

# 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

陈君毅, 蒋德琛, 王智铭, 曹佳禾, 王 勇

(浙江大学信息与电子工程学院, 浙江杭州 315000)

**摘 要:** 本文提出了一种基于调频连续波(Frequency Modulated Continuous Wave, FMCW)雷达回波信号的手势识别算法: 首先, 提出一种双维度滤波算法, 在距离和速度维度对手势回波信号进行滤波, 有效地降低了系统的静态噪声; 其次, 将数据经过动目标检测(Moving Target Indicator, MTI)算法滤除时间维度噪声; 然后, 提出了时间自适应定长化的方法, 在减少手势信息损失的前提下保证了每个手势样本帧数的一致性; 最后, 建立距离多普勒网络(Range Doppler Net, RD-Net)进行训练分类. 该算法在谷歌开源的deep-soli数据集中取得了98.28%的准确率, 比数据集提出者的算法的准确率提升了11.11%. 该算法在实时推理实验中取得了90.8%的准确率, 具有更好的泛化能力.

**关键词:** 毫米波雷达; 手势识别; 深度学习; 数据处理; 实时推理

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)08-2179-09

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

DOI: 10.12263/DZXB.20211410

## Research on a Gesture Recognition Algorithm for FMCW Radar Based on Bidimensional Filtering and Adaptive Fixed Length

CHEN Jun-yi, JIANG De-chen, WANG Zhi-ming, CAO Jia-he, WANG Yong

(College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang 315000, China)

**Abstract:** This paper proposes a gesture recognition algorithm based on frequency modulated continuous wave (FMCW) radar echo signals. Firstly, a two-dimensional filtering algorithm is proposed to filter the gesture echo signals in the distance and speed dimensions, which effectively reduces the static noise of the system. Secondly, the data is filtered by the moving target indicator (MTI) algorithm to filter out the noise in the time dimension. Then a time-adaptive fixed-length method is proposed, which ensures the consistency of the frame number of each gesture sample on the premise of reducing the loss of gesture information. Finally, a range Doppler net (RD-Net) is established for training and classification. The algorithm achieved 98.28% accuracy in Google's open source deepsolis data set, which is 11.11% higher than the algorithm proposed by the data set. The algorithm achieves 90.8% accuracy in real-time reasoning experiments and has better generalization ability.

**Key words:** millimeter-wave radar; gesture recognition; deep learning; data processing; real-time reasoning

### 1 引言

随着信息技术的发展, 人机交互的频率不断提高<sup>[1]</sup>. 手势识别技术以其非接触的方式使人机交互更为便捷, 具有其独特的应用场景. 常见的应用在于智能家居、手机终端控制、智能座舱等领域.

目前主流的手势识别方式是基于视觉的处理. 这类研究通过单目或者多目摄像头对手势动作进行采集, 通过手势检测、动作追踪和分类识别等一系列处理流程达到手势识别的效果<sup>[2]</sup>. 然而基于视觉的方式对

成像质量要求较高, 容易受到光照强度、摄像头像素等客观条件的影响, 在光线不足的情况下无法使用; 且很容易出现用户隐私泄露<sup>[3]</sup>等问题.

相比基于视觉的方式, 毫米波雷达手势识别的优点在于: (1) 数据来源于雷达回波信号, 光照对其的影响较小, 适用的场景范围更大; (2) 回波信号相比与图像信号, 不会暴露过多用户的周边细节, 能够更好地保障用户隐私; (3) 回波信号的数据量相比图像信号量级更小, 易于处理. 而FMCW雷达相较于传统雷达, 穿透性强, 拥有更高的发射频率和更低的功耗; 可以准确测

量距离、速度和角度信息,因此衍生了一系列应用。

手势识别有静态识别和动态识别的区分。静态是指仅采集手势完成时刻的数据,而动态是指采集手势整个动作的数据。在静态目标的检测识别上,Yeo团队<sup>[4]</sup>在2016年使用FMCW雷达进行了物体识别和身体部位区分这两项工作。物体识别是将不同的物体放置于FMCW雷达上方,通过分析回波信号进行区分,实现26种物品98.76%的识别准确率。身体部位区分是将雷达放置在不同的身体部位上进行,实现11个部位90.27%的识别准确率。物体识别和身体部位区分说明不同的物体在FMCW雷达上反射的回波信号的差异是相对明显的。Yeo团队<sup>[5]</sup>的另一项工作则更为深入,使用FMCW雷达对不同的卡片进行区分,甚至可以区分同一张卡片的不同摆放位置。这项工作证明了即使拥有相同的形状,FMCW雷达依旧可以根据内在结构的细小差别区分不同的物体甚至相同的物体的不同空间位置。Qualcomm公司的研究团队<sup>[6]</sup>采用基于神经网络的自动编码器通过阵列雷达信号进行人脸识别。这些实验都证明了FMCW雷达对细微事物的分辨率十分精准。

