

DOI:10.3969/j.issn.1672-2337.2024.02.003

# 基于三通道CNN-GSAM-LSTFEM网络的 雷达人体切向动作识别

屈乐乐, 祝诗卉

(沈阳航空航天大学电子信息工程学院, 辽宁沈阳 110136)

**摘要:** 为了提高干涉雷达对人体切向动作的识别性能,本文提出一种基于三通道CNN-GSAM-LSTFEM网络的人体切向动作识别方法。首先利用一发二收的调频连续波(FMCW)雷达搭建干涉雷达平台采集人体切向动作回波数据,之后对每个接收通道的回波数据进行预处理,得到每个接收通道的多普勒时频图(DTFM)和双通道的干涉时频图(ITFM),然后将这3种时频图分别送入到3个并行的CNN-GSAM-LSTFEM网络进行训练,利用全局空间注意力模块(GSAM)和长短时特征提取模块(LSTFEM)增强卷积神经网络(CNN)的特征提取能力,最后将三通道提取的特征进行融合实现人体切向动作识别。实验结果表明,所提方法可有效提高人体切向动作的识别准确率,平均准确率高达98.77%。

**关键词:** 人体动作识别; 干涉雷达; 注意力机制; 卷积神经网络; 特征融合

中图分类号: TN958.95

文献标志码: A

文章编号: 1672-2337(2024)02-0135-10

**引用格式:** 屈乐乐, 祝诗卉. 基于三通道CNN-GSAM-LSTFEM网络的雷达人体切向动作识别[J]. 雷达科学与技术, 2024, 22(2): 135-144.

QU Lele, ZHU Shihui. Radar Human Tangential Activity Recognition Based on Three-Channel CNN-GSAM-LSTFEM Network[J]. Radar Science and Technology, 2024, 22(2): 135-144.

## Radar Human Tangential Activity Recognition Based on Three-Channel CNN-GSAM-LSTFEM Network

QU Lele, ZHU Shihui

(College of Electronic Information Engineering, Shenyang Aerospace University, Shenyang 110136, China)

**Abstract:** In order to improve the performance of interferometric radar for human tangential activity recognition, a human tangential activity recognition method based on three-channel CNN-GSAM-LSTFEM network is proposed in this paper. Firstly, an interferometric radar platform is constructed using a frequency modulated continuous wave (FMCW) radar with one transmitter and two receivers to collect the human tangential motion echo data. Subsequently, the echo data are preprocessed to obtain the Doppler time-frequency map (DTFM) for each receiving channel and the two-channel interferometric time-frequency map (ITFM). Then, the three obtained time-frequency maps are separately fed into three parallel CNN-GSAM-LSTFEM networks for training. The global spatial attention module (GSAM) and long-short time feature extraction module (LSTFEM) are used to enhance the feature extraction ability of convolutional neural network (CNN). Finally, the features extracted from the three channels are fused to achieve human tangential activity recognition. The experimental results show that the proposed method can effectively improve the recognition accuracy of human tangential activities and the average accuracy is as high as 98.77%.

**Key words:** human activity recognition; interferometric radar; attention mechanism; CNN; feature fusion

## 0 引言

随着科技的快速发展,人体动作识别和行为感知技术在众多领域中都扮演着十分重要的角色,在智能家居、睡眠监测、安防监控、医疗监护和

人机交互等领域中都起着不可替代的作用<sup>[1]</sup>。人体动作识别方式主要有接触式和非接触式两类,接触式传感器需要随身佩戴且识别准确率低,非接触式传感器包括摄像头、WiFi和雷达等<sup>[2]</sup>。雷达通过发射电磁波来探测和感知人体目标,与其

收稿日期: 2023-10-26; 修回日期: 2023-11-16

基金项目: 国家自然科学基金(No.61671310); 航空科学基金(No.2019ZC054004); 辽宁省兴辽英才计划项目基金(No.XLYC1907134); 辽宁省百千万人才工程项目基金(No.2018B21)

他传感器相比,可以有效防止隐私泄露<sup>[3]</sup>,同时雷达可以在不同的天气和光线条件下工作,具有较强的环境鲁棒性,因此雷达在人体动作识别领域中具有很大的应用前景<sup>[4]</sup>。

在基于雷达的人体动作识别研究中,绝大部分工作都集中在基于微多普勒特征的各种分类算法上。文献[5]中研究了基于微多普勒特征对不同人体动作进行分类的可行性。文献[6]中利用多普勒频移特征提取算法对具有径向位移的人体动作进行特征提取,将提取到的特征谱图送入卷积神经网络(Convolutional Neural Network, CNN)中进行训练和测试,平均准确率高达 99%。基于微多普勒特征的人体动作分类主要缺点是过于依赖人体目标的径向速度<sup>[7]</sup>,当运动方向与雷达视线之间的角度大于 60°时,人体运动产生的微多普勒特征会减少,从而降低分类算法的性能<sup>[8]</sup>。当人相对于雷达视线切向运动时,由人体目标径向速度引起的多普勒频移则降为零<sup>[9]</sup>。干涉雷达配备一个发射天线和两个宽间距接收天线,可以同时提供径向和切向速度信息<sup>[10]</sup>,进而可以测得物体运动的切向速度,从而获得物体的切向运动特征。

近年来,Wang 等<sup>[11]</sup>利用干涉雷达获得手势的多普勒时频图(Doppler Time-Frequency Map, DTFM)和干涉时频图(Interferometric Time-Frequency Map, ITFM),输入到 Resnet-50 中提取手势的径向和切向运动特征,可以同时区分出水平对称和垂直对称的手势动作。而且在文献[12]中利用运动捕捉数据模拟人体不同部位的运动并生成雷达回波信号,将获得的 DTFM 和 ITFM 输入到深度卷积神经网络(Deep Convolution Neural Network, DCNN)中,对人体切向动作进行分类,但雷达实测数据集中只有行走模式,并且提出的网络性能有限,识别准确率不高。

为了提高人体切向动作的识别准确率,本文对 FMCW 干涉雷达回波数据进行预处理得到 DTFM 和 ITFM,分别输入到 3 个并行的 CNN-GSAM-LSTFEM 网络中,在每个通道的 CNN 后都添加了全局空间注意力模块(Global Spatial Attention Module, GSAM)以及长短时特征提取模块(Long-Short Time Feature Extraction Module, LSTFEM)来提高特

