江苏南京人。现为南京邮电大学计算机学院硕士研究生。主要研究方向为无线感知。  
E-mail: 1023040912@njupt.edu.cn



孙力娟 女, 1963年12月生, 江苏南京人。现为南京邮电大学计算机学院教授、博士生导师。主要研究方向为演化计算、物联网和无线感知。  
E-mail: sunlj@njupt.edu.cn