



研究与开发

基于毫米波雷达的脚踢动作识别

周晓昀, 张远辉, 陈邵靖, 郑超群

(中国计量大学机电工程学院, 浙江 杭州 310018)

摘要: 近年来, 基于毫米波雷达的脚踢动作识别作为一种免手操作的人机交互技术, 在智能家居和车载应用等领域展现出重要应用价值。然而, 复杂环境下的静态干扰(如墙壁、石柱及天线耦合干扰)和动态干扰(如行人运动、肢体微动)仍对识别精度构成挑战。为实现高精度、稳健识别, 提出一种融合干扰抑制与深度学习的识别方法。该方法通过向量均值相消与动目标显示(moving target indication, MTI)抑制静态干扰, 并结合多普勒加权、恒虚警率(constant false alarm rate, CFAR)检测、密度聚类(density-based spatial clustering of applications with noise, DBSCAN)以及连通域约束消除动态干扰。随后, 提取多帧距离-多普勒图(range-Doppler map, RDM)和距离-角度图(range-angle map, RAM)作为模型输入, 构建基于卷积神经网络(convolutional neural network, CNN)、多头自注意力(multi-head self-attention, MHSA)以及简化时序卷积网络(simplified temporal convolutional network, STCN)的双流CNN-MHSA-STCN模型, 用于完成动作识别。实验表明, 该方法在自采数据集上的识别精度超过98%, 在复杂环境中具备较高的精度与鲁棒性。

关键词: 毫米波雷达; 脚踢动作识别; 干扰抑制; 双流CNN-MHSA-STCN

中图分类号: TN958; TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2026034

Kick motion recognition based on millimeter-wave radar

Zhou Xiaoyun, Zhang Yuanhui, Chen Shaojing, Zheng Chaoqun

College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China

Abstract: In recent years, millimeter-wave radar-based kick motion recognition has attracted significant attention as a hands-free human-computer interaction technology, demonstrating considerable value in applications such as smart home systems and automotive interfaces. However, in complex environments, both static interference (e.g., walls, pillars, and antenna coupling effects) and dynamic interference (e.g., pedestrian movements and subtle limb motions) continue to present substantial challenges to recognition accuracy. To achieve high-precision and robust recognition, an integrated method combining interference suppression and deep learning was proposed. The method employed vector mean cancellation and moving target indication (MTI) to suppress static interference. In contrast, dynamic interference was mitigated through a combination of Doppler weighting, constant false alarm rate (CFAR) detection, density-

收稿日期: 2025-08-12; 修回日期: 2025-11-06

通信作者: 张远辉, zyh@cjlu.edu.cn

基金项目: 浙江省自然科学基金资助项目 (No.LY19F010007)

Foundation Item: The Natural Science Foundation of Zhejiang Province of China (No.LY19F010007)



based spatial clustering of applications with noise (DBSCAN), and connected region constraint. Subsequently, multi-frame range-Doppler maps (RDM) and range-angle maps (RAM) were extracted to serve as model inputs. A dual-stream CNN-MHSA-STCN architecture was constructed, incorporating a convolutional neural network (CNN), multi-head self-attention (MHSA), and a simplified temporal convolutional network (STCN) for comprehensive motion recognition. Experimental results demonstrate that the proposed method achieves a recognition accuracy exceeding 98% on a self-collected dataset, while maintaining high precision and robustness in complex operational environments.

Key words: millimeter-wave radar, kick motion recognition, interference suppression, dual-stream CNN-MHSA-STCN

0 引言

近年来,免手操作技术受到越来越多的关注,脚踢动作识别在智能家居^[1]和汽车应用^[2-3]等领域得到了广泛发展。通过对脚踢动作的识别,用户可以在双手被占用或不便操作的情况下完成设备控制,提升便利性和安全性。然而,现有的人体动作检测传感器,如超声波传感器^[4-5]、电容式传感器^[6-7]、图像传感器^[8]均存在一定的局限性。超声波传感器在雪、冰冻或泥沙环境下易失效;电容式传感器在雨天易误触发;图像传感器对光照敏感,在遮挡或恶劣天气下出现性能下降。相比之下,毫米波雷达凭借抗干扰能力强、环境适应性好及隐私保护优势,更适合脚踢动作识别任务。

在基于雷达的脚踢动作识别领域,已有多项研究成果。文献[9]利用超宽带(ultra-wideband, UWB)雷达生成时间-距离-多普勒踢腿动作序列,并将其输入由卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)组成的时空神经网络进行识别。文献[10]提出了一种基于24 GHz多普勒雷达的二维多普勒频谱图特征提取方法,并将提取的特征输入隐马尔可夫模型(hidden Markov model, HMM)对踢腿动作进行分类。然而,这些方法依赖的特征集较为有限,难以充分捕捉人体动作的复杂性,削弱了区分不同动作的能

力,从而限制了识别精度的进一步提升。

随着深度学习的广泛发展,大量先进模型不断涌现,大幅提升了各类识别任务的精度与效率。研究者们通过融合并改进不同架构的模型,如将CNN和长短期记忆(long short-term memory, LSTM)网络相结合构建混合模型以高效处理复杂的时空特征,从而提升识别精度。例如,文献[11]将多头自注意力机制与一维卷积网络相结合,实现了高效的实时手势识别。文献[12]设计了双流三维卷积神经网络(dual-stream three-dimensional convolutional neural network, DS-3DCNN)和LSTM网络相结合的DS-3DCNN-LSTM模型,从距离-多普勒图时间序列(range-Doppler-map time sequence, RDMTS)和距离-角度图时间序列(range-angle-map time sequence, RAMTS)中提取并融合特征,手势识别精度达到了97.66%。文献[13]提出了一种基于CNN-Transformer的动态手势识别方法,利用调频连续波(frequency-modulated continuous wave, FMCW)毫米波雷达在干扰场景下的识别精度超过98%,在随机动态干扰场景下的识别精度超过96%。文献[14]将FMCW毫米波雷达与CNN相结合,通过训练与测试微多普勒频谱图,手势识别效果获得显著提升。文献[15]基于毫米波雷达提取手势的距离和速度特征,利用残差网络进行特征融合,在不同测试环境和用户位置下实现了平均90%以上的手语分类精度。文献[16]提取微多普勒时域谱图特征,

引入卷积块注意模块 (convolutional block attention module, CBAM) 的残差网络进行识别, 实现了平均 98.17% 的人体行为分类精度。这些研究表明, 提取目标特征并通过网络模型进行处理, 是实现动作识别的有效方式, 但在应对动态干扰, 尤其是类目标干扰 (如行人运动) 时仍显不足, 影响了识别稳定性。因此, 提升毫米波雷达在复杂环境下的识别精度仍是关键挑战。