征提取能力。GSAM 可提高对原始图像特征捕捉能力,而 LSTFEM 可加强对人体动作的时间特征捕捉能力。这两种模块都具有参数少,即插即用的特点,从而进一步提高整体的分类性能。

## 1 回波信号预处理

### 1.1 干涉原理

当人体在雷达视线范围内运动时,会产生频率偏移,这种频率偏移与人体相对于雷达收发天线的径向速度  $v_r$  成正比,即多普勒频移  $f_D$ <sup>[9]</sup>:

$$f_D = \frac{2v_r}{\lambda} = \frac{2v}{\lambda} \cos \varphi \quad (1)$$

式中  $\varphi$  为人体与雷达视线之间的夹角, $\lambda$  为雷达发射电磁波的波长, $v$  为人体运动速度。多普勒频移正比于物体的径向速度,当径向速度较低时,微多普勒频移会显著降低,降低分类的准确性。

干涉雷达具有两个宽间距的接收天线,天线之间的基线为  $D$ ,干涉处理就是将两路接收信号进行相干处理。当人体目标沿着雷达视线的切线方向运动时,两个接收天线接收的人体回波信号的相位不同,存在一个相位差,对两个接收天线的回波信号进行相干处理会产生干涉频移  $f_s$ <sup>[9]</sup>:

$$f_s = \frac{v_t D}{\lambda R} = \frac{v D}{\lambda R} \sin \varphi \quad (2)$$

式中  $v_t$  为物体的切向速度, $R$  为目标与发射天线之间的距离。由此可知干涉频移正比于物体的切向速度,只要进行干涉处理,就可以得到人体的切向运动特征,从而区分出切向运动的人体动作。

### 1.2 回波信号处理

#### 1.2.1 回波信号获取

本文采用一发二收的 FMCW 雷达搭建成干涉雷达进行人体切向动作数据采集,回波信号处理流程图如图 1 所示。FMCW 雷达发射信号的载波频率为  $f_0$ ,带宽为  $B$ 、调频周期为  $T$ 。将 FMCW 雷达第  $n$  次扫频的发射信号与第  $i$  ( $i = 1, 2$ ) 个接收天线接收到的回波信号进行混频并采样,得到离散中频信号<sup>[11]</sup>:

$$s_{Bi}[l] \approx A \exp \left\{ -j4\pi \frac{(f_0 + \mu m T_s)(R_i - n v_r T)}{c} \right\} \quad (3)$$

式中:  $m = 1, \dots, M, n = 1, \dots, N, M$  为每个调频周期

上的采样点个数,  $N$  为总的调频周期个数;  $l = (n - 1)M + m$ ;  $A$  为中频信号幅度值;  $T_s$  为快时间维采样周期;  $T$  为调频周期;  $\mu = B/T$  为啁啾率;  $R_i$  为人体目标与第  $i$  个接收天线之间的距离;  $c$  为雷达发射电磁波在空气中的传播速度。

将中频信号  $s_{Bi}[l]$  的离散样本按列堆叠排列成一个  $M \times N$  的矩阵  $S_{Bi}$ :

$$S_{Bi}[m, n] = s_{Bi}[m + (n - 1)M] \quad (4)$$

矩阵  $S_{Bi}$  是雷达原始回波数据的二维矩阵形式, 横轴为慢时间维, 纵轴为快时间维。

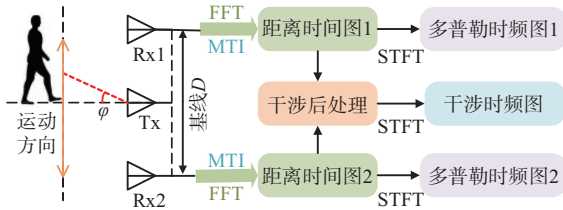


图1 回波信号处理流程图

### 1.2.2 单通道信号处理

由于中频信号的频率与目标的距离为线性关系, 对矩阵  $S_{Bi}$  的每一列即快时间维进行快速傅里叶变换 (Fast Fourier Transform, FFT) 得到距离时间矩阵:

$$R_{Bi}[k, n] = \frac{1}{M} \sum_{m=0}^{M-1} S_{Bi}(m, n) \exp(-j \frac{2\pi}{N} km) \quad (5)$$

式中  $k = 0, \dots, M - 1$  表示人体目标与雷达天线间的距离信息。为了减少背景杂波及直达波的影响, 本文使用巴特沃斯高通滤波器对  $R_{Bi}(k, n)$  进行动目标显示 (Moving Target Indication, MTI) 滤波处理。滤波之后得到距离时间图  $D_{Bi}(k, n)$ , 横轴为时间信息, 纵轴为距离信息。由于切向运动的人体相对于观测雷达之间的距离分布是相同的, 对  $D_{Bi}(k, n)$  选择合适的距离门限, 然后将距离门限内的距离单元进行相干叠加得到  $D_{Bi}(n)$ 。

当人体目标作非匀速运动时, 对  $D_{Bi}(n)$  使用短时傅里叶变换 (Short-Time Fourier Transform, STFT) 算法提取时变的速度信息, 可得到含有径向速度随时间变化信息的 DTFM:

$$T_{Bi}(p, q) = \sum_{n=0}^{N-1} D_{Bi}(n) \omega(n - p) \exp(-j2\pi \frac{nq}{N}) \quad (6)$$

式中  $\omega(n)$  为时间窗函数,  $p$  为窗函数移动的索引值, 对应着时间,  $q$  为多普勒频率。

### 1.2.3 双通道后干涉处理

信号预处理阶段有两种干涉处理方式: 一种是前干涉处理, 将两个接收信号在时域中相乘; 另一种是后干涉处理, 即将两个接收信号在频域中相乘。由于前干涉处理是时域中的非线性乘法, 会增加噪声和不必要的交叉项, 而后干涉处理能够降低频谱图的噪声水平并抑制干扰交叉项<sup>[11]</sup>, 所以本文采用后干涉处理方法。

