

基于MIMO雷达成像图序列的切向人体姿态识别方法

丁传威^① 刘芷麟^① 张力^② 赵恒^① 周庆^① 洪弘^{*①} 朱晓华^{*①}

^①(南京理工大学 南京 210000)

^②(上海航天电子技术研究所 上海 201109)

摘要: 现有的基于雷达传感器的人体动作识别研究主要聚焦于相对雷达径向运动产生的微多普勒特征。当面对非径向,特别是静态姿势或者运动方向与雷达波束中心垂直的切向动作(切向人体姿态)时,传统基于微多普勒的方法无法对径向运动微弱的切向人体姿态进行有效表征,导致识别性能大幅下降。为了解决这一问题,该文提出了一种基于多发多收(MIMO)雷达成像图序列的切向人体姿态识别方法,以高质量成像图序列的形式来表征切向姿态的人体轮廓结构及其动态变化,通过提取图像内的空间特征和图序列间的时序特征,实现对切向人体姿态的准确识别。首先,通过恒虚警检测算法(CFAR)定位人体目标所在距离门,接着,利用慢时滑窗将目标动作划分为帧序列,对每帧数据用傅里叶变换和二维Capon算法估计出切向姿态的距离、俯仰角度和方位角度,得到切向姿态的成像图,将各帧成像图按照时序串联起来,构成切向人体姿态成像图序列;然后,提出了一种改进的多域联合自适应阈值去噪算法,抑制环境杂波,增强人体轮廓和结构特征,改善成像质量;最后,采用了一种基于空时注意力模块的卷积长短期记忆网络模型(ST-ConvLSTM),利用ConvLSTM单元来学习切向人体姿态成像图序列中的多维特征,并结合空时注意力模块来强调成像图内的空间特征和图序列间的时序特征。对比实验的分析结果表明,相比于传统方法,该文所提出的方法在8种典型的切向人体姿态的识别中取得了96.9%的准确率,验证了该方法在切向人体姿态识别上的可行性和优越性。

关键词: MIMO雷达; 切向人体姿态识别; 成像图序列; 图像去噪; 深度学习

中图分类号: TN957.52

文献标识码: A

文章编号: 2095-283X(2024)x-0001-17

DOI: 10.12000/JR24116

引用格式: 丁传威, 刘芷麟, 张力, 等. 基于MIMO雷达成像图序列的切向人体姿态识别方法[J]. 雷达学报(中英文), 待出版. doi: 10.12000/JR24116.

Reference format: DING Chuanwei, LIU Zhilin, ZHANG Li, *et al.* Tangential human posture recognition with sequential images based on MIMO radar[J]. *Journal of Radars*, in press. doi: 10.12000/JR24116.

Tangential Human Posture Recognition with Sequential Images Based on MIMO Radar

DING Chuanwei^① LIU Zhilin^① ZHANG Li^② ZHAO Heng^① ZHOU Qing^①

HONG Hong^{*①} ZHU Xiaohua^{*①}

^①(Nanjing University of Science and Technology, Nanjing 210000, China)

^②(Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China)

Abstract: Recent research on radar-based human activity recognition has typically focused on activities that

收稿日期: 2024-06-05; 改回日期: ; 网络出版: 2024-xx-xx

*通信作者: 洪弘, hongnju@njjust.edu.cn; 朱晓华, zxh@njjust.edu.cn

*Corresponding Authors: HONG Hong, hongnju@njjust.edu.cn; ZHU Xiaohua, zxh@njjust.edu.cn

基金项目: 国家自然科学基金(62201259, 62301255), 中央高校基本科研业务费专项资金资助(30923011006, 30923011026)

Foundation Items: The National Science Foundation of China (62201259, 62301255), Fundamental Research Funds for the Central Universities (30923011006, 30923011026)

责任编辑: 金添 Corresponding Editor: JIN Tian

©The Author(s) 2024. This is an open access article under the CC-BY 4.0 License

(<https://creativecommons.org/licenses/by/4.0/>)

move toward or away from radar in radial directions. Conventional Doppler-based methods can barely describe the true characteristics of nonradial activities, especially static postures or tangential activities, resulting in a considerable decline in recognition performance. To address this issue, a method for recognizing tangential human postures based on sequential images of a multiple-input–multiple-output (MIMO) radar system is proposed. A time sequence of high-quality images is achieved to describe the structure of the human body and corresponding dynamic changes, where spatial and temporal features are extracted to enhance the recognition performance. First, a constant false alarm rate algorithm is applied to locate the human target. A sliding window along the slow time axis is then utilized to divide the received signal into sequential frames. Next, a fast Fourier transform and the 2D Capon algorithm are performed on each frame to estimate range, pitch angle, and azimuth angle information, which are fused to create a tangential posture image. They are connected to form a time sequence of tangential posture images. To improve image quality, a modified joint multidomain adaptive threshold–based denoising algorithm is applied to improve the image quality by suppressing noises and enhancing human body outline and structure. Finally, a spatiotemporal convolution long short-term memory (ST-ConvLSTM) network is designed to process the sequential images. In particular, the ConvLSTM cell is used to extract continuous image features by combining convolution operation with the LSTM cell. Moreover, spatial and temporal attention modules are utilized to emphasize intraframe and interframe focus for improving recognition performance. Extensive experiments show that our proposed method can achieve an accuracy rate of 96.9% in classifying eight typical tangential human postures, demonstrating its feasibility and superiority in tangential human posture recognition.

Key words: MIMO radar; tangential human posture recognition; sequential images; image denoising; deep learning

1 引言

人体动作识别技术在安防监控、搜索救援、老年人看护、人机交互等领域具有较高的研究价值和深远的应用前景,越来越受到国内外学者的关注^[1-5]。现今常用的人体动作识别传感器主要分为两种:一种是基于加速度计、陀螺仪、压力计等可穿戴式传感器,另一种是基于摄像头的视频传感器。然而,前者需要时刻佩戴相应设备,不仅会影响日常生活,还容易遗失,其精确度也存在固有缺陷,虚警误警率高,不适用于长期监测;基于摄像头成像的视频方法虽然实现了非接触式监测,但是视频传感器不仅容易受到光照条件和室内障碍物的影响,更为严重的是,会给隐私安全带来威胁。

雷达传感器由于其在非接触、准确性、鲁棒性和隐私安全性等方面的优势和潜力成为近年来人体动作识别技术的研究热点^[6-11]。它可以通过目标反射回波中的频率变化,分析人体运动过程中产生的多普勒效应,从而对目标的运动状态进行估计和判断。这构成了基于雷达传感器的人体动作识别的理论基础^[12-15]。

近期基于雷达传感器的人体动作识别研究主要聚焦在沿着雷达波束径向方向朝向雷达或者背离雷达运动的径向人体动作^[16-19]。然而,实际的应用情景中需要处理的更多是非径向的人体动作。其运动方向与雷达波束中心存在夹角,这导致了测量到的

径向多普勒信息出现波动乃至失真,无法准确描述目标动作的真实特征^[20,21]。为了解决这一问题,目前的方法主要分为多站雷达方案和单站雷达方案。

基于多站雷达的解决方案通过在不同地点和方位布置多台雷达来同时录取到不同方向的多视角信息。这些多视角信息可以确保至少有一个合适的观察角度来录取到目标人体动作的有效信息,再利用数据层、特征层和决策层的多层级融合算法,实现了目标人体动作的有效信息获取和特征提取,减少了人体动作运动方向对识别性能的影响^[22-24]。文献[25]搭建了一套三基地的多站雷达组网系统,通过融合不同基地雷达的录取数据实现了对不同方向武装和非武装人员的甄别。文献[26]通过布置双站雷达系统来估计目标人体的运动方向,从而增强多普勒特征的提取。文献[27, 28]通过将多站雷达系统得到的多视角数据分别在特征层和决策层进行融合,极大增强了目标人体动作的识别性能。然而,多站雷达方案对设备布置、场地占用和计算复杂度等方面要求很高,多站系统间的校准也会直接影响其性能。这些因素都限制了多站雷达方案在实际场景中的应用。

基于单站雷达的解决方案主要依赖于多输入多输出雷达(Multiple-Input Multiple-Output, MIMO)。利用MIMO雷达的测角功能,来获取目标人体动作的角度信息,通过多次录制不同视角的信息来完备