基于上述背景, 本文提出了一种基于毫米波雷达的脚踢动作识别方法, 该方法针对目标信号易受静态和动态干扰影响的问题, 提出了一套干扰抑制方法, 以提升信号质量。此外, 针对现有方法特征集较为有限、难以充分表征动作复杂性的问题, 本文提取多帧距离-多普勒图 (range-Doppler map, RDM) 和距离-角度图 (range-angle map, RAM) 特征, 并构建双流 CNN-MHSA-STCN 深度学习模型, 最终实现了复杂环境下脚踢动作的高精度识别。

1 基于毫米波雷达的脚踢动作识别方法框架

本文提出的基于毫米波雷达的脚踢动作识别方法总体框架如图 1 所示。该方法由雷达数据采集、雷达信号处理和基于深度学习的动作分类 3 个阶段组成。在数据采集阶段, 采用毫米波雷达和 LA1010 逻辑分析仪作为数据采集设备。采

集到的雷达数据需要经过多阶段信号处理流程: 对距离维度进行一维快速傅里叶变换 (1D fast Fourier transform, 1D-FFT) 以提取目标距离信息, 采用向量均值相消法抑制静态干扰; 对多普勒维度进行二维快速傅里叶变换 (2D fast Fourier transform, 2D-FFT) 以获取目标多普勒信息, 结合动目标显示 (moving target indication, MTI) 算法抑制天线耦合干扰, 至此, 获得了各接收通道的 RDM 后沿接收通道维度进行非相干累加, 生成最终的 RDM; 应用一系列干扰抑制算法抑制动态干扰; 对天线维度进行三维快速傅里叶变换 (3D fast Fourier transform, 3D-FFT) 以提取目标角度信息。由于雷达为垂直排列的 2 发 2 收天线结构, 这 4 个通道包含了目标的方位角和俯仰角信息。本文中, RAM 为方位角与俯仰角的联合能量分布, 该融合方法有效降低了数据维度和计算复杂度。处理后的 RDM 和 RAM 被输入双流 CNN-MHSA-STCN 模型进行训练, 以完成动作分类。

2 干扰抑制算法和深度学习模型

在毫米波雷达脚踢动作识别过程中, 信号会受到多种干扰, 可分为静态干扰和动态干扰。静态干扰包括由墙壁或柱子等产生的静态杂波以及天线耦合干扰。针对静态杂波, 雷达信号在经过 1D-FFT 后采用向量均值相消法去除, 而经过 2D-FFT 后再通过 MTI 算法抑制天线耦合干扰。动态

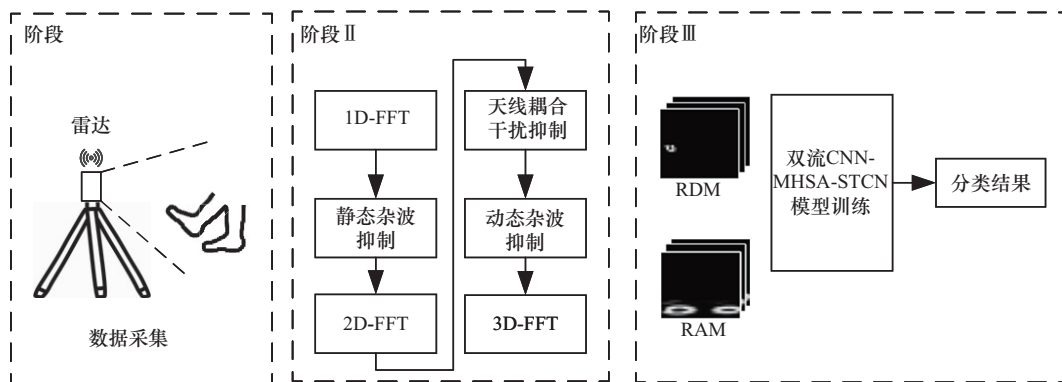


图 1 基于毫米波雷达的脚踢动作识别方法总体框架



干扰包括背景噪声、离散噪声以及类目标干扰（如行人运动）。背景噪声通过多普勒加权结合恒虚警率（constant false alarm rate, CFAR）算法滤除，剩余的离散噪声采用密度聚类（density-based spatial clustering of applications with noise, DBSCAN）算法去除，而类目标干扰则通过连通域约束算法滤除。经过干扰抑制后的信号进一步进行3D-FFT处理，以提取多帧RDM和RAM特征，最终输入CNN-MHSA-STCN模型，实现脚踢动作识别。

2.1 静态杂波抑制

为了抑制环境中静态物体（如墙壁和石柱）产生的静态杂波，本文在对原始雷达信号进行1D-FFT处理后，采用向量均值相消法进行静态杂波抑制，通过沿距离维度计算均值来估计静态杂波成分，并将其从原始信号中去除：

$$S'[a, m, r] = S[a, m, r] - \frac{1}{N_c} \sum_{n=1}^{N_c} S[a, n, r] \quad (1)$$

其中， S 表示雷达信号的三维数据立方体， a 表示接收天线的索引， m 表示每帧中的Chirp索引， r 表示距离单元索引， n 为Chirp的遍历索引变量， N_c 表示每帧中的Chirp总数。

静态杂波抑制前后的信号对比如图2所示。图2（a）展示了脚踢动作的原始时间-距离频谱图，可以看出，近距离区域存在明显的天线耦合干扰。此外，静态杂波和动态杂波严重遮蔽了目标动作。使用向量均值相消法后，有效抑制了静态杂波（在近距离处，并在时间轴上保持固定位置），使得脚踢动作的轮廓初步显现。经过静态杂波抑制后的信号如图2（b）所示。

天线耦合是指发射天线的信号未经目标反射而直接被接收天线接收的现象。这种直接耦合信号会在雷达系统中产生强烈的能量干扰，可能掩盖近距离目标的回波信号，因此需要加以抑制。由于天线耦合干扰在相邻帧之间几乎保持不变，呈现静态特性，本文采用MTI算法进行抑