后干涉处理是对两个接收通道得到的距离时间矩阵  $D_{B1}(k, n)$  和  $D_{B2}(k, n)$  共轭相乘, 即

$$D_1[k, n] = D_{B1}(k, n) D_{B2}^*(k, n) \quad (7)$$

将干涉输出的  $D_1(k, n)$  目标所在距离单元也进行相干叠加得到  $D_1(n)$ , 再使用 STFT 算法从中提取目标切向运动的 ITFM:

$$F_1(p, q) = \sum_{n=0}^{N-1} D_1(n) \omega(n - p) \exp(-j2\pi \frac{nq}{N}) \quad (8)$$

后干涉处理的本质是频率(距离)域进行带通滤波, 滤除其他距离门的噪声, 这样能够降低噪声水平和抑制交叉项。

## 2 模型框架

本文所提出的网络模型由特征提取、特征融合和分类器 3 个模块组成, 总体框图如图 2 所示。特征提取模块由 3 个并行的 CNN、GSAM 和 LSTFEM 组成, 三通道的输入图像分别是两个通道的 DTFM 和 ITFM。特征提取模块提取三通道特征后, 输入特征融合模块进行融合, 再送入分类器进行输出分类。

### 2.1 特征提取模块

#### 2.1.1 CNN

VGG-16 网络模型对图像数据有良好的特征提取能力, 但它的参数量较大, 网络收敛速度较慢。CNN 擅长处理二维图像, 常采用该网络模型学习时频图的特征进而实现人体动作的分类识别<sup>[13]</sup>。本文基于 VGG-16 网络进行改进, 设计了一种仅包含 5 层卷积的 CNN 网络作为主干网络, 能够有效提取数据特征并减少训练时间。网络包括 5 个 2 维卷积层, 每一个卷积层后分别连接 1 个修正线性单元 (Rectified Linear Unit, ReLU) 激活函数和 1 个最大池化层, 5 层 CNN 网络结构如表 1 所示。

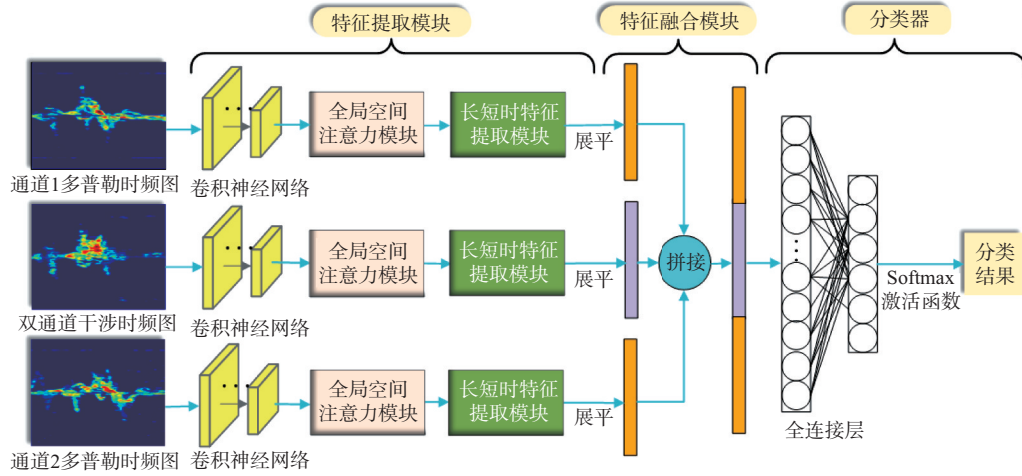


图2 网络模型总体框图

表1 CNN模型结构

层数	层结构	卷积核尺寸	卷积核数量	输出维度
1	卷积层1、ReLU 最大池化层1	3×3	8	8×112×112
2	卷积层2、ReLU 最大池化层2	3×3	16	16×56×56
3	卷积层3、ReLU 最大池化层3	3×3	32	32×28×28
4	卷积层4、ReLU 最大池化层4	3×3	64	64×14×14
5	卷积层5、ReLU 最大池化层5	3×3	128	128×7×7

2.1.2 GSAM

DTFM 和 ITFM 中除了存在人体运动产生的微多普勒和干涉频移之外,通常还存在背景噪声以及未滤除干净的静止杂波。为了使改进后的网络更加关注人体运动的关键特征,本文在不同层的卷积模块后添加了注意力机制模块。注意力机制的目的是按照特征的重要程度对输入特征进行加权,使网络重点关注重要特征而抑制无关特征<sup>[14]</sup>,使预测更加明显。受文献[15]中提出的全局注意力机制的启发,本文将GSAM应用在CNN中,该模块联合全局信息来突出中间层中感兴趣的局部区域,并将注意力细化的特征向量与全局特征向量聚合。GSAM的结构框图如图3所示。全局特征指CNN最后一层输出的特征,中间特征指最后一层之前输出的特征。给定任意中间层  $l$  的特征  $F^l \in \mathbb{R}^{H \times W \times C}$ ,  $H$ 、 $W$  和  $C$  分别是特征图的高度、宽度和通道数量。在提出的CNN架构中,特征映射逐