不同方向人体动作的数据集，并在此基础上结合深度学习网络，利用多视角数据集训练出对角度不敏感的分类识别模型。文献[29]中利用MIMO雷达测量到的角度信息来构建目标的距离角度图来实现细微手势动作的识别。文献[30]通过搭建一个4D毫米波雷达来提取目标的距离、多普勒、角度、能量等信息，并结合多层次分类结构，实现了对不同角度人体动作87.1%的识别准确率。文献[31]和文献[32]构建了一个图像翻译框架，通过录取的径向动作数据来生成不同方向的动作数据，从而减轻了微多普勒信息对人体动作运动方向的敏感度。文献[33]提出了一种小样本学习模型，利用元数据增强策略和基于局部-全局相似性评分的度量学习框架，通过对径向动作微多普勒信息的训练学习，来实现对不同方向人体动作的识别。

现有的多站和单站雷达方案主要聚焦于人体动作的微多普勒特征，通过补充多视角数据和搭建深度学习网络模型来降低人体动作微多普勒特征的角度敏感度。这种方法增大了数据录取的负担和模型复杂度。而且，实际应用场景中经常出现静态的人体姿势或者运动方向垂直于雷达波束中心的人体动作，我们将其定义为切向人体姿态。这些切向人体姿态在雷达径向上的运动十分微弱，传统的微多普勒特征难以准确描述其动作特征，从而导致切向人体姿态的识别性能大幅下降。近年来随着集成电路技术的发展，雷达传感器在小型化的同时，可以容纳更多的阵列天线，使其距离和角度分辨力都得到了显著提高。这为基于雷达成像的切向人体姿态识别方案提供了硬件可能。其中，雷达图像的质量直接关系到识别性能。传统的人体动作识别算法中，经常采用动目标显示算法(Moving Target Indication, MTI)，利用环境杂波的时间平稳性，通过时间轴相消，来实现人体动作图像噪声去除^[1]。然而，在面对切向人体姿态时，切向动作或者静止姿态相对于雷达的运动十分微弱，很难与静止杂波区分开来。这对基于雷达的人体成像去噪算法提出了新的挑战。三维块匹配协同滤波算法(Block-Matching and 3D Filtering, BM3D)作为图像去噪算法领域的集大成者，利用非局部均值去噪算法(Non-local Means, NLM)的块匹配思想，对参考块进行相似块匹配并组合成三维块，接着对三维块进行空域变换后进行协同滤波处理，最后将得到的图像块的估计值加权平均代替参考块，以实现图像去噪^[34]。该算法在处理图像细节及高斯白噪声时效果较好，但基于匹配块的方法容易丢失边缘细节^[35]。为此，可以通过结合平稳小波变换(Stationary Wavelet Trans-

form, SWT)，利用其平移不变性来增强切向人体姿态图像的边缘细节特征^[36]。

因此，本文提出了一种基于MIMO雷达成像图序列的切向人体姿态识别方法。首先，通过MIMO雷达的多通道回波信号中的距离、俯仰角和方位角信息，对人体姿态进行三维成像，并结合时间滑窗引入图像帧之间的时序信息，得到切向人体姿态的成像图序列；然后，通过改进的多域联合自适应阈值去噪算法，提高图像信噪比，改善成像质量；最后，设计了一种基于空时注意力模块的卷积长短期记忆网络模型，利用卷积长短期记忆网络单元来处理成像图序列的多维特征，并通过加入空时注意力模块来强调成像图内的空间特性和图序列之间的时序特性。充足完备的实验结果验证了本文所提方法在切向人体姿态识别上的可行性和优越性。本文的主要贡献如下：

1. 据我们所知，这是首个针对基于雷达传感器的切向人体姿态识别的研究。切向人体姿态被定义为静止姿势或运动方向垂直于雷达波束中心的切向动作。传统聚焦于动作微多普勒的方法无法对径向运动微弱的切向人体姿态进行有效表征和识别。本文提出了一种基于MIMO雷达成像图序列的切向人体姿态识别方法。以高质量成像图序列的形式来表征切向姿态的人体轮廓结构及其动态变化，通过提取图像内的空间特征和图序列间的时序特征，实现对切向人体姿态的准确识别。

2. 针对传统去噪算法难以将切向人体姿态与静止环境噪声区分的问题，提出了一种改进的多域联合自适应阈值去噪算法。利用基于自适应阈值的平稳小波变换滤波算法保留切向人体姿态图像的边缘细节特征，再结合三维块匹配协同滤波算法，进一步增强其结构特征，改善成像质量。

3. 搭建了一种基于空时注意力模块的卷积长短期记忆网络模型(Spatial-Temporal-ConvLSTM, ST-ConvLSTM)。在利用ConvLSTM单元学习人体成像图中多维特征的同时，结合空时注意力模块来强调切向人体姿态在空间上的分布特性和时间上的连贯性。

2 理论与方法

基于MIMO雷达的回波信号可以重排为快时间维、慢时间维和通道维的三维矩阵。在传统的基于距离和多普勒变化的人体动作识别方法中，对回波信号沿着快时间维进行傅里叶变换可以得到目标的时间-距离图。在时间-距离图上的目标所在距离门内沿着慢时间维度进行短时傅里叶变换可以得到目

标的时间-多普勒图。传统方法可以根据人体动作的距离和多普勒随着时间变化的规律来进行动作的识别,然而当目标动作静止的姿态或者垂直于雷达中心波束的切向动作时,它们微弱的距离和多普勒变化给传统方法带来了挑战。

图1为作为示例的“切向弯腰”姿态的传统特征谱图。其中,图1(a)为时间-距离图,可以观察到人体目标一直保持在距离雷达约1.7 m的位置,几乎没有显著的距离变化。图1(b)为时间-多普勒图,多普勒频率基本都集中在零频附近,没有其余的多普勒分量。这是因为“切向弯腰”姿态的运动方向垂直于雷达中心波束,在径向距离上的变化和径向速度接近于零。这使得传统方法中的距离和多普勒信息无法表征目标动作具有区分度的有效信息,给后续的动作识别带来困难。因此本文提出使用MIMO雷达成像图序列特征来描述切向人体姿态。

2.1 方法流程介绍

本文提出的基于MIMO雷达成像图序列的切向人体姿态识别方法流程图如图2所示。在数据采集部分,MIMO雷达接收人体反射的回波信号,每个通道内的数据存储方式是具有慢时间轴和快时间轴的二维矩阵,并和通道维一起构成三维数据矩阵。在雷达信号处理部分,首先在预处理中,采用恒虚

警检测算法(Constant False-Alarm Rate, CFAR)对人体目标所占据的距离门进行定位,接着,利用慢时滑窗将目标动作按时序进行划分,对每帧数据用二维Capon算法估计出人体姿态的俯仰角度和方位角度,并结合距离信息得到人体姿态成像图,将各帧的成像图按照时序串联起来,构成人体姿态成像图序列。然后,通过改进的多域联合自适应阈值去噪算法,增强人体轮廓细节和结构信息,改善成像图序列的质量。最后,在姿态识别部分,通过搭建基于空时注意力的卷积长短期记忆网络,在学习成像图序列多维特征的同时,强调图像内部的空间特性和图序列间的时序特性,得到切向人体姿态的识别结果。

2.2 基于MIMO雷达的切向人体姿态成像图序列

基于MIMO雷达的人体切向姿态成像图序列流程图如图3所示。

雷达目标回波信号是由快时间轴、慢时间轴和通道维构成的三维信号矩阵 $S(\tau, t, c)$ 。在预处理部分,首先在快时间轴上作傅里叶变换得到距离维信息,然后利用邻近单元平均恒虚警检测算法(Cell-Average Constant False-Alarm Rate, CA-CFAR),通过计算待测距离门单元周围的参考距离门单元的平均值来得到环境杂波的平均功率以确定阈值,从而