制。该算法通过对连续帧进行差分，从而抑制天线耦合干扰。天线耦合干扰抑制后的信号如图3所示。从图3可以看出，耦合伪影显著减少，且目标信号得到增强。

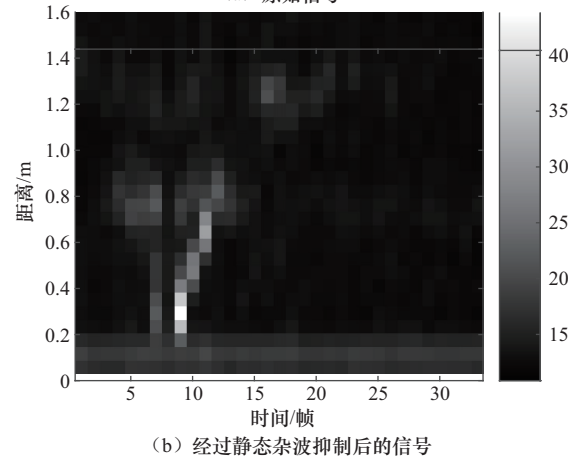
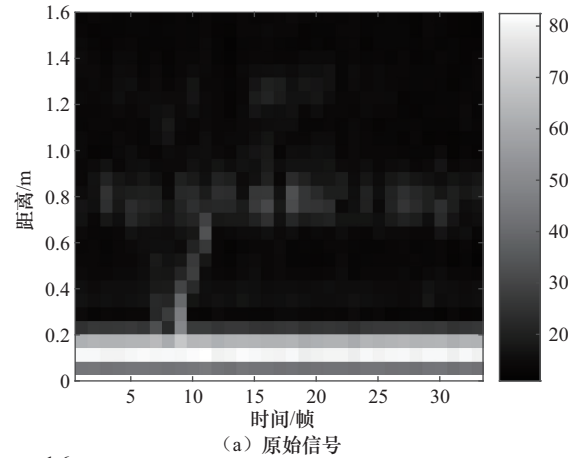


图2 静态杂波抑制前后的信号对比

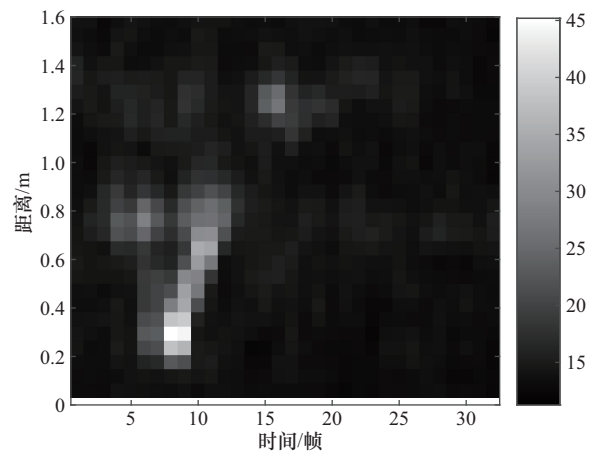


图3 天线耦合干扰抑制后的信号

2.2 动态杂波抑制

2.2.1 背景噪声抑制和目标检测

在抑制了静态杂波和天线耦合干扰后，目标信号仍然受到背景噪声、踢腿过程中其他肢体微动以及附近行人干扰等动态因素的影响，这些动态干扰降低了识别性能。为进一步提升目标检测的准确性，本文采用了多普勒加权抑制算法、CFAR^[17]算法和DBSCAN^[18]算法。

针对传统CFAR算法在复杂运动场景下易出现漏检的不足，本文提出了一种多普勒加权抑制算法。

在人体踢腿过程中，非运动部位（如站立腿）会产生低速高幅值信号，而主动踢腿部位则表现出高速高幅值特征，背景噪声则始终呈现低幅值特征。因此，本文通过设置幅值加权以抑制低幅值信号，并调整多普勒加权以增强高速信号同时抑制低速信号，具体步骤如下。

步骤1 计算2D-FFT得到的距离-多普勒矩阵的幅值，以提取每帧距离-多普勒单元的回波强度：

$$A[c, r] = |D[c, r]| \quad (2)$$

其中， D 表示通过对每帧雷达信号执行2D-FFT后得到的复数距离-多普勒矩阵， A 表示该矩阵的幅值矩阵， $c \in [1, N_d]$ 表示多普勒单元索引， $r \in [1, N_r]$ 表示距离单元索引， N_d 为多普勒单元总数， N_r 为距离单元总数，零多普勒单元位置为 $c_0 = [N_d/2]$ 。

步骤2 对每个距离单元 r 的幅值进行归一化处理：

$$\tilde{A}[c, r] = \frac{A[c, r]}{\max_{1 \leq k \leq N_d} A[k, r] + \varepsilon} \quad (3)$$

其中， k 为索引变量，用于遍历第 r 个距离单元下的所有多普勒单元（即该列中的每个元素）， $\varepsilon = 2.22 \times 10^{-16}$ 为机器精度常数，用以避免除零错误。

步骤3 计算每个多普勒单元相对于零多普勒单元的归一化距离：

$$\tilde{D}_v[c, r] = \frac{|c - c_0|}{\max_{1 \leq k \leq N_d} |k - c_0| + \varepsilon} \quad (4)$$

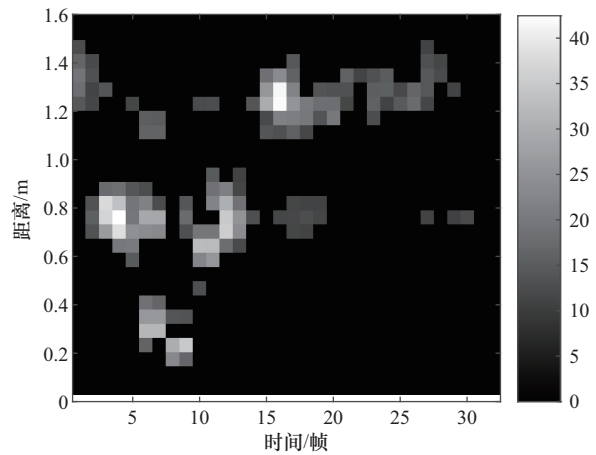
步骤4 通过线性组合归一化幅值和多普勒偏移，生成复合权重矩阵，权重系数设为 $\alpha=1$ 、 $\beta=1.5$ ：