在动态目标的检测识别上,Tahmouh等人<sup>[7]</sup>使用将信号提取为微多普勒图进行步态识别。Vandersmissen等人<sup>[8]</sup>则是考虑到遮挡对FMCW雷达信号的影响,进行了有墙壁阻挡的雷达步态识别,取得了不错的效果。而Google团队<sup>[9]</sup>推出了Soli模块,并将之应用到手势识别领域。Suh等人<sup>[10]</sup>将长短期记忆神经网络(Long Short-Term Memory, LSTM)应用到24 GHz FMCW雷达手势识别领域。Smith<sup>[11]</sup>利用FMCW雷达手势识别进行车内设备控制。重庆邮电大学的王勇等人<sup>[12]</sup>设计出一种具有端到端结构的RDA-T多维参数卷积神经网络(Convolutional Neural Networks, CNN)进行手势识别。复旦大学的夏朝阳等人<sup>[13]</sup>实现了基于多通道FMCW雷达的微动手势识别。由此可见,FMCW雷达进行手势识别是可行的,有着相对成熟的条件。

与现有的研究不同,本文提出一种新的手势识别算法,对雷达回波信号进行距离和速度双维度滤波。相比目前准确率最高的方法——直接通过两次快速傅里叶变换(Fast Fourier Transform, FFT)计算得到距离多普勒图(Range Doppler Image, RDI)的方法<sup>[14]</sup>,本文算法更有效地滤除了静态噪声。不同人员的习惯差异导致了采集的手势数据帧数不一致。为解决该问题,与传统的截断补零不同,本文采用时间自适应定长化的方法,在减少信息损失的前提下保证了数据时间维度的一致性。本文最终将经过上述算法处理后的多通道数据进行融合,通过RD-Net进行分类。本文算法在开源的deep-soli数据集<sup>[1]</sup>上准确率提升了11.11%,在实时推理情况

下取得了90.8%的准确率。

## 2 数据集建立

本文搭建了一个采集数据的实验平台进行手势识别的相关实验。本次实验的系统架构整体分为数据采集和数据处理两大部分。数据采集部分通过计算机发送雷达参数配置文件控制雷达采集回波信号数据,该数据通过串口读取到PC机上,在PC机上进行数据处理并输出对应结果。数据处理部分采用下文中的算法,在PyTorch框架下进行训练和推理,整个平台可以实现手势识别流程的一体化、实时化。具体的手势识别流程如图1所示。



图1 手势识别流程

本次的实验平台如下。台式机:Windows10 64位操作系统, Inter®Core™ i7-8700K CPU @3.70 GHz, 48 GB内存,显卡型号NVIDIA GeForce GTX1080 Ti)。选用的FMCW雷达是Infineon公司的BGT60TR13C毫米波雷达模块。实验采用的雷达配置如表1所示,最终性能指标如表2所示。

表1 雷达参数配置

配置项目	配置参数
TX最低频率	58.13 GHz
TX最高频率	62.96 GHz
带宽B	4.84 GHz
Chirp数量	32
每个Chirp的Sample数量	64
TX power	31
IF gain	50 dB
激活天线数	3
A/D采样速率	1 000 kHz

数据采集实验的流程如下:首先通过PC机给雷达配置相应的参数,然后持续读取雷达传回的数据;志愿者在雷达上面进行规定的手势动作的操作,PC机将回

表 2 雷达性能指标

配置项目	配置参数
最远探测距离	0.096 m
距离分辨率	0.03 m
检测最大速度	1.2 m/s
速度分辨率	0.05 m/s
帧率	31.1 fps

波数据保存至磁盘作为数据集,通过文件名区分类别.具体的数据采集场景如图2所示.

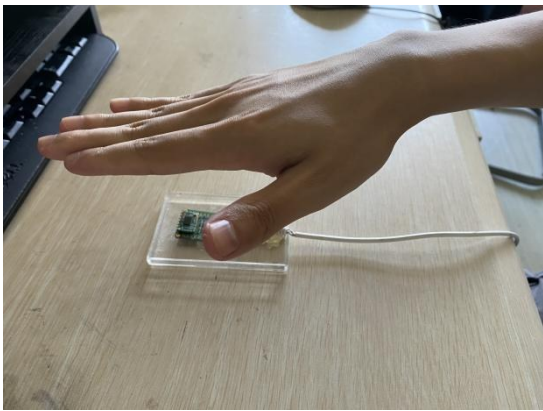


图2 数据采集场景

本文设定了如图3所示的5种基础手势,分别为转圈、双指按压、抓取、呼叫、挥手.实验招募了8位志愿者,对于5种手势各自采集10次数据,总计 $8 \times 5 \times 10 = 400$ 条数据.每条数据包含3个通道的回波信号.数据集的开源地址为 [https://gitee.com/xiao\\_chens\\_classmates/Radar\\_Gesture\\_Data.git](https://gitee.com/xiao_chens_classmates/Radar_Gesture_Data.git).



图3 手势动作示意图

### 3 整体算法流程

本文整体算法流程如下:首先,在距离和速度维度对手势回波信号进行滤波,有效地降低了系统的静态噪声;其次,将数据经过MTI算法滤除时间维度噪声;然后,提出了时间自适应定长化的方法,在减少手势信息损失的前提下保证了每个手势样本帧数的一致性;最后,建立RD-Net进行训练分类.