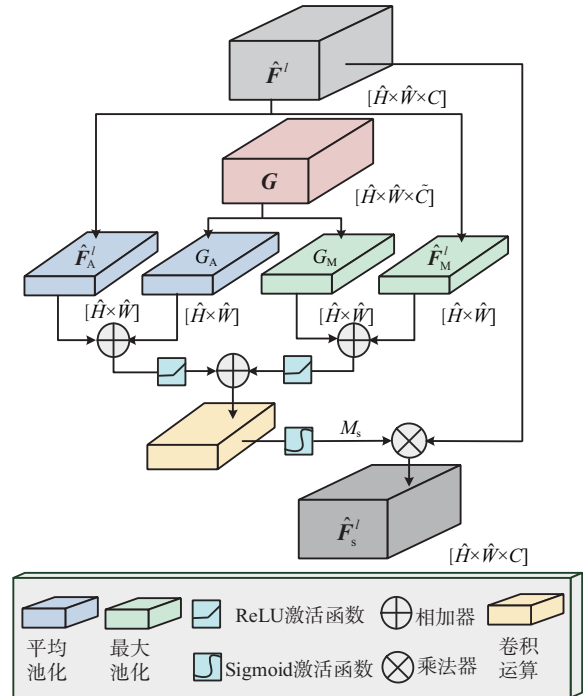


图3 GSAM结构框图

渐下采样以捕获足够大的感受野。因此,得到的每一层的空间分辨率不同,需要将中间层特征尺寸与全局特征尺寸相匹配。本文使用双三次插值方法对中间层特征  $F^l$  下采样,得到下采样后的中间层特征  $\hat{F}^l \in \mathbb{R}^{\hat{H} \times \hat{W} \times C}$ , 其宽和高同全局特征  $G \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$  一致。将下采样得到的  $\hat{F}^l$  和  $G$  沿两者的通道轴分别做平均池化和最大池化,这些操作产生了  $\hat{F}_A^l$  和  $G_A$ , 还有最大池化特征  $\hat{F}_M^l$ 、 $G_M$ , 这4个特征图的尺寸都是  $\hat{H} \times \hat{W} \times 1$ 。

将平均池化特征  $\hat{F}_A^l$  和  $G_A$  相加,并通过非线性激活函数 ReLU 得到激活后的平均池化特征,来关注  $\hat{F}^l$  中相对于  $G$  的感兴趣区域。对  $\hat{F}_M^l$  和  $G_M$  也重复相同的操作,再将激活后的平均池化和最大池化特征相加,并通过卷积层生成全局空间注意力图,最后再经过归一化激活函数 Sigmoid 将全局空间注意力图中的数值固定在 0~1 之间,这样就获得了一个全局空间注意力权重  $M_s \in \mathbb{R}^{\hat{h} \times \hat{w} \times 1}$ :

$$M_s = \sigma_2(f^{1 \times 1}(\sigma_1(\hat{F}_A^l + G_A) + \sigma_1(\hat{F}_M^l + G_M))) \quad (9)$$

式中  $\sigma_1$  表示 ReLU 激活函数,  $f^{1 \times 1}$  表示卷积核尺寸为  $1 \times 1$  的卷积运算,  $\sigma_2$  表示 Sigmoid 激活函数。最后将得到的  $M_s$  的通道维扩充到  $C$  再与  $\hat{F}^l$  逐元素相乘得到最终输出的全局空间注意力细化特征  $\hat{F}_s^l \in \mathbb{R}^{\hat{h} \times \hat{w} \times C}$ :

$$\hat{F}_s^l = M_s \otimes \hat{F}^l \quad (10)$$

GSAM 利用最后一层的全局特征从空间维度学习细化中间层特定于类的特征,克服了普通卷积层有限的感受野,从而使它们能够学习除了空间局部化特征之外的全局特征。

### 2.1.3 LSTFEM

DTFM 和 ITFM 能够反映人体行走时微多普勒及干涉频率随时间变化的过程,不同的时间段对应的特征有所差异。在行人重识别领域中,文献 [16] 提出了水平金字塔匹配方法,将人体图片从上至下切分成不同的水平部分,利用全局平均池化 (Global Average Pooling, GAP) 和全局最大池化 (Global Max Pooling, GMP) 分别提取各个部分的局部细粒度特征,然后将局部特征与全局特征进行融合以增强特征表示。文献 [17] 在此基础上提出了长短时特征提取模块将多普勒时频图沿时间维度 (从左至右) 切分成不同的份数,对其中的一份提取特征,得到短时的“局部”特征,对整个时频图进行特征提取能够得到长时的“全局”特征,最后将两部分特征进行融合以增强特征表示,进行人体步态识别。本文结合文献 [16] 中的水平金字塔匹配方法对文献 [17] 中的 LSTFEM 进行改进,同时使用 GAP 和 GMP 进行特征提取,可以提高特征提取能力,改进后的 LSTFEM 结构框图如图 4 所示。

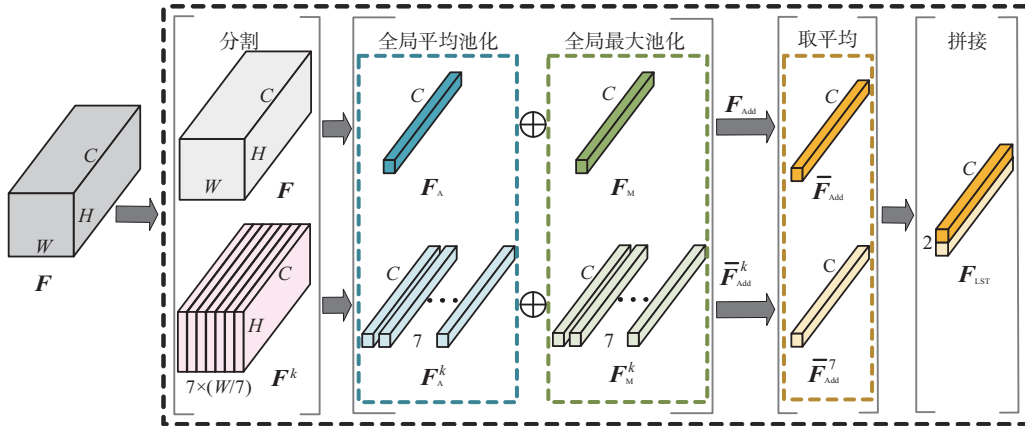


图4 LSTFEM结构框图

首先将LSTFEM输入特征图  $F \in \mathbb{R}^{C \times H \times W}$  沿时间维度 (宽度) 分割成不同的份数,  $H$ 、 $W$  和  $C$  分别是特征图的高度、宽度和通道数量。本文  $F$  的宽度为 7, 因此将  $F$  沿时间维度分割成 7 份, 分割后的特征图表示为  $F^k \in \mathbb{R}^{C \times H \times W/7}$  ( $k = 1, 2, \dots, 7$ )。对划分后的 7 个特征  $F^k$  分别使用 GAP 和 GMP, 得到长度为  $C$  的特征向量  $F_A^k, F_M^k$ , 然后将  $F_A^k$  和  $F_M^k$  进行逐元素相加得到求和后的特征向量  $F_{Add}^k$ 。不同时频图中人体的起始状态可能不同, 为了弱化时频图起