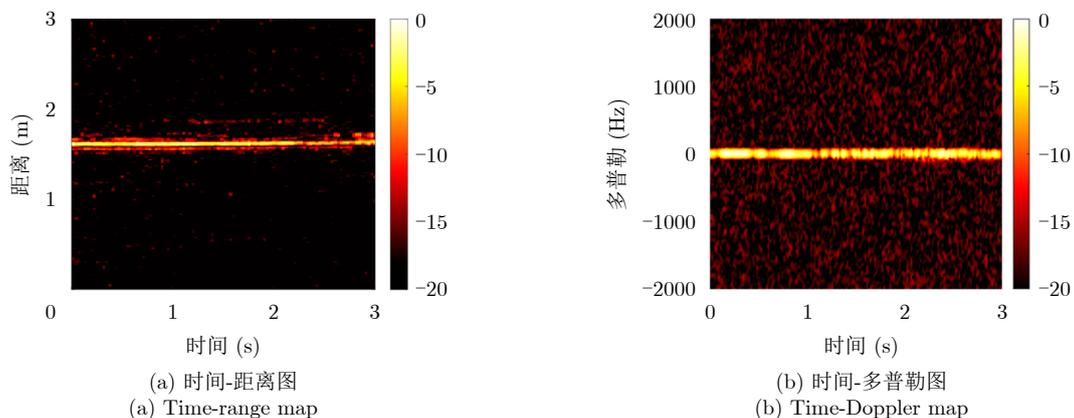


图1 “切向弯腰”姿态的传统特征谱图

Fig. 1 Conventional feature maps of the example "tangential bending" posture

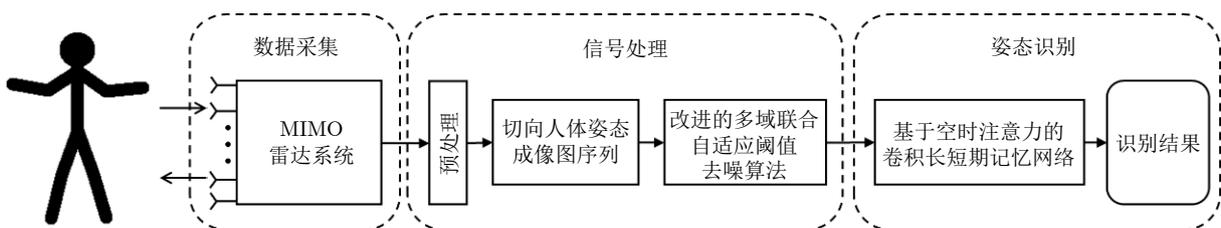


图2 基于MIMO雷达成像图序列的切向人体姿态识别方法流程图

Fig. 2 The framework of the proposed tangential human posture recognition with sequential images based on MIMO radar.

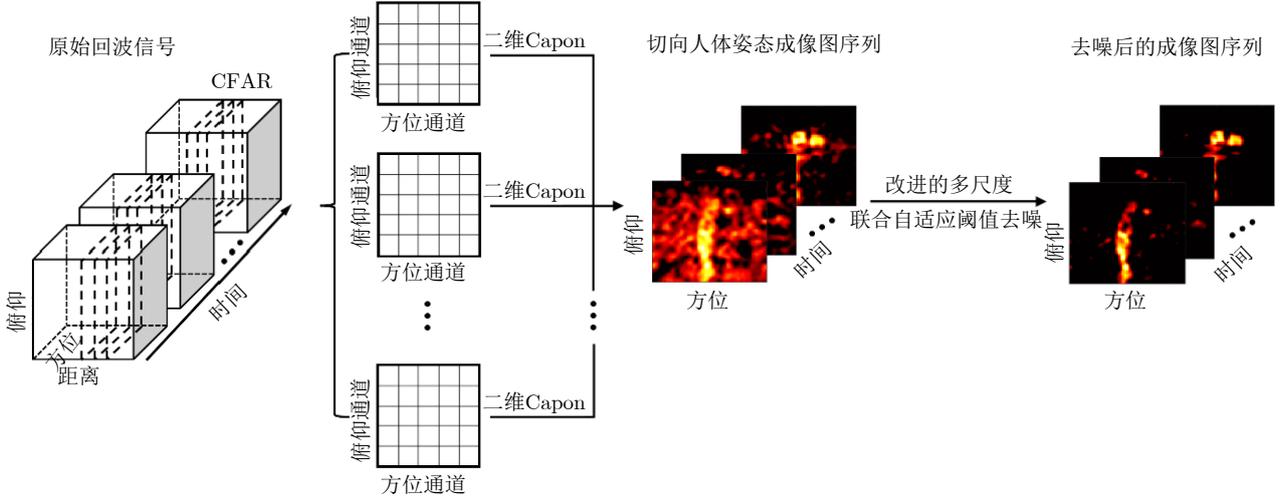


图 3 基于MIMO雷达的切向人体姿态成像图序列流程图

Fig. 3 The flow chart of tangential human posture sequential images based on MIMO radar

识别出人体目标所在的距离单元。CA-CFAR的检测阈值可以表示为

$$T_i = \alpha \cdot \frac{1}{2R} \left(\sum_{j=i-Q-R}^{j=i-Q-1} x_j + \sum_{j=i+Q+1}^{j=i+Q+R} x_j \right) \quad (1)$$

其中, T_i 为第 i 个待测距离单元的检测阈值, α 为阈值因子, R 为参考距离单元数目, Q 为保护距离单元数目, x_j 为第 j 个距离单元回波信号。当待测距离单元的功率大于检测阈值时, 则确定其为目标所在距离单元。由于人体自身具有一定的厚度, 所在距离单元不唯一, 将所有潜在的目标距离单元求和, 作为人体目标所在的距离门。接着, 将目标所在距离门信号在慢时间轴上利用滑窗划分为按照时序排列的帧序列, 并在通道维度上区分出俯仰向通道和方位向通道。这样, 可以得到目标所在距离的按照时序排列的方位通道和俯仰通道二维矩阵序列 S_p , 其中, $p = [1, 2, \dots, P]$ 为成像图索引, 即帧数。对每一帧序列上的方位-俯仰二维矩阵进行二维Capon角度估计^[37]。

二维Capon的角度估计问题可表示为

$$\begin{cases} \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{SS} \mathbf{w} \\ \mathbf{w}^H \mathbf{a}(\theta_d, \phi_d) = 1 \end{cases} \quad (2)$$

其中, \mathbf{w} 为加权矢量, $\mathbf{R}_{SS} = E \{S_p S_p^H\}$ 为接收阵列信号的协方差矩阵, $\mathbf{a}(\theta_d, \phi_d)$ 为信号的导向矢量。利用Lagrange乘子法求得上述带有约束条件最小化问题的解为

$$\mathbf{w}_{\text{Capon}} = \frac{\mathbf{R}_{SS}^{-1} \mathbf{a}(\theta_d, \phi_d)}{\mathbf{a}^H(\theta_d, \phi_d) \mathbf{R}_{SS}^{-1} \mathbf{a}(\theta_d, \phi_d)} \quad (3)$$

最终可以得到二维Capon空间谱函数为

$$P_{\text{Capon}}(\theta, \phi) = \frac{1}{\mathbf{a}^H(\theta_d, \phi_d) \mathbf{R}_{SS}^{-1} \mathbf{a}(\theta_d, \phi_d)} \quad (4)$$

之后在 $\theta \in [-90^\circ, 90^\circ]$ 和 $\phi \in [-90^\circ, 90^\circ]$ 的范围内对 $P_{\text{Capon}}(\theta, \phi)$ 进行谱峰搜索, 峰值点即对应目标信号的方位角 θ 和俯仰角 ϕ 的估计值。结合预处理阶段的目标所在距离, 即可以得到切向人体姿态的方位-俯仰成像图序列。

2.3 改进的多域联合自适应阈值人体成像图去噪算法

由于切向人体姿态在径向上的距离变化和多普勒都很微弱, 这导致基于运动目标和静止杂波区别的传统去噪算法难以将切向人体姿态从静止环境杂波中区分出来。本文提出了一种改进的多域联合自适应阈值人体成像图去噪算法。所提出的算法充分利用平稳小波滤波能够保持目标边缘的优势, 对人体成像图的细节特征进行保留。接着将其与改进的三维块匹配协同滤波算法BM3D相结合, 对人体成像图做进一步结构特征增强。具体算法流程如图4所示。

改进的多域联合自适应阈值人体成像图去噪算法分为3个阶段, 分别是平稳小波基础滤波阶段、BM3D基础估计阶段和BM3D最终估计阶段。

(1) 平稳小波基础滤波阶段

此阶段首先对原始成像图进行平稳小波变换分解。SWT首先通过Symlet4小波基函数进行小波变换。确定了小波基函数之后, 需要确定分解层数。通过对输入的原始人体成像图进行滑窗, 计算每个像素点周围的局部方差。再根据局部方差的中位数, 估计图像噪声的标准差。将噪声标准差和预设的阈值进行比较, 来自适应调整小波分解的层数。最后得到不同频率范围的小波系数。二维平稳小波变换公式为