### 3.1 双维度滤波计算RDI

FMCW 雷达得到的原始回波信号是发射信号和接收信号混频后的中频信号,中频信号的频率是接收和发射信号的频率差.

动态的手势识别相比静态的识别,是具有时间维度的.具体而言,针对雷达的单个接收通道,对于一次手势动作实验可以获取到多帧数据.由于FMCW 雷达信号的特性,一帧数据由多个线性调频脉冲(Chirp)组成.Chirp 其中独立的一个部分,经过ADC 采样后变成多个采样(Sample)值.

3个通道同一时刻对应Chirp的数据情况如图4所示.可以看出,3个通道的数据差异较大,这主要是由天线的排列问题,即接受天线的位置不同造成的.

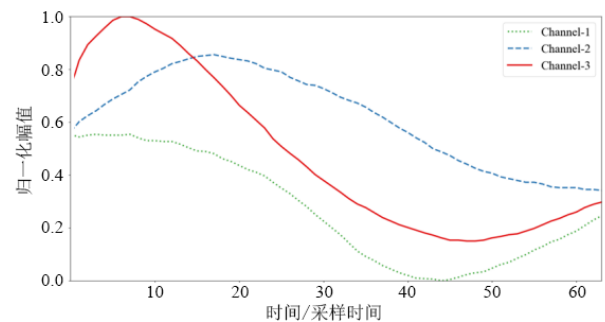


图4 3通道雷达数据

实验将采集的原始数据集进行随机森林分类,可以发现其准确率很低,5种手势的准确率均在20%左右.这说明原始数据中存在大量干扰信息,导致分类器没有起到作用,因此本文需要对数据进行一定的预处理.

实验将一帧数据按照Chirp数目与Sample数目相乘排列组成一个矩阵 $D(c, s)$ ,通过对该矩阵运算得到RDI,具体操作如图5所示,对每个Chirp对应的数字化采样点执行距离FFT(Range FFT),输出结果以连续行的形式存储在矩阵中.处理器接收并处理单帧中各个Chirp后,开始对整个Chirp串序列进行多普勒FFT(Doppler FFT).Range FFT和Doppler FFT的联合操作可视作每帧对应数字化采样点的二维FFT.

二维FFT可同时分辨出目标的距离和速度.纵坐标表示距离,横坐标表示速度,如图6所示,RDI的原点

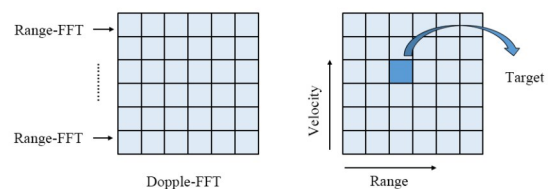


图5 RDI处理流程

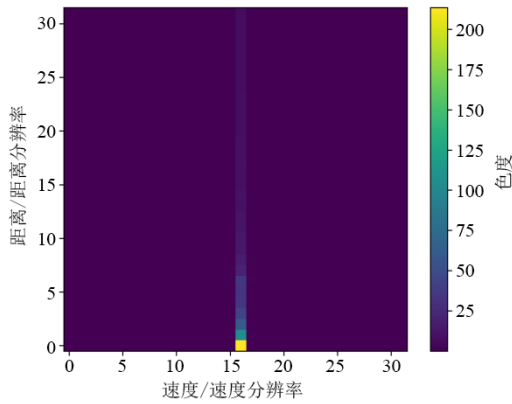


图6 原始RDI

通常位于底部的中间位置,因为距离没有负值,但速度可以有正负值.速度值为正表示该物体正在远离雷达系统,速度值为负表示该物体正在接近雷达系统.色度值的大小表示检测到对应距离对应速度的强度大小.二维FFT的峰值位置往往对应着雷达待测目标的距离和速度.

可以看出基于传统的两次FFT计算得到的RDI存在很强的静态分量,在此基础上进行手势识别实验是不可取的.因此,本文对该计算RDI的方式进行了优化,提出双维度滤波计算RDI的算法.本文RDI的计算流程选取如图7所示的关键节点特征图进行展示,主要分为以下6个步骤.

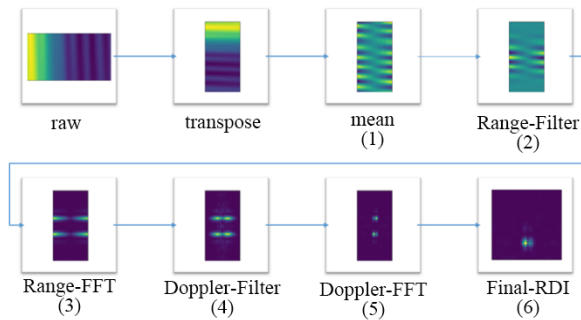


图7 双维度滤波计算RDI示意图

### (1) 去均值化

将  $D(c, s)$  首先沿 Chirp 维度取平均,然后扩展一个维度至二维,再将其按列放置后复制 Chirp 数目的份数,最后将原始输入数据转置后减去这一平均数据得到去均值化数据.去均值的作用主要在于减少不同人员手部反射强度的差异,从而提高手势识别在不同人员上的泛化性.