$$W[c, r] = \alpha \tilde{A}[c, r] + \beta \tilde{D}_v[c, r] \quad (5)$$

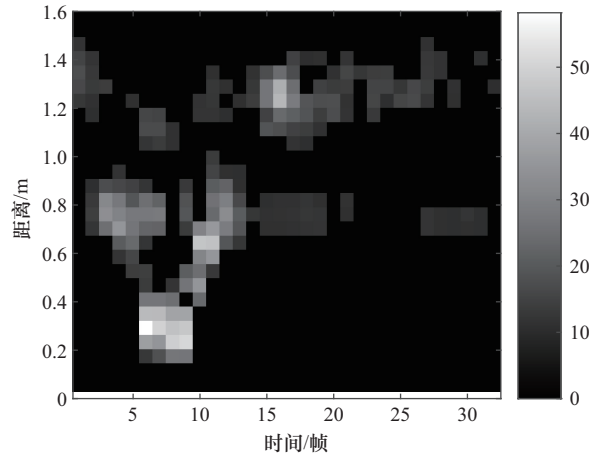
步骤5 将复合权重矩阵 W 与原始距离-多普勒矩阵 D 按元素相乘，得到加权后的距离-多普勒矩阵：

$$D'[c, r] = D[c, r] \cdot W[c, r] \quad (6)$$

CFAR算法通过动态调整检测阈值，以适应环境噪声的变化。在引入多普勒加权抑制后，目标信号更加突出，使阈值选择更为合理，从而有效降低误警率和漏检率。有无多普勒加权处理的CFAR检测结果对比如图4所示。从图4中可以看出，若直接将CFAR算法应用于未经多普勒加权抑制的信号，部分目标区域会被错误滤除，影响目标的完整性。



(a) 未经过多普勒加权处理



(b) 经过多普勒加权处理

图4 有无多普勒加权处理的CFAR检测结果对比



本文采用有序统计恒虚警率 (ordered statistic-CFAR, OS-CFAR) 算法进行目标检测与背景噪声抑制。该方法在距离维度和多普勒维度分别独立应用, 将两个维度的检测结果结合, 只有在两者中均被判定为目标的信号才被视为真实目标, 从而提升了检测的准确性与鲁棒性。

虽然 CFAR 算法能有效检测目标并抑制大部分背景噪声, 但仍可能存在残余的离散噪声点。为进一步优化信号处理, 本文采用了 DBSCAN 进行后续处理。经过 DBSCAN 处理后的信号如图 5 所示, 尽管离散噪声成分减少, 但由肢体微动和行人干扰引起的动态干扰依然存在。

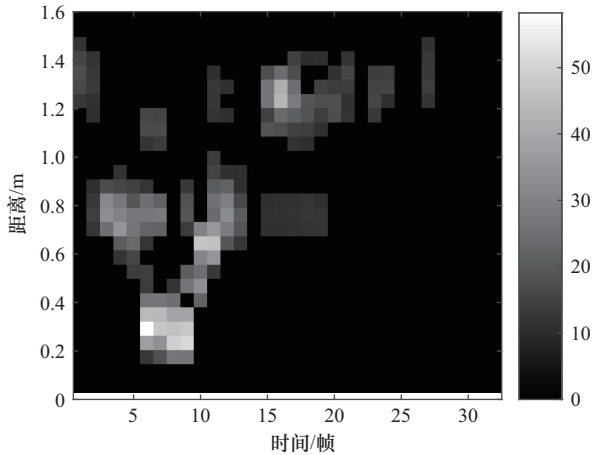


图5 经过DBSCAN处理后的信号

2.2.2 连通域约束算法

尽管上述背景噪声抑制和目标检测算法能够有效滤除大量噪声, 但其对动态干扰, 尤其是行人运动和肢体微动等类目标干扰的抑制能力仍显不足。因此, 本文提出了一种连通域约束算法。

当人体执行特定动作 (如踢腿或行走) 时, 目标在距离和时间维度上的变化通常具有一定的连续性, 不会出现突变。基于这一特性, 本文在算法设计中引入了距离和时间维度的连续性约束, 通过筛选符合连续变化特征的目标区域, 有效剔除潜在的动态干扰信号, 从而提升目标检测的准确性。

连通域约束算法示意图如图 6 所示。可以看

出, 目标信号在时间和距离维度上均表现出一定程度的连续性。连通域约束算法的步骤总结如下。

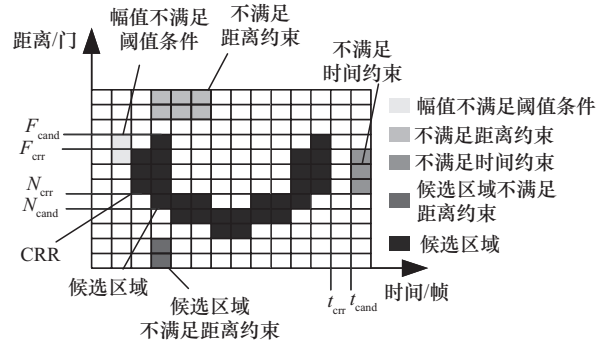


图6 连通域约束算法示意图

步骤 1 雷达信号幅值累积: 采集到的雷达信号数据形状为 $N_f \times N_d \times N_r$, 其中 N_f 为帧数, N_d 为多普勒单元数, N_r 为距离单元数。通过遍历所有帧, 对每一帧 $t \in [1, N_f]$ 的信号 $S(c, r)$ 在多普勒维度上进行累加:

$$T(r) = \sum_{c=1}^{N_d} S(c, r), r \in [1, N_r] \quad (7)$$

其中, $T(r)$ 表示第 t 帧的信号强度。

步骤 2 连通区域检测: 使用二值掩码 $M_T(r)$ 标记有效信号:

$$M_T(r) = \begin{cases} 1, & T(r) > 0 \\ 0, & T(r) \leq 0 \end{cases} \quad (8)$$

对 $M_T(r)$ 中非零区域进行连通区域检测, 对每帧中的每个连通区域分配唯一标识符。

步骤 3 当前参考区域 (current reference region, CRR) 种子选择: 遍历所有连通区域, 计算该连通区域的最大幅值:

$$\text{Amplitude}(k) = \max_{r \in \text{Region}(k)} T(r) \quad (9)$$

其中, k 表示当前帧中连通区域的标识符, $\text{Region}(k)$ 表示 k 连通区域包含的距离单元集合。然后, 选择满足以下条件的连通区域作为 CRR:

$$\text{Amplitude}(k) \geq \gamma \quad (10)$$

其中, γ 是 CRR 的幅值阈值。该阈值通过多个代

表性样本的幅值分布经验确定，以确保仅选择幅值显著的区域。

当检测到多个满足条件的连通区域时，优先选择距离雷达最近的区域作为CRR，其他不符合条件的区域则被过滤掉。

步骤4 CRR 候选区域的时空关联与更新：选定第一个CRR后，依次检查后续帧中的连通区域，候选区域同样需满足式(10)中的约束条件。计算候选连通区域与CRR之间的最小边缘距离，分为以下两种情况：