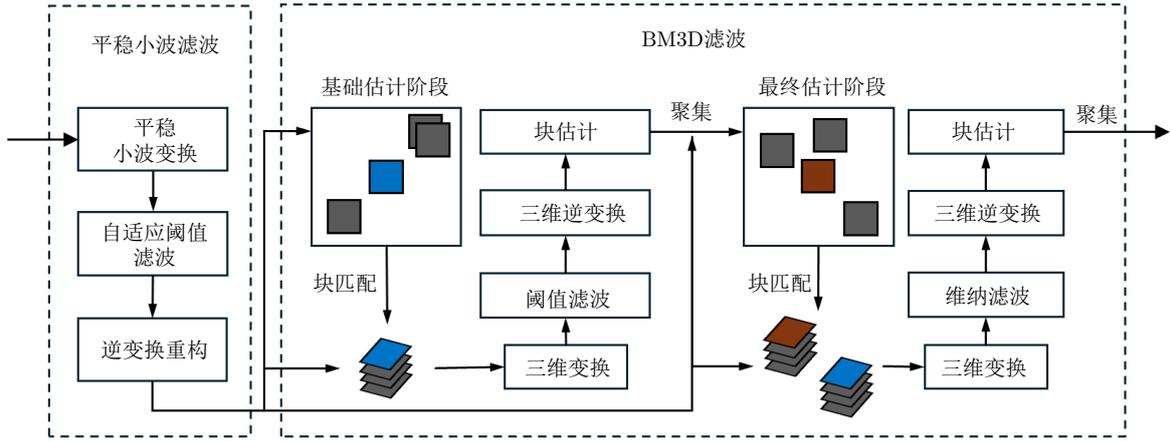


图4 改进的多域联合自适应阈值人体成像图去噪算法流程图

Fig. 4 The flow chart of modified joint multi-domain adaptive threshold-based image denoising algorithm

$$\mathbf{W}_{\varphi}^k(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j,m,n}^k(x, y) \quad (5)$$

其中, $\varphi_{j,m,n}^k(x, y)$ 为小波基函数, j 表示小波层数, m, n 表示二维小波尺度, k 表示不同方向的子带。

然后对人体成像图经平稳小波变换分解成不同尺

$$\omega_{\lambda_j} = \begin{cases} \tanh(\omega_j) \left\{ |\omega_j| - \lambda_j \left[1 - \exp\left(\frac{-h}{|\omega_j| - \lambda_j}\right) \right] \right\}, & |\omega_j| \geq \lambda_j \\ \tanh(\omega_j) \left[|\omega_j| - \frac{\lambda_j}{\exp(|\omega_j| - \lambda_j)} \right], & |\omega_j| < \lambda_j \end{cases} \quad (6)$$

其中, ω_{λ_j} 为小波估计系数, ω_j 为小波变换系数, λ_j 表示随层数 j 变换的阈值, h 为调节因子。为了更好地适应实际情况, 可将 h 的取值范围调整为 0 到 1。

针对阈值 λ_j 的选取, 本文采用了一种利用噪声特征自适应计算阈值的方法。可以根据不同的分解层自动选择合适的阈值, 更好地抑制和去除每一层的噪声, 修改后的阈值如下:

$$\lambda_j = \frac{\sigma_j \sqrt{2 \lg N_j}}{\ln(j + 2^{j-1})} \quad (7)$$

其中, N_j 表示图像的尺寸, σ_j 表示图像估计噪声的标准差。最后再经过平稳小波逆变换对信号进行重构, 得到平稳小波滤波后的人体成像图。

(2) BM3D 基础估计阶段

为了方便说明 BM3D 基础估计阶段和 BM3D 最终估计阶段的不同参数, 基础估计阶段的参数用“t”标识, 最终估计阶段的参数用“wie”标识。在 BM3D 基础估计阶段, 首先对平稳小波滤波后的成像图进行相似块匹配, 按照固定步长, 对成像图遍历行和列, 选中参考块。在参考块周围的指定邻域内, 继续遍历每一个点, 将每一个点都再作为中心点, 向四周扩散出对应的匹配块, 之后计算参考

度上的近似系数和细节系数进行阈值滤波处理。针对阈值处小波系数的复杂性, 提出了一种改进的阈值函数, 既能使得图像中有效小波系数保持不变, 保留图像的边缘信息; 又能在阈值处要保持良好的连续性, 这样可以使去噪后图像更加平滑, 更好地减少图像的伪吉布斯现象。改进的阈值函数表达式为

块和匹配块之间的相似度, 符合相似度要求的块为该参考块的相似块。

参考块和待匹配块之间的相似度计算采用的是欧式距离。 $\hat{\mathbf{F}}$ 代表平稳小波基础滤波后的人体成像图, \mathbf{Z} 为 $\hat{\mathbf{F}}$ 中的参考块, \mathbf{Z}_x 是参考块 \mathbf{Z} 中的一个匹配块, 则 \mathbf{Z} 和 \mathbf{Z}_x 之间的欧式距离表示为

$$d(\mathbf{Z}, \mathbf{Z}_x) = \frac{\|\mathbf{Z} - \mathbf{Z}_x\|_2^2}{(N^t)^2} \quad (8)$$

其中, $(N^t)^2$ 表示图像块尺寸, $\|\cdot\|_2$ 表示 L_2 范数。当匹配块与参考块的距离小于固定阈值时, 则将该匹配块归类于参考块的相似块。否则, 就不是相似块。依据该判断准则, 可以求出 \mathbf{Z} 的所有匹配块集合为

$$N_{\max}^t = \{x \in \hat{\mathbf{F}} : d(\mathbf{Z}, \mathbf{Z}_x) \leq \tau_{\text{match}}^t\} \quad (9)$$

其中, τ_{match}^t 表示设定的相似块间的最大欧式距离阈值, N_{\max}^t 表示每个匹配块集合中相似块的数目。将找到的相似块聚集到一起, 这样就构成了若干个匹配块集合。

接着对匹配块集合按照距离由小到大排序, 得到三维矩阵 $\mathbf{Z}_{N_{\max}^t}$, 前两维进行二维离散余弦变

换，对第3个维度进行哈达玛变换。利用自适应阈值函数滤波减弱噪声后，通过逆变换得到图像匹配块集合的基本估计值为

$$\hat{\mathbf{Y}}_{N_{\max}^t}^t = \mathbf{T}_{3D}^{t-1} (\gamma (\mathbf{T}_{3D}^t (\mathbf{Z}_{N_{\max}^t}))) \quad (10)$$

其中， \mathbf{T}_{3D}^t 表示三维变换， \mathbf{T}_{3D}^{t-1} 表示三维逆变换， γ 表示阈值滤波运算。在三维变换之后，还要计算聚集时的权重 w^t 。如果用 N_{har} 表示匹配块集合中所有非零元素数目的总和，将所有非零元素的数目作为该块的权重值，则权重公式可以表示为

$$w^t = \begin{cases} \frac{1}{\sigma^2 N_{\text{har}}}, & N_{\text{har}} \geq 1 \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

其中， σ^2 表示噪声方差。在遍历参考块的过程中，每个参考块会对应多个估计值，对同一参考块的不同估计值进行加权平均作为该参考块的最终估计值，并将其称之为聚集过程。通过聚集可以得到基础估计的结果，表达式为

$$\hat{\mathbf{Y}}^{\text{basic}}(x) = \frac{\sum_{x \in I} \sum_{x_m \in N_{\max}^t} w^t \hat{\mathbf{Y}}_{x_m}^{t,x}(x)}{\sum_{x \in I} \sum_{x_m \in N_{\max}^t} w^t \chi_{x_m}(x)} \quad (12)$$

其中， χ_{x_m} 是图像相似块的特征函数， x 表示一个匹配块， $\hat{\mathbf{Y}}_{x_m}^{t,x}(x)$ 为基础估计阶段图像匹配块集合中 x 位置上的估计结果。

(3)BM3D最终估计阶段

在BM3D的最终估计阶段，利用BM3D基础估计值 $\hat{\mathbf{Y}}^{\text{basic}}(x)$ 进行最终估计处理。将基础估计的人体成像图经过滑窗划分成固定大小的参考块，块设置与BM3D基础估计阶段相同，对每个参考块进行块匹配和协同维纳滤波。从而得到参考块的匹配块集合为

$$N_{\max}^{\text{wie}} = \left\{ x \in \hat{\mathbf{F}} : \frac{\|\hat{\mathbf{Y}}^{\text{basic}} - \hat{\mathbf{F}}\|_2^2}{(N^{\text{wie}})^2} < \tau_{\text{match}}^{\text{wie}} \right\} \quad (13)$$

其中， $\tau_{\text{match}}^{\text{wie}}$ 为设定的最大欧式距离阈值， $(N^{\text{wie}})^2$ 为图像块尺寸。