### (2) 距离维度滤波

接下来需要进行滤波,首先采用一个距离窗函数对数据进行处理.窗函数选取 Blackman 窗.对距离特征而言,该窗函数旁瓣小这一特点有利于解决误触发

问题.式(1)为该窗的频谱函数:

$$W[\omega] = 0.42W_R(\omega) - \frac{0.5}{2} \left[ W_R\left(\omega - \frac{2\pi}{N}\right) + W_R\left(\omega + \frac{2\pi}{N}\right) \right] + \frac{0.08}{2} \left[ W_R\left(\omega - \frac{4\pi}{N}\right) + W_R\left(\omega + \frac{4\pi}{N}\right) \right] \quad (1)$$

其中,  $W_R(\omega)$  如下所示:

$$W_R(\omega) = \exp\left(-j\frac{N-1}{2}\omega\right) \frac{\sin(N\omega/2)}{\sin(\omega/2)} \quad (2)$$

生成的窗函数通过复制的方式扩展至 Chirp 数目与 Sample 数目相乘的二维大小,将去均值化后的数据经过该窗函数滤波.

### (3) 距离维度 FFT

将上述数据进行距离维度 FFT,并将零频点移到频谱的中间.

### (4) 速度维度滤波

速度维度采用切比雪夫窗函数进行滤波.对于微多普勒特征而言,该窗函数的最大振幅比能够起到较好的效果.其频谱函数如式(3)所示:

$$W(k) = \frac{\cos\left\{M \cos^{-1}\left[\beta \cos\left(\frac{\pi k}{M}\right)\right]\right\}}{\cosh\left[M \cosh^{-1}(\beta)\right]} \quad (3)$$

其中,  $M$  为 Sample 的数目,  $0 \leq \text{abs}(k) \leq M-1$ ,  $\beta$  如式(4)所示:

$$\beta = \cosh\left[\frac{1}{M} \cosh^{-1}\left(10^{\frac{A}{20}}\right)\right] \quad (4)$$

其中,  $A$  是输入的增益 dB 值.

生成的窗函数通过复制的方式扩展至 Chirp 数目与 Sample 数目相乘的二维大小,将上述数据经过该窗函数滤波.

### (5) 速度维度 FFT

将上述数据进行速度维度 FFT,并将零频点移到频谱的中间.

### (6) 最终结果

经过两次滤波后,算法将距离和速度维度的静态噪声滤除.由于 FFT 计算存在对称性,实验保留上述流程输出数据的上半部分作为特征提取结果.最终得到的 RDI 如图 8 所示.

从图 8 可以看出,优化后的计算方法通过去均值化、距离和速度维度滤波等手段,使手势动作更为突出,特征信息更为明显.

将本文提出的双维度滤波算法和传统的二维 FFT 算法提取 RDI 的方式进行对比实验.实验将两种方式提取的单通道 RDI 数据作为手势特征,采用截断补零的方式,通过简单的三层卷积神经网络进行分类.最终的分类准确率如表 3 所示.

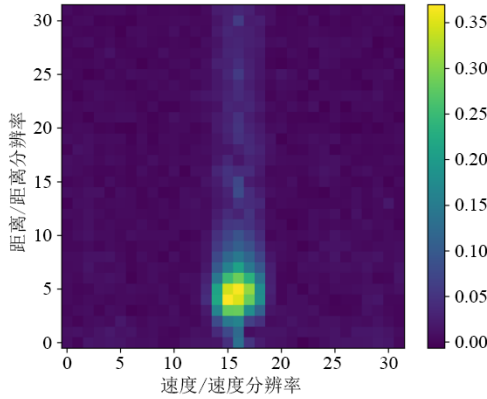


图8 双维度滤波后的RDI

表3 RDI提取方式准确率对比

RDI提取方式	准确率/%
二维FFT	72.5
双维度滤波	77.5

实验结果表明,相比传统的二维FFT计算RDI的方式,本文提出的双维度滤波算法的准确率提升了5%。

在上述RDI运算中,本文通过去均值化和距离速度双维度滤波的操作,使单帧数据构建的RDI中减少了大量静态噪声.但是在时间这一维度上应当考虑如何滤除静态分量,使动态的手势信息更为突出。