- 当连通区域在距离维度上发生重叠，即满足条件 $N_{\text{cand}} \leq F_{\text{crr}}$ 且 $F_{\text{cand}} \geq N_{\text{crr}}$ 时，最小边缘距离 d_{edge} 为0。
- 当连通区域在距离维度上不重叠时，最小边缘距离为：

$$d_{\text{edge}} = \min(|N_{\text{cand}} - F_{\text{crr}}|, |F_{\text{cand}} - N_{\text{crr}}|) \quad (11)$$

其中， N_{cand} 和 F_{cand} 分别为候选区域的近距离边界和远距离边界， N_{crr} 和 F_{crr} 分别为CRR的近距离边界和远距离边界。近距离边界表示距离雷达较近的位置，远距离边界表示距离雷达较远的位置。

若 $d_{\text{edge}} \leq \Delta d$ (最大允许边缘距离)，则保留该候选连通区域，并将其更新为新的CRR，同时更新其近距离和远距离边界。其中 Δd 的取值根据经验确定。

若当前帧未检测到有效的候选区域，则保留最近一次CRR的时间戳 t_{crr} ，在后续帧继续搜索满足上述条件的候选区域。

考虑目标应在连续时间内存在，进一步引入时间约束，新候选区域的时间戳 t_{cand} 必须满足：

$$\Delta t = |t_{\text{cand}} - t_{\text{crr}}| \leq \Delta T \quad (12)$$

其中， ΔT 为最大允许时间间隔，其取值根据实验经验确定。若该条件不满足，则后续帧中的候选区域将被丢弃。

连通域约束算法处理后的信号如图7所示。

从图7可以看出，连通区域约束算法能够有效抑制干扰，脚踢动作在处理后的信号变得更加清晰可见。

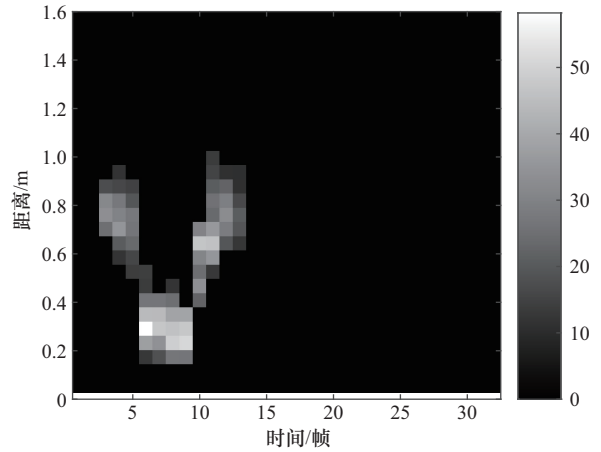


图7 连通域约束算法处理后的信号

2.3 深度学习模型

双流 CNN-MHSA-STCN 模型结构如图8所示。模型以每帧的RDM和RAM为输入，分别通过两个独立的CNN分支提取空间特征，并在时间维度上拼接降维。融合后的特征经多头自注意力 (multi-head self-attention, MHSA) 建模全局时序依赖，再由简化时序卷积网络 (simplified temporal convolutional network, STCN) 提取局部时序特征。最后，通过全局平均池化与全连接层完成动作分类。

2.3.1 CNN 结构

CNN能够通过卷积层自动学习输入数据的局部特征，有效处理复杂的空间信息。本文采用双流CNN架构，该架构分别由2个分支从RDM和RAM中提取空间信息。

分支1处理每帧的RDM数据，包含2个卷积模块和1个最大池化层，用于提取局部特征并降低空间分辨率。卷积模块的结构如图9所示。卷积模块由3×3卷积层 (Conv2d)、批量归一化层 (BatchNorm2d)、LeakyReLU激活函数和Dropout2d层组成。最大池化层使用2×2的卷积核和步长为2的参数设置，逐步对特征图进行下采样。

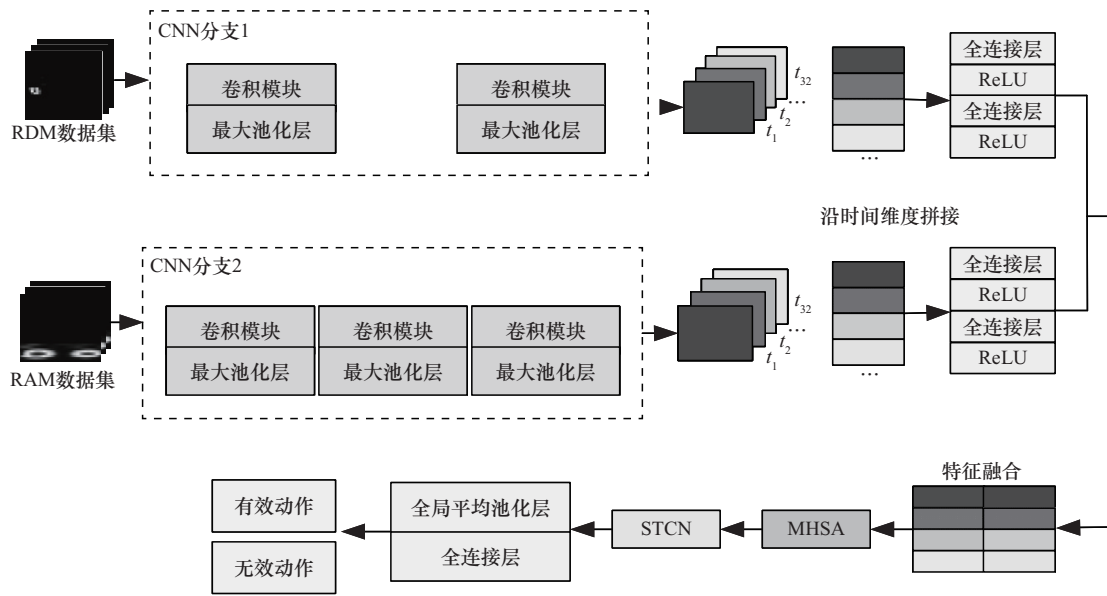


图8 双流CNN-MHSA-STCN模型结构

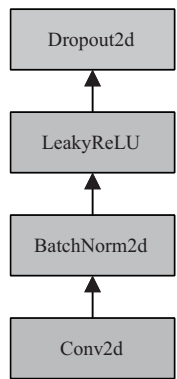


图9 卷积模块的结构

分支2处理每帧的RAM数据，其结构与分支1相似，但包含更多的卷积模块和池化层（共3个），以提取更深层的特征。

2.3.2 STCN 结构

时序卷积网络（temporal convolutional network, TCN）最初被设计用于序列建模任务，如今被广泛应用于时间序列信号处理和动作识别等领域^[19-20]。与传统的序列网络（如RNN和LSTM网络）相比，TCN通过扩张卷积捕捉长时依赖，有效避免了梯度消失问题。此外，TCN支持并行计算，提高了训练效率。