将匹配块集合进行排序得到三维矩阵 $\mathbf{Z}_{N_{\max}^{\text{wie}}}$ ，经三维变换后，采用维纳滤波将图像形成的三维矩阵进行系数收缩。维纳收缩系数表达式为

$$w_{N_{\max}^{\text{wie}}} = \frac{\left| \mathbf{T}_{3D}^{\text{wie}} (\hat{\mathbf{Y}}_{N_{\max}^{\text{wie}}}^{\text{basic}}) \right|^2}{\left| \mathbf{T}_{3D}^{\text{wie}} (\hat{\mathbf{Y}}_{N_{\max}^{\text{wie}}}^{\text{basic}}) \right|^2 + \sigma^2} \quad (14)$$

其中， $\hat{\mathbf{Y}}_{N_{\max}^{\text{wie}}}^{\text{basic}}$ 为基础估计阶段估计值。对 $\mathbf{Z}_{N_{\max}^{\text{wie}}}$ 作三维变换后与维纳收缩系数 $w_{N_{\max}^{\text{wie}}}$ 相乘，再利用三维逆变换得到位于 $x \in N_{\max}^{\text{wie}}$ 处匹配块集合的估计值为

$$\hat{\mathbf{Y}}_{N_{\max}^{\text{wie}}}^{\text{wie}} = \mathbf{T}_{3D}^{\text{wie}-1} (w_{N_{\max}^{\text{wie}}} \mathbf{T}_{3D}^{\text{wie}} (\mathbf{Z}_{N_{\max}^{\text{wie}}})) \quad (15)$$

其中， $\mathbf{T}_{3D}^{\text{wie}}$ 表示三维变换， $\mathbf{T}_{3D}^{\text{wie}-1}$ 表示三维逆变换。

和基础估计阶段相似，在完成三维协同滤波之后需要对其进行聚集处理。此时加权的权重取决于维纳滤波的收缩系数和噪声方差，权重公式为

$$w^{\text{wie}} = \sigma^2 \|w_{N_{\max}^{\text{wie}}}\|_2^{-2} \quad (16)$$

最后将所有参考块基本估计的结果聚集就可以得到人体成像图的最终估计结果：

$$\hat{\mathbf{Y}}^{\text{final}}(x) = \frac{\sum_{x \in I} \sum_{x_m \in N_{\max}^{\text{wie}}} w^{\text{wie}} \hat{\mathbf{Y}}_{x_m}^{\text{wie},x}(x)}{\sum_{x \in I} \sum_{x_m \in N_{\max}^{\text{wie}}} w^{\text{wie}} \chi_{x_m}(x)} \quad (17)$$

图5为作为示例的“切向弯腰”人体姿态成像图序列。其中，图5(a)为用摄像头录取的光学参考图像序列，图5(b)为“切向弯腰”原始雷达成像图序列，图5(c)为“切向弯腰”去噪后雷达成像图序列。其中，在时间维度上的滑窗长度设置为0.5s，窗与窗之间有50%的重叠，一共截取了6帧的成像图序列，来覆盖“切向弯腰”整个动作过程。可以观察到，经过本文所提出的改进的多域联合自适应阈值去噪后，雷达成像图中的静止环境噪声得到了较大程度的滤除，图像中人体的结构得到了保留和增强，轮廓边缘更加清晰，整体成像质量得到了改善。能够从6帧图像的变化中显著地观察到弯腰动作中人体从站立到上半身弯曲至90度的整个过程，并且与光学参考图序列中的人体姿态吻合度高。

2.4 基于空时注意力模块的卷积长短期记忆网络

从图5的示例中可以观察到，切向人体姿态的特征不仅取决于每一帧图像中能量的空间分布，帧之间的时序变化信息也十分关键。因此，在姿态识别部分，本文提出了一种基于空时注意力模块的卷积长短时记忆网络的识别模型来同时增强帧内的空间特征和帧间的时序特征的学习。ST-ConvLSTM网络模型结构如图6所示。

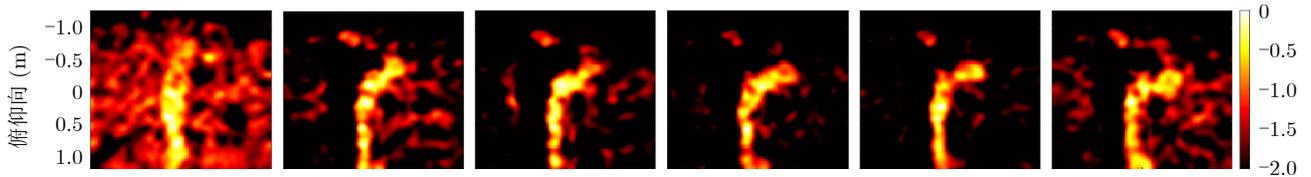
ST-ConvLSTM模型利用ConvLSTM神经网络单元在保留了人体成像图序列空间信息处理能力的同时，还能够捕获长短期时间信息。它相比于传统的三维卷积神经网络(3D Convolutional Neural Network, 3DCNN)网络更加灵活^[38]，相比于卷积+长短期记忆网络(Convolutional Neural Network + Long short term memory, CNN+LSTM)网络，通过用门单元取代全连接层使得可以更有效地表示时间序列中的空间结构^[39]。

ConvLSTM单元的结构图如图7所示，其公式可以表示为



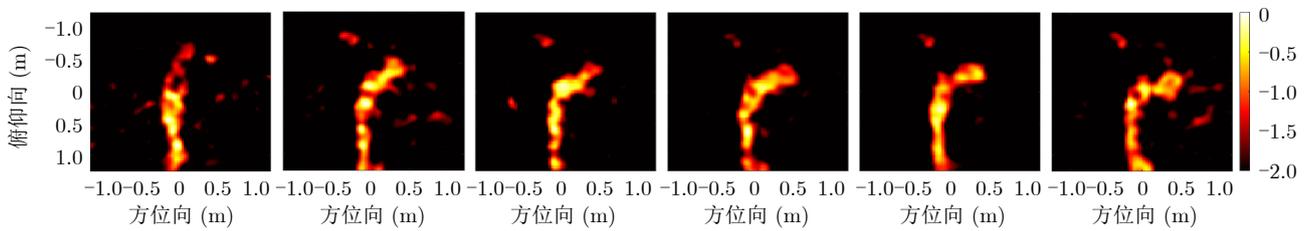
(a) 光学参考图像序列

(a) Referenced camera-based sequential images



(b) “切向弯腰”原始雷达成像图序列

(b) Radar-based Original sequential images



(c) “切向弯腰”去噪后雷达成像图序列

(c) Radar-based sequential images after denoising

图5 “切向弯腰”人体姿态成像图序列

Fig. 5 Sequential images of "tangential bending" human posture

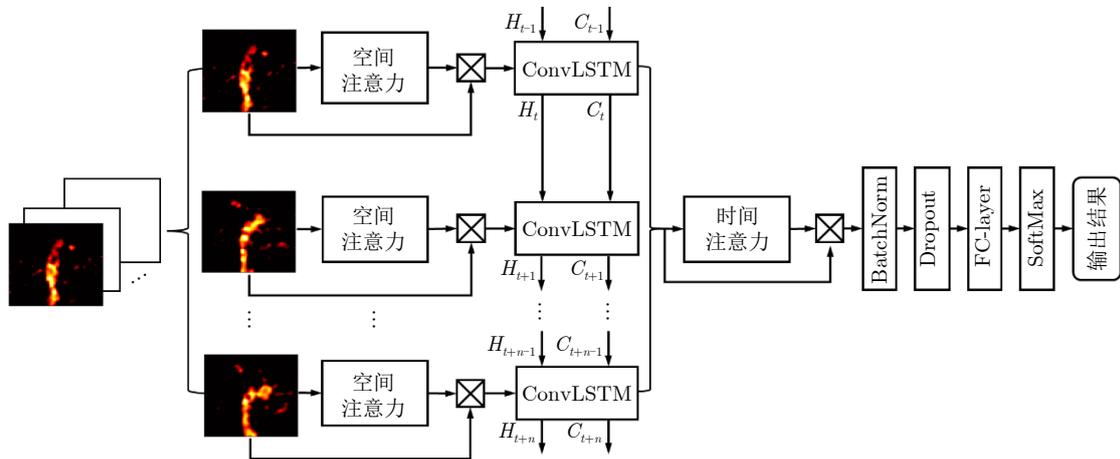


图6 ST-ConvLSTM网络结构图

Fig. 6 ST-ConvLSTM network structure diagram

$$\begin{aligned}
 i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \otimes C_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf} * X_t + W_{hf} * K_{t-1} + W_{cf} \otimes C_{t-1} + b_f) \\
 C_t &= f_t \otimes C_{t-1} + i_z \otimes \tanh(W_{xc} * \chi_t \\
 &\quad + W_{hc} * C_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo} * X_t + W_{ho} * K_{t-1} + W_{co} \otimes C_t + b_o) \\
 H_t &= o_t \otimes \tanh(C_t)
 \end{aligned} \quad (18)$$