### 3.2 静态目标滤除

MTI算法是一种有助于区分运动目标、静态目标和杂波的雷达目标检测算法。MTI算法存在不同的实现方式,其中二维动目标检测(Two Dimensional Moving Target Indicator, 2DMTI)算法的核心思想在于,通过将每一帧的数据减去一定比例 $\alpha$ 的时间域平均数据(过往时间叠加得到的平均数据)的方法来排除静态目标,具体如式(5)所示:

$$D_{MTI} = D - \alpha D_{avg} \quad (5)$$

其中, $D$ 为当前输入算法的数据; $D_{avg}$ 为过往时间叠加得到的平均数据; $D_{MTI}$ 为经过算法滤除后的输出数据。

和传统2DMTI相比,本文所使用的算法的优势在于,其平均RDI的计算方式中,过去的的数据帧存在一个衰减过程,即越早的时间帧对当下的影响越小,符合客观规律。算法将每帧的RDI减去当前累计的平均RDI,该平均RDI的计算方式不再是单纯地将每帧数据累计取平均值,而是通过一个比例系数 $\beta$ 进行均衡,具体如式(6)、式(7)所示:

$$D_{MTI} = D - D_{avg} \quad (6)$$

其中, $D$ 为每帧实时输入算法的RDI; $D_{MTI}$ 为经过2DMTI算法滤除后的RDI; $D_{avg}$ 为累次运算得到的平均RDI。

$$D'_{avg} = \beta D + (1 - \beta) D_{avg} \quad (7)$$

其中, $D'_{avg}$ 为更新后的平均距离多普勒图,过滤器比例

系数 $\beta$ 需要在使用算法时指定,本文取0.95。

算法滤除后得到的RDI如图9所示,经过2DMTI算法处理后的RDI中明显减少了时间维度存在的静态噪声.这有助于神经网络更好地提取到手势的特征信息,后续实验都将使用2DMTI算法处理后的RDI。

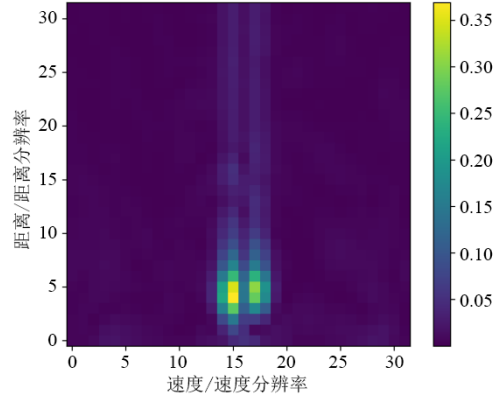


图9 静态目标滤除后的RDI

### 3.3 时间自适应定长化

动态手势识别的准确率,主要依赖手自动作的的时间维度信息.而这样采集的数据会存在时间不定长的问题.具体而言,由于手势采集过程中志愿者做动作的速度不同,最终采集的数据帧数会有一定差异.作为深度学习的输入,往往需要的是一个维度一致的数据,因此需要在尽可能减少信息损失的前提下,将手势特征数据维度一致化。

本文采用类似内插法的方式进行时间自适应定长化.内插法的作用就是采用一组已知的变量及其函数值来拟合一个近似函数,将这个函数的值作为该区间内所求未知函数的近似值.常用的内插法为多项式插值法,其中拉格朗日插值法的具体流程如下所述。

在求解平面上存在 $n$ 个点,分别为 $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ .假设该 $n$ 个点的 $x_i$ 均不相同,可以推出拉格朗日插值多项式如式(8)所示:

$$L(x) = \sum_{i=0}^n y_i l_i(x) \quad (8)$$

其中, $l_i(x)$ 为拉格朗日基本多项式,其具体如式(9)所示:

$$\begin{aligned} l_i(x) &= \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j} \\ &= \frac{(x - x_0)}{(x_i - x_0)} \times \dots \times \frac{(x - x_{i-1})}{(x_i - x_{i-1})} \\ &\quad \times \frac{(x - x_{i+1})}{(x_i - x_{i+1})} \times \dots \times \frac{(x - x_n)}{(x_i - x_n)} \end{aligned} \quad (9)$$

上述拉格朗日基本多项式的 $l_i(x)$ 特点如式(10)

所示:

$$l_i(x_k) = \begin{cases} 1, & k=i \\ 0, & k \neq i \end{cases} \quad (10)$$

拟合出  $L(x)$  后, 在该区间内的  $f(x)$  其他点的值都用这条拟合函数  $L(x)$  的值作为近似值.

在内插法中函数区间往往是不会变动的, 拟合近似函数的目的在于获取该区间内近似值. 但在具体手势识别场景下, 实验的目的在于能够将数据帧数从不定长转化为定长. 当预设了一个帧数阈值  $T_{th}$  时, 真实的手势数据帧数  $T_{re}$  会出现大于或小于阈值这两种需要处理的情况, 因此本文设计的算法流程图如图 10 所示.

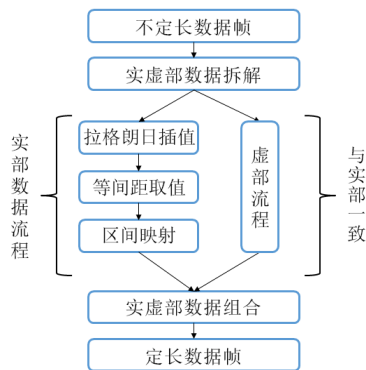


图 10 时间自适应定长化

将帧数不一致的 RDI 特征分解为实部和虚部两部分, 对这两部分分别进行此后的操作. 以实部为例, 此时在求解空间内存在  $T_{re}$  个点, 其中自变量空间的集合  $T_n = [1, 2, \dots, T_{re}]$ , 函数变量值为对应帧实部的数据  $I_n$ , 将这组数据通过拉格朗日插值法拟合出对应的多项式  $L(x)$  作为近似函数. 根据目标帧数  $T_{th}$  将当前的  $[1, T_{re}]$  区间进行等分. 等分后对每一段的始末点编号, 将其区间映射成为  $[1, T_{th}]$ , 并代入  $L(x)$  获取对应的函数值作为近似解. 此时实部的自适应定长化完成, 对于 RDI 虚部的做法也是如此.