为进一步优化计算效率，本文对TCN结构进

行了改进，去除了残差连接和权重归一化，仅保留每层中的因果卷积、扩张卷积、LeakyReLU激活函数和Dropout层。STCN结构如图10所示。

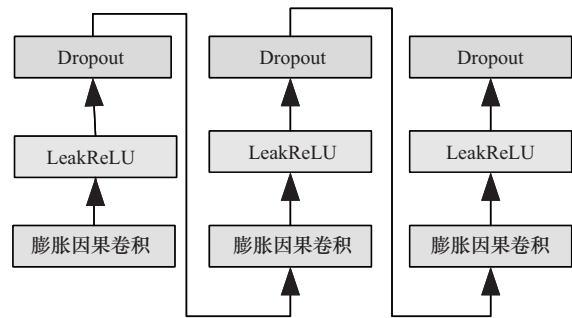


图10 STCN结构

2.3.3 MHSA 结构

在时序特征提取阶段，由于不同帧对脚踢动作识别的贡献不同，本文引入多头自注意力机制，以动态调整各帧的重要性，并捕捉时间序列中的关键特征。该机制广泛应用于Transformer模型^[21]中，通过在不同子空间并行计算注意力权重，从而更全面地建模时序依赖关系。本文将注意力头数量 h 设置为2。MHSA结构如图11所示，图11中 Q （Query）表示查询向量， K （Key）表示键向量， V （Value）表示值向量。

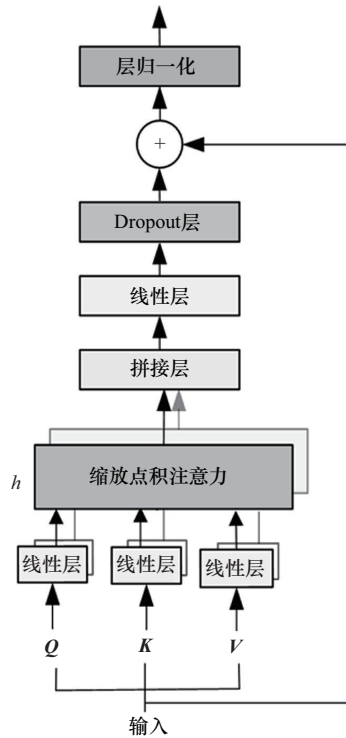


图 11 MHSa 结构

3 实验验证与分析

3.1 实验平台

实验设备如图 12 所示。实验的硬件系统由杭州岸达科技有限公司开发的 ADT3102 毫米波雷达模块、LA1010 逻辑分析仪和 J-Link 数据线组成。雷达配备了 2 个发射天线 (TX0, TX1) 和 2 个接收天线 (RX0, RX1), 采用时分复用多输入多输出 (time-division multiplexing multiple-input multiple-output, TDM-MIMO) 技术进行信号传输。ADT3102 雷达参数见表 1。



(a) ADT3102 雷达 (b) LA1010 逻辑分析仪 (c) J-Link 数据线

图 12 实验设备

表 1 ADT3102 雷达参数

雷达参数	数值
起始频率/GHz	77
带宽/GHz	4
Chirp 数/帧	32
采样点数/Chirp	256
角度范围	-60°~60°
ADC 采样频率/MHz	8.33
距离分辨率/m	0.05
速度分辨率/(m·s ⁻¹)	0.33

雷达模块负责信号采集及后续的模数转换 (analog-to-digital conversion, ADC) 处理。LA1010 逻辑分析仪配置为 500 MSa 采样深度和 100 MHz 采样率, 用于对采集的信号进行时序验证和协议分析。处理后的数据通过串行外设接口 (serial peripheral interface, SPI) 协议传输。同时, 模型训练在配备 NVIDIA RTX 3090 GPU 和 Intel Xeon Gold 6142 CPU 的高性能计算服务器上进行, 其强大的计算能力保证了训练的高效性。

3.2 数据采集

实验中, 雷达设备被安装在距地面 45 cm 的高度, 共有 12 名参与者 (7 名男性和 5 名女性) 参与数据采集。所有参与者被要求在雷达前方 1.6 m 范围内完成指定动作, 每个动作持续时间不超过 5 s, 共采集 1 904 个样本, 其中踢腿样本 443 个、摆腿样本 359 个、抬腿样本 374 个、下蹲样本 368 个、走路样本 360 个。

在数据处理阶段, 每帧的 RDM 和 RAM 特征被用作深度学习模型的输入特征。由于深度学习算法要求输入数据大小一致, 数据被截断或补齐至 32 帧, 其中 RDM 尺寸为 32×32, RAM 尺寸为 32×128。实验动作示意图如图 13 所示。

3.3 脚踢动作识别结果

为评估模型的性能, 本文将数据集划分为正类 (有效踢腿) 和负类 (摆动、抬腿、走路、下蹲)。由于有效踢腿样本数量远少于无效样本, 数据存在明显的类别不平衡, 传统交叉熵损失函

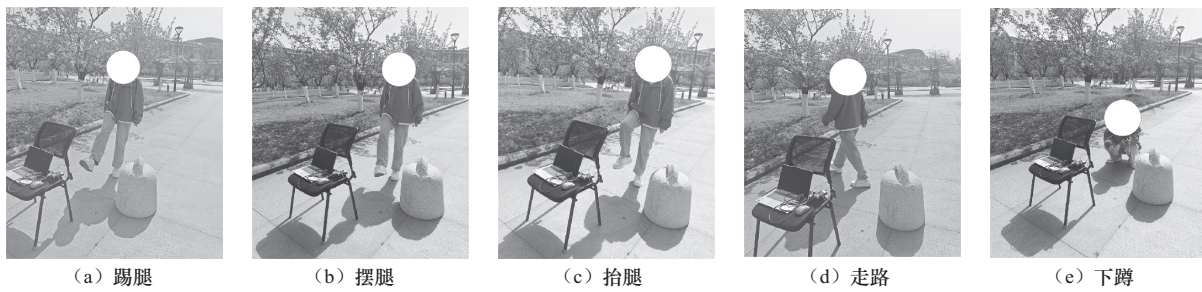


图13 实验动作示意图

数在训练中易受多数类主导，从而降低模型对少数类的识别能力。为解决类别不平衡问题，本文采用了 Focal Loss^[22]作为损失函数，定义如下。

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \ln(p_t) \quad (13)$$

其中， α_t 为类别平衡权重，用于调整正、负样本在损失计算中的相对重要性， $(1-p_t)^\gamma$ 为调节因子，用于调整难易样本的权重， p_t 为真实类别的预测概率， γ 为聚焦参数。