其中, σ 为Sigmoid函数, W_x , W_h 和 W_c 为加权矩

阵, b 为偏置参数, $*$ 为卷积运算, \otimes 为矩阵对应位置元素乘积。

通过在ConvLSTM网络中加入空间和时间注意力模块,以综合分析切向人体姿态成像图序列帧内和帧间的信息,利用空间注意力模块来增强人体姿态的空间分布特征,设计时间注意力模块来获得人体姿态连续变化的最优时间权重。整个注意力过程可以概括为

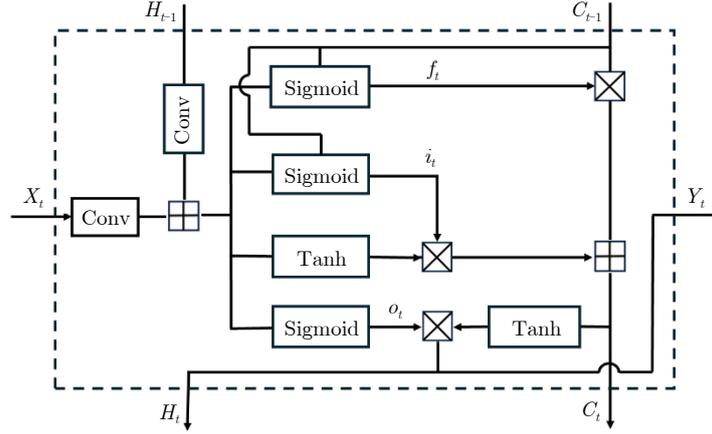


图 7 ConvLSTM单元结构图

Fig. 7 The architecture of ConvLSTM cell

$$\mathbf{F}'_{S/T} = \mathbf{M}_{S/T}(\mathbf{F}) \otimes \mathbf{F} \quad (19)$$

其中， \mathbf{F} 为中间特征映射，被用作输入， \mathbf{M}_S 和 \mathbf{M}_T 分别代表一维时间注意力图和二维空间注意力图。

空间注意力模块用于表示每个切向人体姿态成像图帧内的空间关系，重点关注图像内部的“位置”分布信息。空间注意力模块结构如图8所示。

将空间注意力模块与卷积层结合，直接应用于每帧切向人体姿态成像图，然后通过Sigmoid函数处理，空间注意力图 $\mathbf{M}_s(\mathbf{F})$ 用于增强帧内图像的人体轮廓信息，表达式为

$$\mathbf{M}_S(\mathbf{F}) = \sigma(\text{Conv}^{p \times p}(\mathbf{F})) \quad (20)$$

其中， $\text{Conv}^{p \times p}$ 表示卷积核为 $(p \times p)$ 的卷积操作。

时间注意力模块利用切向人体姿态帧之间的时序关系，重点关注整个人体姿态变化中特征最显著的时刻。时间注意力模块结构如图9所示。

通过最大池化和平均池化操作提取空间信息，

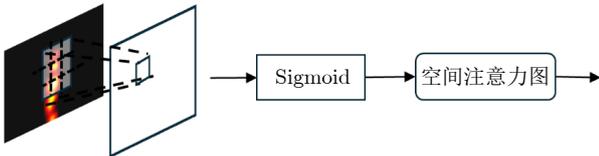


图 8 空间注意力模块结构图

Fig. 8 The architecture of the spatial attention module

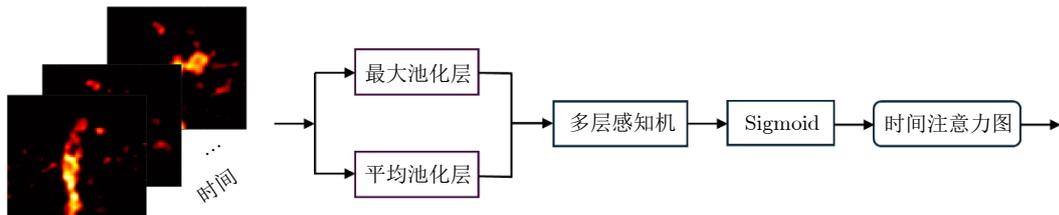


图 9 时间注意力模块结构图

Fig. 9 The architecture of the temporal attention module

并得到相应的最大池化特征和平均池化特征 $\mathbf{F}_{\text{avg}}^c$ 和 $\mathbf{F}_{\text{mac}}^c$ ，然后将它们传到多层感知器(Multi-layer perceptron, MLP)模块中，得到时间注意力图 \mathbf{M}_T 。MLP模块包括MLP和一个隐藏层。最后通过元素求和将输出的特征向量合并，计算出构成时间注意图的加权系数如下：

$$\mathbf{M}_T(\mathbf{F}) = \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F}) + \text{MaxPool}(\mathbf{F}))) \quad (21)$$

完成上述神经网络组件后，将提取的特征图输入以下分类模块进行最终的运动识别。第1层是批处理标准化层，它可以允许使用更高的学习速率；第2层是Dropout层，旨在防止过拟合，减少参数的数量；第3层为全连接层(Full Connected Layer, FC layer)；最后使用softmax函数作为激活函数，输出分类结果。

3 实验设置

3.1 硬件平台和数据录取

本文使用的雷达硬件平台是Mini-Circuits公司的毫米波雷达IMAGEVK-74，雷达实物图和系统结构如图10所示。

IMAGEVK-74雷达使用MIMO架构，天线阵列由20个发射天线和20个接收天线组成，可以合成

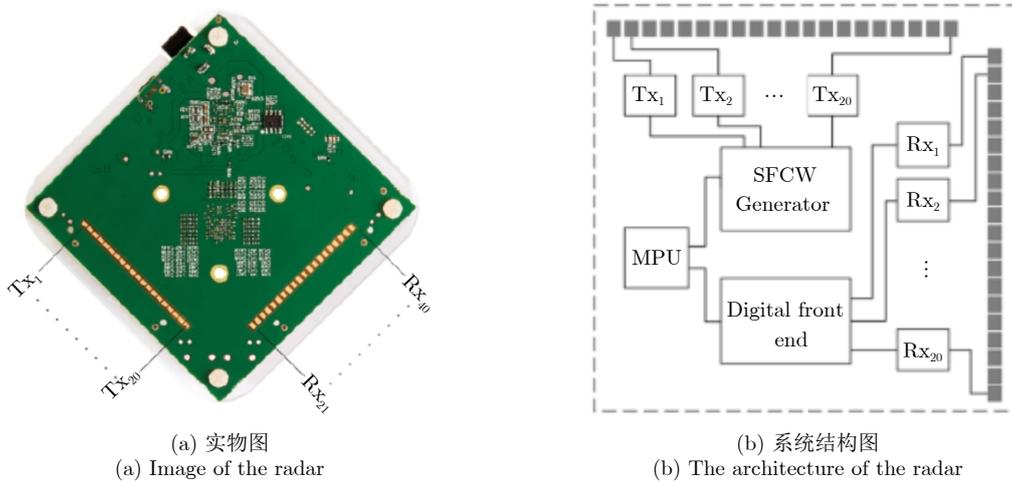


图 10 IMAGEVK-74雷达示意图
Fig. 10 IMAGEVK-74 radar schematic

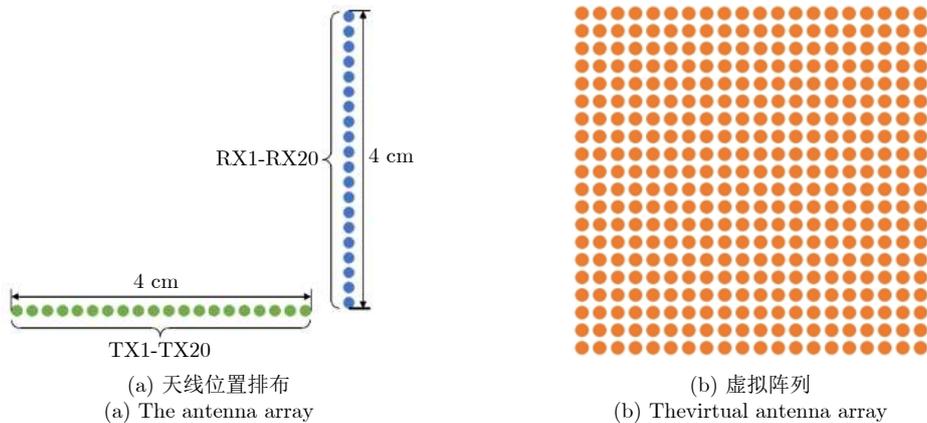


图 11 IMAGEVK-74天线示意图
Fig. 11 IMAGEVK-74 antenna array

为 20×20 的400个虚拟天线二维面阵,通过雷达信号角度估计算法,可以获得高分辨率的人体姿态成像数据。天线位置排布与虚拟阵列几何位置如图11所示。

本文实验场景如图12所示。实验地点为空旷的走廊,雷达放置在距离地面约1 m高度位置,受试者站在雷达正前方约1.5 m处。雷达系统的参数设置如表1所示。

为了验证本文方法的通用性和稳健性,本文选取了8种典型的切向人体姿态,分别是“站立”、“张开双臂”、“挥手”、“切向蹲下”、“切向弯腰”、“切向踏步”、“切向坐下”以及“切向踢腿”。其中,“站立”和“张开双臂”为静止姿态,其余的为切向动作。图13是这8种切向人体姿态的示意图,表2是其对应的具体描述。