始状态不对齐对特征提取的影响, 将 7 个特征向量  $F_{Add}^k$  取平均值得到  $\bar{F}_{Add}^7$ 。对没有分割的  $F$  也重复相同的操作得到  $\bar{F}_{Add}$ , 最后将“短时”特征向量  $\bar{F}_{Add}^7$  和“长时”特征向量  $\bar{F}_{Add}$  按图 4 所示的拼接方式进行拼接, 得到长短时运动特征  $F_{LST} \in \mathbb{R}^{C \times 2}$ 。

### 2.1.4 CNN-GSAM-LSTFEM

由于DTFM和ITFM的时频特征没有优先级排列,因此特征提取模块中3个并行的网络模型具有相同的结构和参数,一个通道的CNN-GSAM-LST-

FEM 特征提取结构如图 5 所示。由于 CNN 的前两层输出代表的都是低层次特征,没有足够的判别性,不需要添加注意力机制。本文将 CNN 三、四层输出的中间特征图与第五层(最后一层)输出的全局特征图一起通过 GSAM,得到两个注意力细化特征图,最后再将这些注意力细化特征与全局特征按图中所示方式拼接在一起,作为 LSTFEM 的输入。

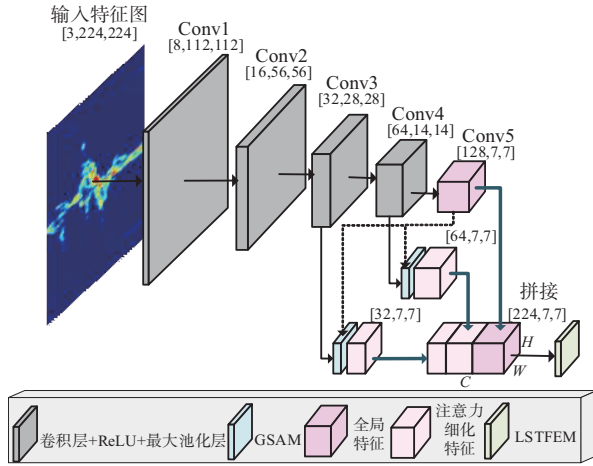


图 5 CNN-GSAM-LSTFEM 特征提取结构框图

## 2.2 特征融合及分类

将特征提取模块提取到的三通道特征展开成一维特征向量后,通过特征融合模块将 3 个特征向量进行拼接融合,拼接后的一维向量作为分类器的输入。分类器包括全连接层和 Softmax 层,由于融合特征中还包含大量的神经元,在分类器中还使用了 Dropout 层来减少隐藏节点的相互作用,可以明显地减少过拟合现象,分类器的最后一层利用 Softmax 激活函数实现了从全局特征到分类输出的非线性映射,分类器的参数结构如表 2 所示。

表 2 分类器参数结构

层数	层结构	内部细节
1	全连接层 1+ReLU	隐藏层单元=2 880
2	丢弃层	丢弃率=0.1
3	全连接层 2+ReLU	隐藏层单元=256
4	全连接层 3+ReLU	隐藏层单元=256
5	全连接层 4+Softmax	隐藏层单元=6

## 3 实验结果与分析

### 3.1 实验环境和平台

本文搭建的干涉雷达平台由 Ancortek SDR-

KIT 2400AD2 型号的 K 波段 FMCW 雷达、两个接收天线、一个发射天线和主机组成。3 个天线沿着直线水平排列,发射天线位于两个接收天线的中间,两个接收天线之间的基线长 1 m。FMCW 雷达的参数设置如表 3 所示。

表 3 FMCW 雷达系统参数

参数	参数值
发射天线数量	1 个
接收天线数量	2 个
载频	24 GHz
带宽	2 GHz
调频周期	1 ms
基带采样频率	128 kHz
天线增益	10 dBi
发射功率	18 dBm

### 3.2 数据集构建

实验在室内进行,人体距雷达约 2 m,每个动作的采集时间为 3 s。实验设计了 6 种常见的人体动作:(a)行走,(b)慢跑,(c)迈步,(d)蹲坐,(e)跳远,(f)跌倒,所有动作都是相对于雷达视线方向切向运动的,切向运动的方向包括从左向右和从右向左。实验采集对象为 5 男 3 女,编号为 A 至 H。编号 A 每种动作重复测量 40 次,编号 B~H 每种动作重复测量 30 次,即每个动作共得到 250 个雷达回波数据,从左向右运动和从右向左运动的数量各占其中一半,最终共测得  $6 \times 250 = 1\,500$  组样本。为了提取不同动作在径向和切向上的时变运动特征,对这 6 种人体动作测得的 1 500 组原始回波数据进行第 1 节中的单通道处理和双通道后干涉处理,每组样本都会得到对应的两通道 DTFM 和 ITFM。6 种动作在从左向右和从右向左两种情况下对应的通道一 DTFM 和 ITFM 如图 6 所示,DTFM 和 ITFM 分别展示了不同动作的径向和切向运动特征。

从图 6 可以看到,对于同一动作(从左向右运动和从右向左运动)的 DTFM 是相似的,因为 DTFM 表现人体径向运动特征,这两种运动方式在径向方向具有相同的表现;但同一动作的 ITFM 在两种运动方式上表现不同,ITFM 表现的是人体切向运动特征,从左向右运动的干涉频率为正,而从右