本文将正样本的类别权重 α_t 设为 0.25，负样本相应设为 $1-\alpha_t=0.75$ ，聚焦参数 γ 设为 2.0。训练时采用批量大小为 32，学习率为 0.000 01，使用 Adam 优化器自适应更新模型参数。

测试集包含 120 个有效踢腿样本和 360 个无效踢腿样本，其中无效样本由 4 种非踢腿动作组成，每种 90 个样本。脚踢动作识别混淆矩阵如图 14 所示。图 14 展示了模型在测试集上的分类结果，有效踢腿动作的识别准确率达到 99.17%，仅有 1.11% 的无效踢腿样本被误判为有效；无效踢腿动作的识别准确率为 98.89%，仅有 0.83% 的有效踢腿动作被误判为无效。结果表明，双流 CNN-MHSA-STCN 模型在踢腿动作的区分上表现优异，误判率极低。

为了评估所提模型在动作识别任务中的性能，本文在相同数据集上对多个模型进行了对比实验。模型包括 CNN、STCN、CNN-MHSA、CNN-STCN、CNN-LSTM 和 CNN-TCN，其中前 4 个也用于消融实验研究。本文选取准确率、精确率和 F1 分数作为评估指标。其中，准确率用

于衡量模型整体分类的正确性，精确率反映模型对正类预测的可靠性，而 F1 分数则综合考虑了精确率与召回率的平衡性，用于更全面地评价模型性能。消融实验及不同模型的对比结果见表 2。

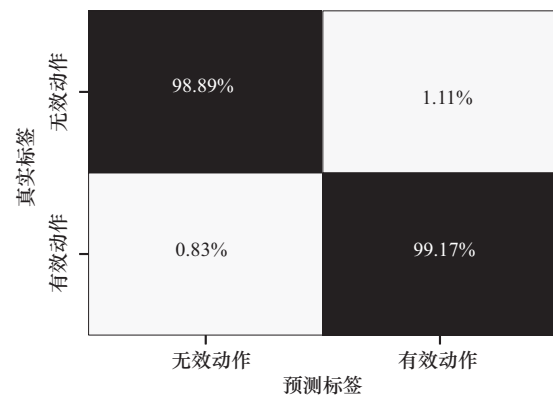


图14 脚踢动作识别混淆矩阵

表2 消融实验及不同模型的对比结果

模型	准确率	精确率	F1 分数
CNN	95.79%	87.00%	0.906 2
STCN	95.33%	95.00%	0.883 7
CNN-MHSA	98.36%	97.75%	0.961 3
CNN-STCN	98.13%	96.67%	0.956 0
CNN-LSTM	96.26%	94.19%	0.910 1
CNN-TCN	94.16%	80.73%	0.875 6
本文模型	98.96%	96.75%	0.979 4

实验结果表明，仅使用 CNN 或 STCN 进行动作识别时，准确率分别为 95.79% 和 95.33%，准

表3 其他参考文献中脚踢动作识别准确率的对比

模型	数据集	雷达型号	准确率
CNN ^[2]	距离、速度和角度	FMCW 雷达 (BGT60ATR24C)	97.41%
CNN+Bi-RNN ^[9]	时间-距离-多普勒	UWB 雷达 (X4M02)	99.86%
AlexNet ^[23]	高压压缩雷达特征图	多普勒雷达	98.64%
本文模型	RDM、RAM	FMCW 雷达 (ADT3102)	98.96%

准确率稍低,这是因为单一模块在建模时只能关注空间或时间维度的特征,缺乏对时空信息的联合建模能力。因此,本文在CNN的基础上引入了MHSA模块,以增强时间建模能力,并进一步探索了结合CNN与STCN的混合结构。由此得到的CNN-MHSA和CNN-STCN模型准确率分别提升至98.36%和98.13%,这表明联合建模时空信息能显著提升识别性能。本文提出的模型在基于CNN进行空间特征提取的基础上,进一步融合了MHSA和STCN模块,不仅实现了更强的时间建模能力,还使模型能够聚焦关键时间帧,识别准确率达到98.96%,优于其他结构。

此外,本文还比较了2种常见的时空建模方法:CNN-LSTM和CNN-TCN。CNN-LSTM模型基于CNN与LSTM的结合,针对本文的数据特性进行了调整,采用双分支CNN提取空间特征,随后输入LSTM捕捉时间特征,识别准确率达到96.26%。CNN-TCN模型则采用TCN提取时间信息,准确率为94.16%。这两种方法的结果均略低于本文提出的模型。对比结果充分验证了所提模型在动作识别任务中的有效性。其他参考文献中脚踢动作识别准确率的对比见表3。文献[2]利用距离、速度和角度序列作为CNN输入特征,准确率为97.41%。文献[9]基于UWB信号,采用时间-距离-多普勒特征,并结合CNN与双向循环神经网络(bidirectional recurrent neural network, Bi-RNN),准确率达到99.86%。文献[23]使用一种新型高压压缩雷达特征图和AlexNet模型进行足部

姿势识别,准确率为98.64%。相比之下,本文方法以每帧的RDM和RAM作为输入特征,融合空间与时间信息,准确率达98.96%。本文方法有效结合了RDM和RAM的特征,增强了时空建模能力,提高了识别性能和鲁棒性,展现出较强的动作特征捕捉能力,具有一定的创新性和参考价值。

4 结束语

本文提出了一种基于毫米波雷达和深度学习的脚踢动作识别方法。首先,针对雷达信号中的干扰问题,设计了一套干扰抑制算法,有效抑制了静态和动态干扰的影响;随后,从干扰抑制后的雷达信号中提取多维特征,具体为RDM和RAM,并将其输入双流CNN-MHSA-STCN模型进行脚踢动作识别。为了验证所提方法的有效性,本文基于雷达传感器进行了实验。实验结果表明,双流CNN-MHSA-STCN模型在动作识别任务中表现优异,动作识别准确率超过98%。可见,基于毫米波雷达的脚踢动作识别在智能交互应用中具有广阔的应用前景,尤其适用于智能车尾门控制等免手操作场景。