时间阈值  $T_{th}$  选取的值决定手势动作的数据量. 本文进行实验判定如何合理地选取该值. 实验将提取的单通道 RDI 数据作为手势特征, 设置不同的阈值参数后, 通过简单的三层卷积神经网络进行分类. 最终的分类准确率如图 11 所示.

阈值参数的选取上, 实验以手势数据集中最长的手势持续帧数  $L_{max}$  为基准. 实验表明: 当帧数阈值选取为  $L_{max} \times 75\%$  时, 分类的准确率已经达到 85% 以上, 且之后增加阈值提升的准确率并不明显. 而选取的阈值越大, 算法需要的运算时间也就越长. 因此本文最终选择  $L_{max} \times 75\%$  的长度作为自适应定长化的阈值. 在此阈

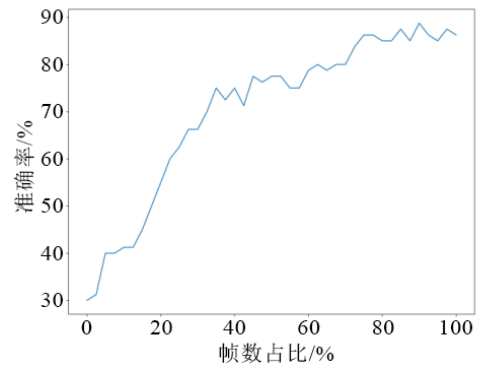


图 11 帧数阈值选取

值情况下, 传统截断补零算法和自适应定长算法对手势识别率的实验结果如表 4 所示.

表 4 手势帧数处理方式准确率对比

手势帧数处理方式	准确率/%
截断补零	77.50
自适应定长化	86.25

截断补零算法的准确率为 77.50%, 相比于此, 本文提出的自适应定长化算法的准确率提升了 8.25%.

### 3.4 神经网络算法设计

在最后的分类算法上, 相比多层感知机这种简单直接的形式, 本文设计了一种比传统神经网络更加适用于雷达信号处理的 RD-Net, 该网络的主体采用的是 2D-CNN 的结构, 2D-CNN 是一种提取特征的有效手段. RD-Net 有以下 2 个优势.

(1) 本次实验采用的 FMCW 雷达的中频信号是多通道的, 与传统的机器学习和多层感知机等算法的 CNN 结构相比, 本文的 CNN 结构对多通道信号的处理方式更加符合底层逻辑, 使前面分别根据多通道提取的信号特征在通过分类算法时不受到损失.

(2) 本文采用的 RDI 的特征类比于图像, 通过 CNN 处理的方式更加丰富, 可以借鉴计算机视觉相关的理论提升实验的效果. 模型整体架构如图 12 所示.

每个方块代表一个或一组特定操作, ReLU 是激活函数, Conv 是二维卷积层. Dil-Conv 是二维空洞卷积层. 模型的输入一次有效手势的双通道数据. 其中一个通道采用 convolution kernel 大小为 3 的网络层和 max pooling 操作快速地对输入数据进行降维, 另一个通道采用二维空洞卷积的方式, 没用使用 pooling 操作, 空洞卷积<sup>[15]</sup>的作用在于扩大感受野以及捕获多尺度上下文信息, 不同的 dilation 值旨在获取不同的感受野, 兼顾整体特征和局部特征. 两个通道互为补充, 最终将这些数据 concat 到一起, 经过 flatten layer 拉平, 送入 full connect layer 进行分类任务. 实验设置的学习率为  $10^{-4}$ , 优

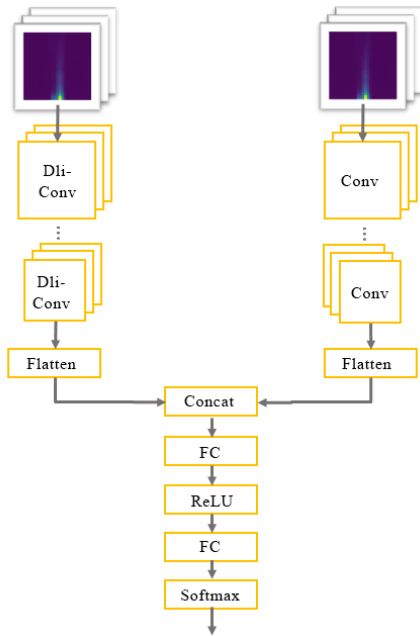


图 12 模型整体结构

化器采用 Adam<sup>[16]</sup>, batch size 为 4, 损失函数采用交叉熵损失函数 CrossEntropyLoss.