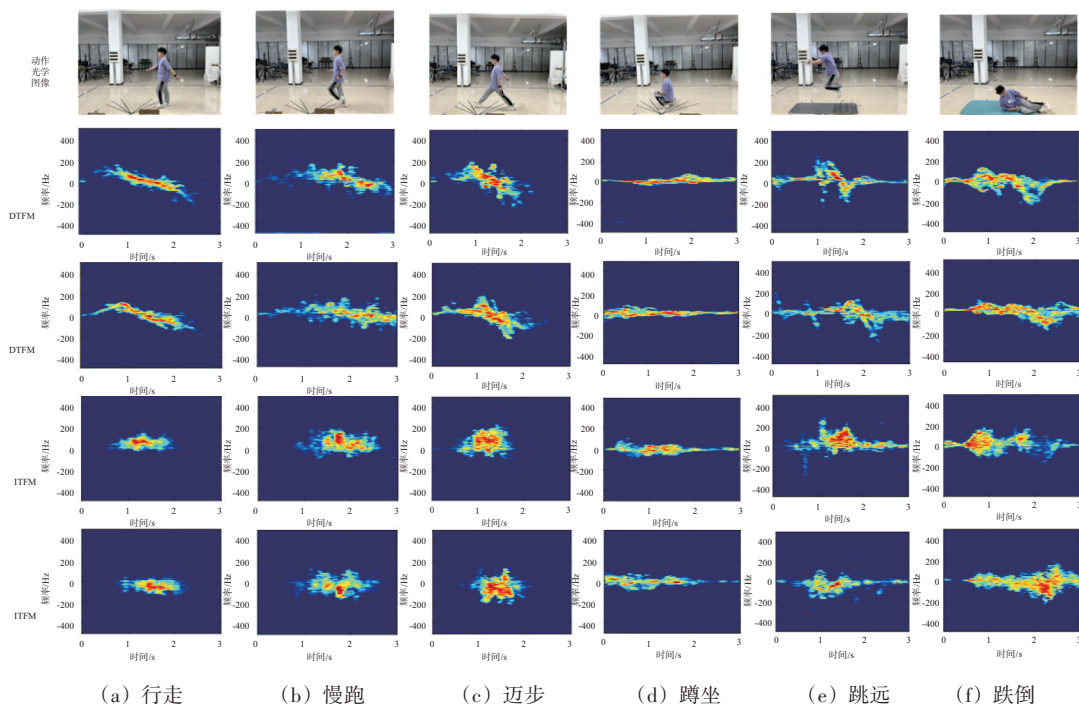


图6 6种人体动作光学图像及对应的通道一DTFM和ITFM(第2、4行为从左向右运动,第3、5行为从右向左运动)

向左运动的干涉频率为负,因为它们沿着切向方向的运动是相反的。

数据预处理后,通过数据增强来扩展数据集,以提高模型的泛化能力。本文应用了一种简单的数据增强方案,将均值为0,方差为随机数的加性高斯白噪声(Additive White Gaussian Noise, AWGN)添加到测量的数据集中,将数据集大小扩充为原始数据集的两倍,最终共得到 $2 \times 1\,500 = 3\,000$ 组样本。

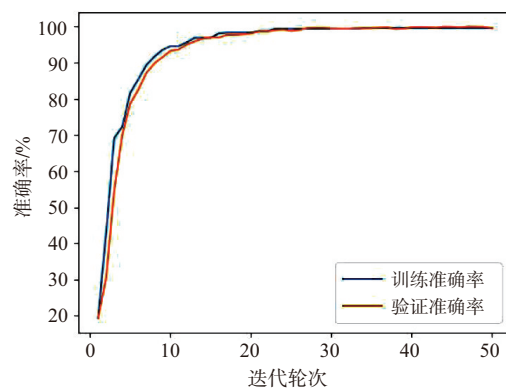
### 3.3 实验结果

#### 3.3.1 网络训练结果

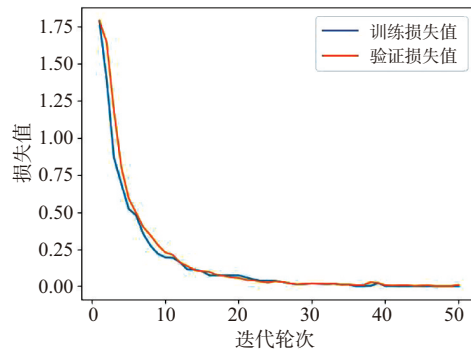
本文使用的深度学习框架为Pytorch, CPU为Intel(R) Xeon(R) Gold 6132,同时使用NVIDIA Quadro P5000的显卡加速训练。为了使训练结果更具有普遍性,本文对测得的数据集进行五折交叉验证实验,将全部的数据随机分为5份,每次取其中的一份作为测试集,其余4份以4:1的比例被随机划分为训练集和验证集,最终结果为5次实验的平均值。

首先将6种人体动作获得的两通道DTFM和ITFM输入到搭建的三通道CNN-GSAM-LSTFEM网

络中。为了适应网络结构的输入要求,将每个通道的RGB彩色时频图像尺寸缩放为 $224 \times 224 \times 3$ ,然后再进行归一化操作,通过将每个图像的像素值除以255,将其范围转换在 $[0,1]$ 之间。网络模型采用Adam梯度下降优化器,并将分类交叉熵设置为损失函数,设置0.0001恒定大小的学习率,训练一共50轮迭代,每次迭代的批样本大小为16。在3000组样本的实验中,对训练集( $320 \times 6 = 1\,920$ )进行模型训练,验证集( $80 \times 6 = 480$ )用于调整模型超参数。三通道CNN-GSAM-LSTFEM的训练验证准确率和损失值曲线如图7所示。



(a) 准确率曲线



(b) 损失值曲线

图7 三通道CNN-GSAM-LSTFEM准确率和损失值曲线

从图7可以看出,本文所提方法的准确率在第27轮之前随着迭代次数持续增加,27轮之后趋于平稳,损失值曲线也在迭代27轮之后趋向于0,说明此时模型参数达到最优。当网络模型训练完成后,调用保存好的模型对测试集数据(100×6=600)进行测试,5次测试的平均识别准确率为98.77%。

### 3.3.2 识别性能分析与对比

为了证明所提方法的有效性,本文在子网络结构相同的前提下,采用3种不同的组合作为网络输入图像,不同输入组合的平均识别准确率如表4所示。

表4 不同输入组合的平均识别准确率

输入组合	平均识别准确率/%
两通道DTFM	92.36
ITFM	94.02
两通道DTFM+ITFM	98.77

从表4可以看出,基于两通道DTFM+ITFM的输入组合平均识别准确率最高,输入为两通道DTFM的平均识别准确率最低。这是因为当人体切向运动时,ITFM能够提供显著的人体切向运动特征,DTFM仅提供一部分径向特征,分类特征不足。当输入为两通道DTFM+ITFM时,两者提供的特征互为补充,能够达到最好的分类效果。