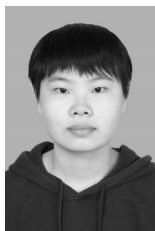
未来研究计划扩展识别动作类型,涵盖挥手、跌倒等,以评估模型在多样化动作条件下的稳定性。此外,还将致力于将干扰抑制算法和双流CNN-MHSA-STCN模型部署到边缘设备上,以提升实时响应能力和实际应用价值。



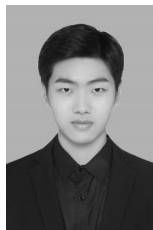
参考文献:

- [1] Chen Z M, Tu H W, Wu H Y. User-defined foot gestures for eyes-free interaction in smart shower rooms[J]. *International Journal of Human-Computer Interaction*, 2023, 39(20): 4139-4161.
- [2] Park C, Baek H I, Chae Y, et al. Deep-learning-based kick motion recognition in millimeter waveband radar system[J]. *IEEE Sensors Journal*, 2024, 24(19): 31395-31407.
- [3] Frank S, Kuijper A. Robust driver foot tracking and foot gesture recognition using capacitive proximity sensing[J]. *Journal of Ambient Intelligence and Smart Environments*, 2019, 11(3): 221-235.
- [4] Saez B, Mendez J, Molina M, et al. Gesture recognition with ultrasounds and edge computing[J]. *IEEE Access*, 2021, 9: 38999-39008.
- [5] Zhang Q, Lin Y Q, Lin Y B, et al. Hand pose estimation with mems-ultrasonic sensors[C]//*Proceedings of the SIGGRAPH Asia 2023 Conference Papers*. New York: ACM Press, 2023: 1-11.
- [6] Noble F, Xu M Q, Alam F. Static hand gesture recognition using capacitive sensing and machine learning[J]. *Sensors*, 2023, 23(7): 3419.
- [7] Liu Y, Guo L C, Makarov V A, et al. Agile gesture recognition for capacitive sensing devices: adapting on-the-job[C]//*Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN)*. Piscataway: IEEE Press, 2023: 1-8.
- [8] Li S B, Liu Y. Human motion recognition based on Nano-CMOS Image sensor[J]. *Mathematical Biosciences and Engineering*, 2023, 20(6): 10135-10152.
- [9] Wang Y H, Yan S, Fu Y L, et al. UWB radar signal kick detection for tailgate unlocking based on spatio-temporal network[C]//*Proceedings of the 2024 IEEE International Conference on Mobility, Operations, Services and Technologies (MOST)*. Piscataway: IEEE Press, 2024: 135-142.
- [10] Shankar Y, Santra A. Valid kick recognition in smart trunks based on hidden Markov model using Doppler radar[C]//*Proceedings of the 2019 International Radar Conference (RADAR)*. Piscataway: IEEE Press, 2020: 1-5.
- [11] Zhang H F, Liu K, Zhang Y H, et al. TRANS-CNN-based gesture recognition for mmWave radar[J]. *Sensors*, 2024, 24(6): 1800.
- [12] Lei W T, Jiang X Y, Xu L, et al. Continuous gesture recognition based on time sequence fusion using MIMO radar sensor and deep learning[J]. *Electronics*, 2020, 9(5): 869.
- [13] Jin B, Ma X, Zhang Z K, et al. Interference-robust millimeter-wave radar-based dynamic hand gesture recognition using 2-D CNN-transformer networks[J]. *IEEE Internet of Things Journal*, 2024, 11(2): 2741-2752.
- [14] Dekker B, Jacobs S, Kossen A S, et al. Gesture recognition with a low power FMCW radar and a deep convolutional neural network[C]//*Proceedings of the 2017 European Radar Conference (EURAD)*. Piscataway: IEEE Press, 2018: 163-166.
- [15] 赵学荣, 王旋, 刘彤, 等. 面向智慧博物馆的基于毫米波雷达稳健的手语识别[J]. *电信科学*, 2023, 39(8): 109-117.
Zhao X R, Wang X, Liu T, et al. mmWave radar based robust sign language recognition for the smart museum[J]. *Telecommunications Science*, 2023, 39(8): 109-117.
- [16] 罗金燕, 常俊, 吴彭, 等. 基于残差网络的FMCW雷达人体行为识别[J]. *计算机科学*, 2023, 50(S2): 174-179.
Luo J Y, Chang J, Wu P, et al. Human behavior recognition of FMCW radar based on residual network[J]. *Computer Science*, 2023, 50(S2): 174-179.
- [17] Diao P S, Alves T, Poussot B, et al. A review of radar detection fundamentals[J]. *IEEE Aerospace and Electronic Systems Magazine*, 2024, 39(9): 4-24.
- [18] Saha P K, Logofatu D. Efficient approaches for density-based spatial clustering of applications with noise[C]//*Artificial Intelligence Applications and Innovations*. Cham: Springer, 2021: 184-195.
- [19] Bijalwan V, Khan A M, Baek H, et al. Interpretable human activity recognition with temporal convolutional networks and model-agnostic explanations[J]. *IEEE Sensors Journal*, 2024, 24(17): 27607-27617.
- [20] Nan M H, Trăscău M, Florea A M, et al. Comparison between recurrent networks and temporal convolutional networks approaches for skeleton-based action recognition[J]. *Sensors*, 2021, 21(6): 2051.
- [21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017(30): 1-11.
- [22] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2017: 2999-3007.
- [23] Song S, Kim B, Kim S, et al. Foot gesture recognition using high-compression radar signature image and deep learning[J]. *Sensors*, 2021, 21(11): 3937.

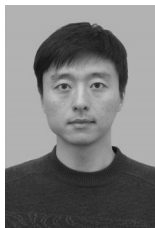
[作者简介]



周晓昀 (2001-), 女, 中国计量大学机电工程学院硕士生, 主要研究方向为毫米波雷达信号处理和深度学习。



陈邵靖 (2003-), 男, 中国计量大学机电工程学院硕士生, 主要研究方向为毫米波雷达信号处理。



张远辉 (1982-), 男, 博士, 中国计量大学机电工程学院副教授, 主要研究方向为毫米波雷达算法、图像处理技术。



郑超群 (2001-), 男, 中国计量大学机电工程学院硕士生, 主要研究方向为毫米波雷达信号处理。