### 4 实验结果分析

选取采集的数据集中任意两个通道的数据, 按照上文的预处理对原始数据进行操作. 训练集和测试集的分割上, 随机选取数据集中 80% 的数据作为训练集, 20% 的数据作为测试集, 分别使用随机森林算法、多层感知机算法和 RD-Net 算法进行测试. 最终得到的准确率如表 5 所示.

表 5 不同算法的准确率对比

算法模型	准确率/%
随机森林	87.5
多层感知机	92.5
RD-Net	100

可以发现, 相比直接使用原始数据, 本文提出的预处理方法更能有效提取手势动作的特征. 而数据集在 RD-Net 中表现为完全正确, 其原因与数据集量级不够大、产生一定程度的过拟合有关.

为了验证算法的泛化性, 本文进一步在开源的 deep-soli 数据集<sup>[16]</sup>上进行测试. 该数据集共有 11 类手势, 每类手势有 475 条数据, 共计  $11 \times 475 = 5225$  条数据, 具体手势类别如图 13 所示.

随机选取数据集中 80% 的数据作为训练集, 20% 的数据作为测试集. 数据集中每帧数据都以一张  $32 \times 32$  大小的 RDI 的形式呈现, 实验将其进行上文中的 2DMTI 预处理和时间自适应定长化, 通过 RD-Net 神经

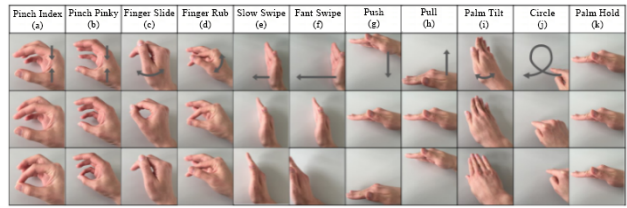


图 13 deep-soli 手势数据集

网络进行分类. 最终的准确率如表 6 所示.

表 6 在 deep-soli 数据集上的准确率

手势 ID	0	1	2	3
准确率/%	88.4	100.0	96.8	94.7
手势 ID	4	5	6	7
准确率/%	100.0	100.0	100.0	100.0
手势 ID	8	9	10	11
准确率/%	100.0	100.0	97.9	98.3

测试结果表明本文算法的平均识别准确率可以达到 98.28%. 而该数据集开源团队采用轻量级端到端 CNN 和 LSTM 架构, 在开源数据集上的准确率为 87.17%. 因此, 本文提出的识别算法在 deep-soli 开源数据集上的准确率提升了 11.1%. 其混淆矩阵如图 14 所示.

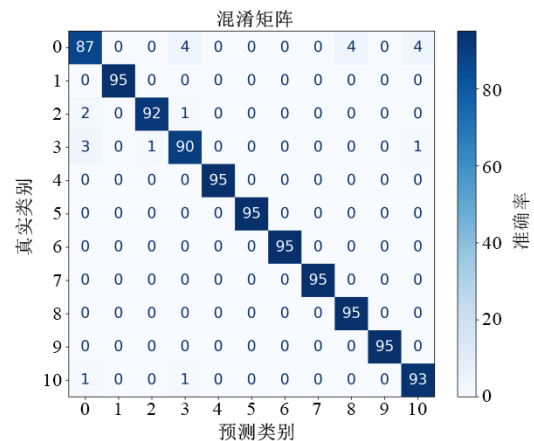


图 14 deep-soli 数据集的混淆矩阵

实验表明, 本文设计的预处理方法与 RD-Net 神经网络分类算法在数据量级较大的开源手势数据集上依然具有良好的识别表现, 并不存在明显的过拟合现象.

### 5 实时推理的实现

本文将上面通过实验采集数据建立的训练集训练得到的神经网络模型及其权重保存用于之后的实时推理实验. 推理的实验场景与数据集采集建立的情况一致. 实验招募了 10 位新的志愿者, 分别编号为 A~J, 对上述 5 种手势各进行 10 次实时推理的测试. 最终的准

准确率如表7所示。

身份ID	0	1	2	3	4	平均值
A	100	90	90	80	90	90
B	100	90	90	100	80	92
C	100	90	90	80	90	90
D	90	100	90	80	90	90
E	90	90	80	90	90	88
F	90	90	100	90	100	94
G	90	90	80	90	90	88
H	100	80	100	90	100	94
I	100	80	90	100	100	94
J	100	80	90	90	80	88
平均值	96	88	90	89	91	90.8

实验分别针对手势和志愿者两个维度,对模型算法的准确率进行分析,达到的手势识别平均准确率为90.8%,相比在数据集上测试的效果降低了9.2%。其混淆矩阵如图15所示。

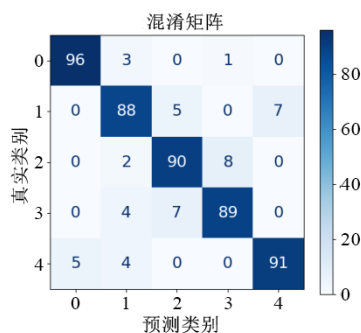


图15 实时推理的混淆矩阵

实验结果表明,手势识别对不同人员给出的手势具有一定的敏感性,例如在识别不同人员的第一类手势时,识别准确率会在90%~100%之间浮动。对比5类手势的不同人员分类的平均准确率,最高达到94%,最低为88%,识别准确率有6%的动态范围。