为了证明所提网络中每个模块的有效性,本文进行了消融实验,其中Base代表三通道CNN主干特征提取网络,结果如表5所示。

表5 消融实验平均识别准确率

网络模型	平均识别准确率/%
Base	94.66
Base+GSAM	97.15
Base+LSTFEM	97.32
Base+GSAM+LSTFEM	98.77

相较于Base,Base+GSAM的识别准确率提高了2.49%,证明了GSAM能够让网络重点关注重要特征而抑制无关特征,提高分类性能。与Base相比,Base+LSTFEM的识别准确率提高了2.66%,证明了LSTFEM将“长时”特征和“短时”特征进行融合,增强了特征表示。最后相较于Base+LSTFEM/GSAM,Base+GSAM+LSTFEM的识别准确率至少提高了1.45%。

为了进一步证明所提网络在人体切向动作识别中的优越性,在数据集相同的条件下,本文将该网络与Resnet-50<sup>[11]</sup>、DCNN<sup>[12]</sup>进行比较,结果如表6所示。

表6 不同网络模型平均识别准确率

网络模型	平均识别准确率/%
Resnet-50 <sup>[11]</sup>	86.27
DCNN <sup>[12]</sup>	94.98
三通道CNN-GSAM-LSTFEM	98.77

由表6看出,三通道CNN-GSAM-LSTFEM的平均识别准确率比其他模型至少提高了3.79%,上述3种网络的训练准确率曲线如图8所示。

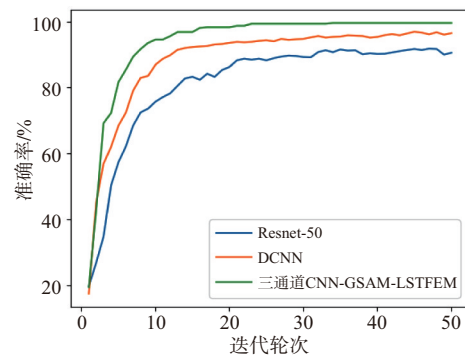


图8 3种网络训练准确率曲线

从图8可以看出,三通道CNN-GSAM-LSTFEM的效果要明显优于DCNN和Resnet-50的效果。

此外,在相同的设备和参数设置下,计算了这3种不同网络模型的数量、模型复杂度及测试平均耗时,如表7所示。

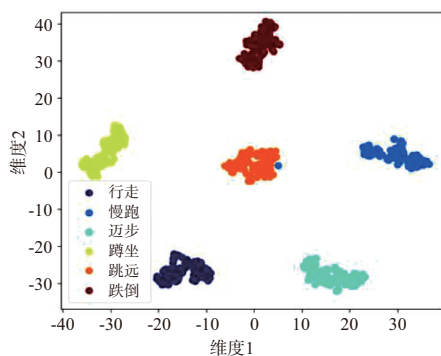
表7 不同网络对比结果

网络模型	参数量 (×10 <sup>6</sup> )	FLOPs (×10 <sup>9</sup> )	测试平均耗时/s
Resnet-50 <sup>[11]</sup>	21.31	62.63	0.005 1
DCNN <sup>[12]</sup>	604.79	14.71	0.000 6
三通道 CNN-GSAM-LSTFEM	0.11	3.3	0.007 7



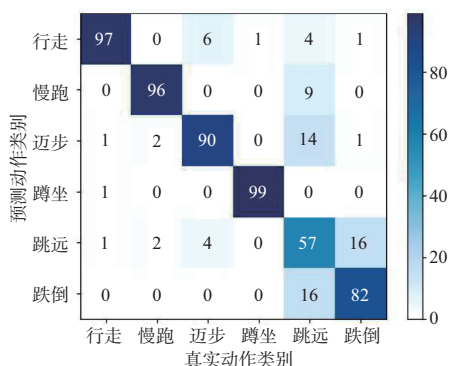
从表7可以看出,本文所提出的网络模型具有最少的参数量和模型复杂度,参数量少的主要原因是CNN中卷积层的卷积核数量要明显少于其他两种模型的卷积核数量;3种网络模型对一个动作的测试平均耗时具有相同的数量级,DCNN具有最短的测试时间主要是因为网络层数较少且输入图像转变为单通道灰度图。综上考虑,本文所提出的Base+GSAM+LSTFEM网络模型更适用于人体切向动作分类。

3种对比网络基于t分布随机邻域嵌入(t-distributed Stochastic Neighbor Embedding, t-SNE)的特征空间可视化散点图和测试混淆矩阵如图9和图10所示。将每个测试数据的输出特征向量转换为二维坐标,即可得到每个数据对应的坐标点,每种类别的人体动作对应点的颜色均不同。从图9可以明显看出,本文提出网络的散点图中每种颜色之间几乎没有重叠,且不同颜色间距离较远,特征可分离度最高。DCNN的散点图中每种颜色之间仅有少量重叠,特征可分离度仅次于本文网络。Resnet-50的散点图中跳远与行走、跌倒、迈步和跌倒均有重叠,这是由于这5种动作的运动幅度都较