本文实验中共采集了8位受试者的数据,包括5名男性和3名女性,年龄范围是23至26岁,体重范围为45公斤至85公斤,身高范围是1.61米至

1.80米。每人在相同场景每种姿态录取10次,使每种切向人体姿态在数据集中均匀分布。数据集总共包含640包数据。训练、验证和测试集比例为3:1:1。最后,利用合成少数过采样算法(Synthetic Minority Oversampling Technique, SMOTE)合成新样本,将训练集扩充为2000包数据,以解决随机过采样可能导致的模型过拟合问题。

3.2 实施细节

在特征提取算法方面,在CA-CFAR定位人体目标距离门时,设置保护距离单元数目为2,参考距离门单元数目为30,虚警概率为0.01;在获取切向人体姿态成像图序列时,在时间维度上的滑窗长度设置为0.5 s,窗之间有50%的重叠,一共截取了6帧的成像图序列来覆盖切向人体姿态的整个过程。

在深度学习网络方面,所有的网络模型都是基于Pytorch 3.10环境搭建的,运行在配置为AMD Ryzen 75800H处理器、NVIDIA GeForce RTX

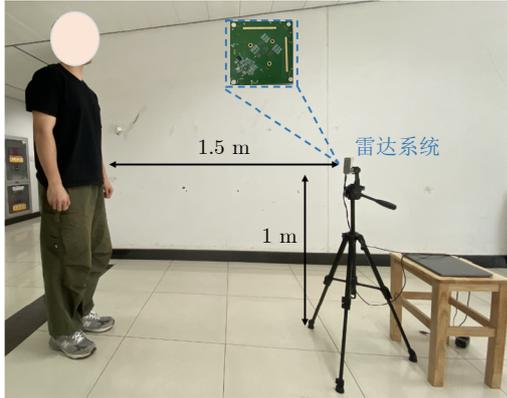


图 12 切向人体姿态识别实验场景图
Fig. 12 The experiment setup

表 1 IMAGEVK-74雷达配置参数
Tab. 1 IMAGEVK-74 Radar configuration parameters

参数	数值
起始频率	62 GHz
终止频率	66 GHz
中频带宽	100 MHz
频率步进	40 MHz
发射功率	-10 dBm
帧率	20 Hz
阵列数	20Tx, 20Rx
频率采样数	64
距离分辨率	3.75 cm
角度分辨率	6.7deg

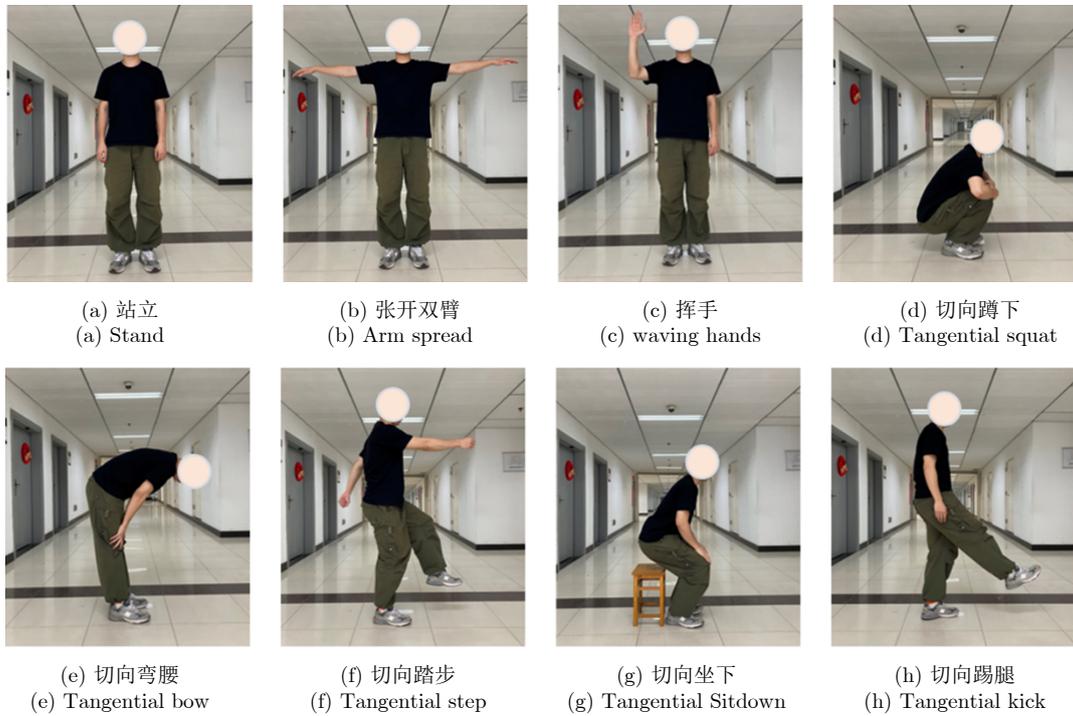


图 13 切向人体姿态示意图
Fig. 13 Illustrations of eight typical tangential human postures

表 2 切向人体姿态描述
Tab. 2 Descriptions of eight typical tangential human postures

种类	人体姿态	具体描述
(a)	站立	面对雷达原地静止站立
(b)	张开双臂	面对雷达原地张开双臂静止站立
(c)	挥手	面对雷达原地站立并左右挥动右手
(d)	切向蹲下	侧对雷达从站立开始蹲下
(e)	切向弯腰	侧对雷达从站立开始弯腰到90°
(f)	切向踏步	侧对雷达双臂和双腿前后摆动踏步
(g)	切向坐下	侧对雷达从站立开始向后坐下
(h)	切向踢腿	侧对雷达从站立开始向前缓慢踢出右腿

3060显卡、16G内存的处理平台。本文所采用的ST-ConvLSTM网络参数设置中,Dropout层的丢弃比设置为0.7,在全连接层采用ReLU函数作为激活函数,训练损失函数为交叉熵损失函数,优化器为Adam优化器。

4 结果与讨论

4.1 去噪性能

本文将典型的传统雷达图像去噪算法动目标显示(Moving Target Indicator, MTI)算法来作为参照组,同时对“张开双臂”这一切向人体姿势进行处理,其结果如图14所示。其中,图14(a)为去噪前的原始成像图,图14(b)为利用MTI算法去噪后的成像图,图14(c)为利用本文所提去噪算法得到的结果。通过对比观察可知,MTI算法处理后的成像图里人体下半身和身体头部区域信息丢失十分严重。这是因为MTI算法是利用环境噪声的时间平稳性,通过时间轴上的对消操作,剔除静态杂波,突出运动的目标信号。然而,“张开双臂”等切向人体姿态在径向上的运动十分微弱,特别是身体的下半身和头部几乎为静止,容易被误认为静态杂波而被剔除。相对的,本文所提的改进的多域联合自适应阈值去噪算法能够在保留几乎绝大部分人体轮廓和结构的基础上,抑制了环境杂波的干扰,改善了成像质量。为后续的姿态识别提供了高分辨的切向人体姿态成像图。

4.2 识别性能

分别将传统的时间-距离图、时间-多普勒图和本文所提的成像图序列作为特征输入到不同的深度学习网络模型中,对其各方面性能进行比较和评估。输入特征、去噪算法、选用模型及各识别方法的比较参数汇总如表3所示。其中,准确率是各识

别方法对本文切向人体姿态数据的平均识别准确率,代表了识别性能;耗时是每一包测试数据从输入到所用模型到输出识别结果的平均时间,反映了实时处理性能;模型尺寸则代表了所用网络模型的存储大小。