邀请新的志愿者进行手势识别的实时推理实验,相比在数据集上测试的效果会差一些。出现这一情况主要是因为,每个志愿者手势动作不一致,以及雷达本身的变化、外界环境的变化等。实时推理实验中准确率能达到90.8%,这已经符合预期的实验要求了。

## 6 结语

基于FMCW雷达的手势识别,不受光照的影响,数据量级小,易于处理,有效地保障了用户隐私,具有很好的发展前景。本文提出的手势识别算法,在开源数据集deep-soli上取得了98.3%的准确率。通过搭建的数据采集实验平台,进行手势识别实时推理实验,也取得

90.8%的平均准确率。这些充分验证了本文算法的有效性和实用性。

**致谢** 本工作得到浙江省协同感知与自主无人系统重点实验室和浙江省重点研发计划(No.2022C01028)的支持。

## 参考文献

- [1] CARD S K, MORAN T P, NEWELL A. The Psychology of Human-Computer Interaction[M]. Boca Raton: CRC Press, 2018.
- [2] RAUTARAY S S, AGRAWAL A. Vision based hand gesture recognition for human computer interaction: A survey [J]. Artificial Intelligence Review, 2015, 43(1): 1-54.
- [3] 陆霖霖, 江春华, 郝宗波. 基于不同光照条件的人体手势识别新方法[J]. 计算机应用, 2015, 35(S1): 273-277, 291. LU L L, JIANG C H, HAO Z B. Human gesture recognition new approach based on different light conditions[J]. Journal of Computer Applications, 2015, 35(S1): 273-277, 291. (in Chinese)
- [4] YEO H S, FLAMICH G, SCHREMPF P, et al. RadarCat: radar categorization for input & interaction[C]//Proceedings of the 29th Annual Symposium on User Interface Software and Technology. New York: ACM, 2016: 833-841.
- [5] YEO H S, MINAMI R, RODRIGUEZ K, et al. Exploring tangible interactions with radar sensing[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(4): 1-25.
- [6] HOF E, SANDEROVICH A, SALAMA M, et al. Face verification using mmWave radar sensor[C]//2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). Fukuoka: IEEE, 2020: 320-324.
- [7] TAHMOUSH D, SILVIOUS J. Radar micro-Doppler for long range front-view gait recognition[C]//2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems. Washington: IEEE, 2009: 1-6.
- [8] VANDERSMISSEN B, KNUDDE N, JALALVAND A, et al. Indoor person identification using a low-power FMCW radar[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(7): 3941-3952.
- [9] LIEN J, GILLIAN N, et al. Soli: Ubiquitous gesture sensing with millimeter wave radar[J]. ACM Transactions on Graphics, 2016, 35(4): 1-19.
- [10] SUH J S, RYU S, HAN B, et al. 24 GHz FMCW radar system for real-time hand gesture recognition using LSTM[C]//2018 Asia-Pacific Microwave Conference (APMC). Kyoto: IEEE, 2018: 860-862.
- [11] SMITH K A, CSECH C, MURDOCH D, et al. Gesture

recognition using mm-wave sensor for human-car interface[J]. IEEE Sensors Letters, 2018, 2(2): 1-4.

- [12] 王勇, 吴金君, 田增山, 等. 基于FMCW雷达的多维参数手势识别算法[J]. 电子与信息学报, 2019, 41(4): 822-829.

WANG Y, WU J J, TIAN Z S, et al. Gesture recognition with multi-dimensional parameter using FMCW radar[J]. Journal of Electronics & Information Technology, 2019, 41(4): 822-829. (in Chinese)

- [13] 夏朝阳, 周成龙, 介钧誉, 等. 基于多通道调频连续波毫米波雷达的微动手势识别[J]. 电子与信息学报, 2020, 42(1): 164-172.

XIA Z Y, ZHOU C L, JIE J Y, et al. Micro-motion gesture recognition based on multi-channel frequency modulated continuous wave millimeter wave radar[J]. Journal of Electronics & Information Technology, 2020, 42(1): 164-172. (in Chinese)

- [14] WANG S W, SONG J, LIEN J, et al. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum[C]//Proceedings of the 29th Annual Symposium on User Interface Software and Technology. New York: ACM, 2016: 851-860.

- [15] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015)[2021]. <https://arxiv.org/abs/1511.07122>.

- [16] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014)[2021]. <https://arxiv.org/abs/1412.6980>.

#### 作者简介



陈君毅 男, 1997年1月出生, 浙江嘉兴人. 浙江大学硕士研究生. 主要研究方向为毫米波雷达感知技术.

E-mail: chen19858874807@163.com



蒋德琛 男, 1997年3月出生, 浙江宁波人. 浙江大学硕士研究生. 主要研究方向为毫米波雷达应用.

E-mail: dechen\_j@163.com