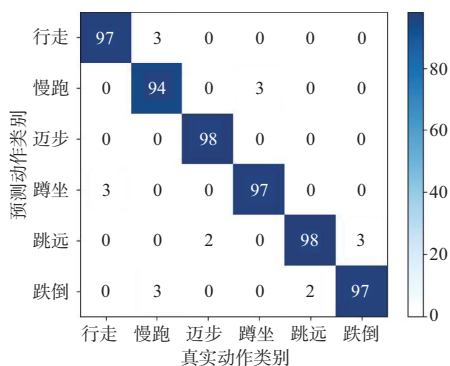


(c) 三通道CNN-GSAM-LSTFEM

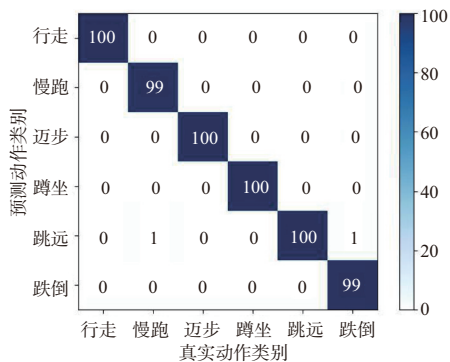
图9 3种对比网络t-SNE可视化散点图



(a) Resnet-50

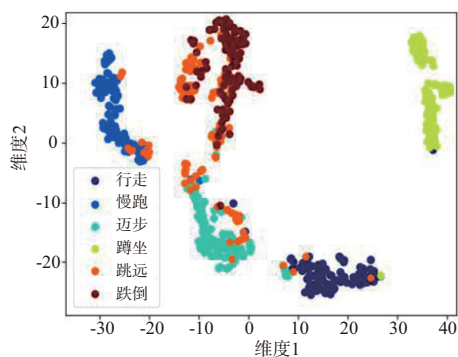


(b) DCNN

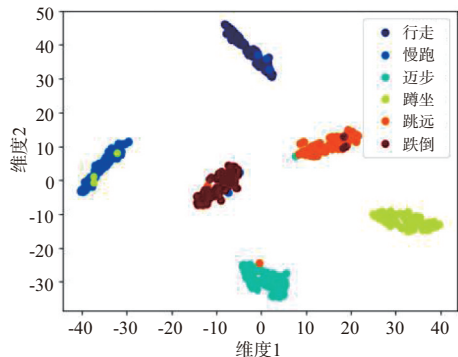


(c) 三通道CNN-GSAM-LSTFEM

图10 3种对比网络测试混淆矩阵



(a) Resnet-50



(b) DCNN

大,因此特征可分离度相对不高。并且蹲坐相较于这 5 种动作的距离较远,这是由于蹲坐相较于其他 5 种动作的运动幅度较小。

以上对比表明,三通道 CNN-GSAM-LSTFEM 更适合从切向动作中提取和优化特征,在人体切向动作识别中取得了良好的效果。

## 4 结束语

本文提出了一种基于三通道 CNN-GSAM-LSTFEM 网络的 FMCW 干涉雷达人体切向动作识别方法。该方法首先采集了 1 500 组包含 6 种人体切向动作的雷达数据集,对每组样本的回波数据进行单通道预处理获得两通道 DTFM,再进行双通道后干涉处理得到 ITFM。为了提高模型的泛化能力,加入 AWGN 将预处理后的数据集扩充为原来的 2 倍。然后,将 3 种时频图并行输入到三通道 CNN-GSAM-LSTFEM 网络,提取三通道数据的运动特征,最后将三通道特征进行融合输出最终的分类结果。实验结果表明,该方法对 6 种人体切向动作的识别准确率高达 98.77%,并且相较于其他网络减少了参数量和计算复杂度。在下一步工作中,将研究原地动作的人体切向动作识别。

### 参考文献:

- [1] LI Xinyu, HE Yuan, FIORANELLI F, et al. Semisupervised Human Activity Recognition with Radar Micro-Doppler Signatures [J]. IEEE Trans on Geoscience and Remote Sensing, 2022, 60(1):1-12.
- [2] LI Wenxuan, ZHANG Dongheng, LI Yadong, et al. Real-Time Fall Detection Using Mmwave Radar [C]//ICASSP, Singapore: IEEE, 2022:16-20.
- [3] CAO Zhongping, LI Zhenchang, GUO Xuemei, et al. Towards Cross-Environment Human Activity Recognition Based on Radar Without Source Data [J]. IEEE Trans on Vehicular Technology, 2021, 70(11):11843-11854.
- [4] 何密,平钦文,戴然.深度学习融合超宽带雷达图谱的跌倒检测研究[J].雷达学报,2023,12(2):343-355.
- [5] KIM Y, LING Hao. Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine [J]. IEEE Trans on Geoscience and Remote Sensing, 2009, 47(5):1328-1337.
- [6] QI Rui, LI Xiuping, ZHANG Yi, et al. Multi-Classification Algorithm for Human Motion Recognition Based on IR-UWB Radar [J]. IEEE Sensors Journal, 2020, 20(21): 12848-12858.
- [7] LIANG Huaiyuan, WANG Xiangrong, GRECO M S, et al. Enhanced Hand Gesture Recognition Using Continuous Wave Interferometric Radar [C]//2020 IEEE International Radar Conference, Washington, USA: IEEE, 2020:226-231.
- [8] NANZER J A, CHEN V C. Microwave Interferometric and Doppler Radar Measurements of a UAV [C]//2017 IEEE Radar Conference, Washington, USA: IEEE, 2017:1628-1633.
- [9] NANZER J A, ZILEVU K S. Dual Interferometric-Doppler Measurements of the Radial and Angular Velocity of Humans [J]. IEEE Trans on Antennas and Propagation, 2014, 62(3):1513-1517.
- [10] WANG Xiangrong, WANG Pengcheng, CHEN V C. Simultaneous Measurement of Radial and Transversal Velocities Using Interferometric Radar [J]. IEEE Trans on Aerospace and Electronic Systems, 2019, 56(4):3080-3098.
- [11] WANG Xiangrong, LI Weiliang, CHEN V C. Hand Gesture Recognition Using Radial and Transversal Dual Micro-Motion Features [J]. IEEE Trans on Aerospace and Electronic Systems, 2022, 58(6):5963-5973.
- [12] HASSAN S, WANG Xiangrong, ISHTIAQ S, et al. Human Activity Classification Based on Dual Micro-Motion Signatures Using Interferometric Radar [J]. Remote Sensing, 2023, 15(7):1752.
- [13] ABDU F J, ZHANG Yixiong, DENG Zhenmiao. Activity Classification Based on Feature Fusion of FMCW Radar Human Motion Micro-Doppler Signatures [J]. IEEE Sensors Journal, 2022, 22(9):8648-8662.
- [14] HE Jiale, REN Zhaoting, ZHANG Wenqiang, et al. Fall Detection Based on Parallel 2DCNN-CBAM with Radar Multidomain Representations [J]. IEEE Sensors Journal, 2023, 23(6):6085-6098.
- [15] VISHWAKARMA S, LI Wenda, TANG Chong, et al. Attention-Enhanced Alexnet for Improved Radar Micro-Doppler Signature Classification [J]. IET Radar, Sonar & Navigation, 2023, 17(4):652-664.
- [16] FU Yang, WEI Yunchao, ZHOU Yuqian, et al. Horizontal Pyramid Matching for Person Reidentification [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1):8295-8302. (下转第 154 页)