可以观察到,传统时间-距离图和时间多普勒图由于无法对切向人体姿态的特征进行准确描述,其对应的识别方法的平均准确率最低,分别为73.08%和70.19%。相对于传统的时频信息,基于图像的特征更加适合描述切向人体姿态。因此,当采用切向人体姿态成像图序列作为输入特征时,识别准确率得到了显著提高。

在去噪算法对比方面,将原始成像图序列、MTI去噪后、SWT去噪后、BM3D去噪后和改进的多域联合自适应阈值去噪后的特征分别输入ST-ConvLSTM模型中进行对比。其中,原始成像图序列的平均识别准确率为91.4%。除了MTI算法外,其余的去噪算法的引入均获得了识别性能的提升。这是因为切向人体姿态信号与静止杂波的在时间平稳性上区别不大,MTI算法反而会将会人体静态部位当作杂波剔除,其识别准确率反而下降,这与4.1节的分析结果吻合。注意到,MTI和SWT的耗时很少,而BM3D由于需要进行块搜索和匹配,运算时间较长。而我们所提的改进的多域联合自适应阈值去噪算法结合了SWT和BM3D的优势,虽然运算时间和BM3D相近,但是在结合相同的网络模型的情况下,识别准确率最高。

在基于成像图序列的识别网络模型对比方面,3DCNN模型的计算耗时最长,达到了2.38毫秒,平均识别准确率为90.6%。同样作为对照组的CNN+LSTM模型的计算耗时最小,但是它的模型尺寸最大,对应的平均识别准确率为91.4%。为了验证注意力模块的效果,这里还对本文所提的ST-ConvL-

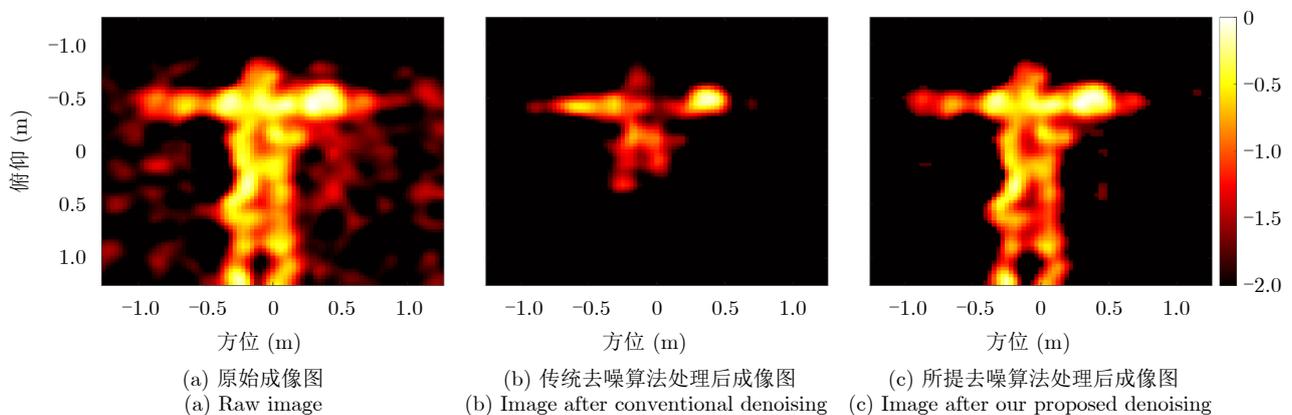


图14 示例“张开双臂”切向姿态去噪前后对比图

Fig. 14 Images of “arm spread” posture before and after denoising processing

表 3 切向人体姿态识别方法分类结果汇总

Tab. 3 Results of tangential human postures recognition methods

方法	输入特征	去噪算法	去噪耗时(s)	模型	模型耗时(ms)	模型尺寸(kB)	准确率
1	时间-距离图	N/A	N/A	CNN ^[40]	2.60	2602	72.7%
2	时间-多普勒图	N/A	N/A	CNN ^[40]	2.60	2602	70.3%
3	成像图序列	N/A	N/A	ST-ConvLSTM	1.93	151	91.4%
4	成像图序列	MTI	0.13	ST-ConvLSTM	1.93	151	89.8%
5	成像图序列	SWT	0.03	ST-ConvLSTM	1.93	151	92.1%
6	成像图序列	BM3D	4.40	ST-ConvLSTM	1.93	151	93.0%
7	成像图序列	Proposed	4.61	3DCNN ^[41]	2.38	3806	90.6%
8	成像图序列	Proposed	4.61	CNN-LSTM ^[42]	0.92	41299	91.4%
9	成像图序列	Proposed	4.61	ConvLSTM	1.29	148	93.8%
10	成像图序列	Proposed	4.61	S-ConvLSTM	1.62	150	94.5%
11	成像图序列	Proposed	4.61	T-ConvLSTM	1.35	148	95.3%
12	成像图序列	Proposed	4.61	ST-ConvLSTM (Proposed)	1.93	151	96.9%

STM进行了消融实验，对比了原始的ConvLSTM模型、加入空间注意力模块的卷积长短期记忆网络(Spatial-ConvLSTM, S-ConvLSTM)、加入时间注意力模块的卷积长短期记忆网络(Temporal-ConvLSTM, T-ConvLSTM)和同时考虑空间和时间注意力的ST-ConvLSTM模型。可以观察到，成像图序列和基于ConvLSTM的模型的组合均获得了较高的切向人体姿态识别准确率，且它们在耗时和模型尺寸等方面的性能都十分接近。其中，ConvLSTM模型作为基准，获得了93.8%的平均识别准确率。当加入一个空间注意力模块，构成S-ConvLSTM后，准确率达到94.5%。同样地，加入一个时间注意力模块构成T-ConvLSTM模型，准确率提高了约1.5%。最后，同时考虑空间和时间注意力，得到本文所提出的ST-ConvLSTM模型，获得了最高的96.9%的平均识别准确率。它对应的混淆矩阵图如图15所示。其中，各切向人体姿态的识别准确率都达到了85%以上。(a)站立、(b)张开双臂、(c)切向蹲下、(d)切向坐下、(e)切向踢腿这五种切向人体姿态没有数据被误分类，识别准确率达到100%。但(f)“切向踏步”姿态的识别准确率仅为87.5%，分别被错误的归类为(g)“切向坐下”和(h)“切向踢腿”姿态，踏步的抬腿和落腿动作使得它们分别在俯仰维度和方位维度上有着相似的特征。

表3中切向人体姿态识别的传统方法和所提方法网络特征的t-SNE(t-distributed stochastic neighbor embedding, t-SNE)二维可视化结果如图16所示。可以观察到，由于本文所研究的切向人体姿态在传统的TR图和TD图上特征不明显，因此基于

TR图和TD图的t-SNE二维可视化结果Fig. 16(a)和Fig. 16(b)的聚类效果最不理想。Fig. 16(g)中，3DCNN网络由于没有考虑到成像图序列之间的时序信息，8种切向人体姿态的二维可视化特征虽然呈现出一定的聚类特性，但各类样本之间的分布较为紧凑乃至重叠，分类边界不清晰。相应地，当深度学习模型通过加入LSTM单元构成CNN-LSTM和ConvLSTM网络后，各类样本的混叠现象得到改善。通过进一步加入空间注意力模块和时间注意力模块来强调成像图序列帧内的空间特征和帧间的时间特征，得到ST-ConvLSTM网络。它的可视化结果的聚类效果得到进一步提高，各类样本之间出现较为

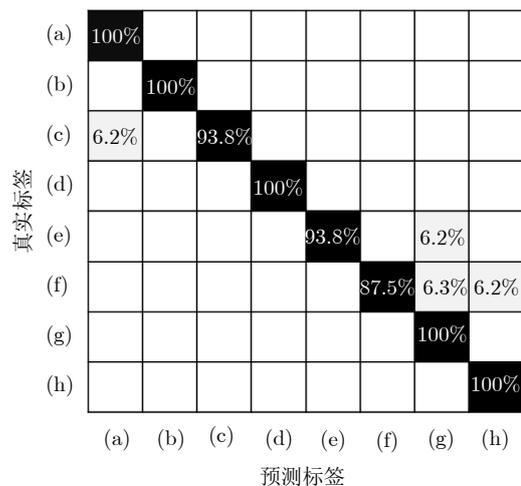


图 15 基于成像图序列特征的ST-ConvLSTM网络切向人体姿态识别结果。(a)~(h)分别对应表2中的8种切向人体姿态

Fig. 15 Recognition results of tangential human postures by ST-ConvLSTM network based on imaging sequence. (a)~(h) indicate tangential human activities from Tab. 2, respectively.

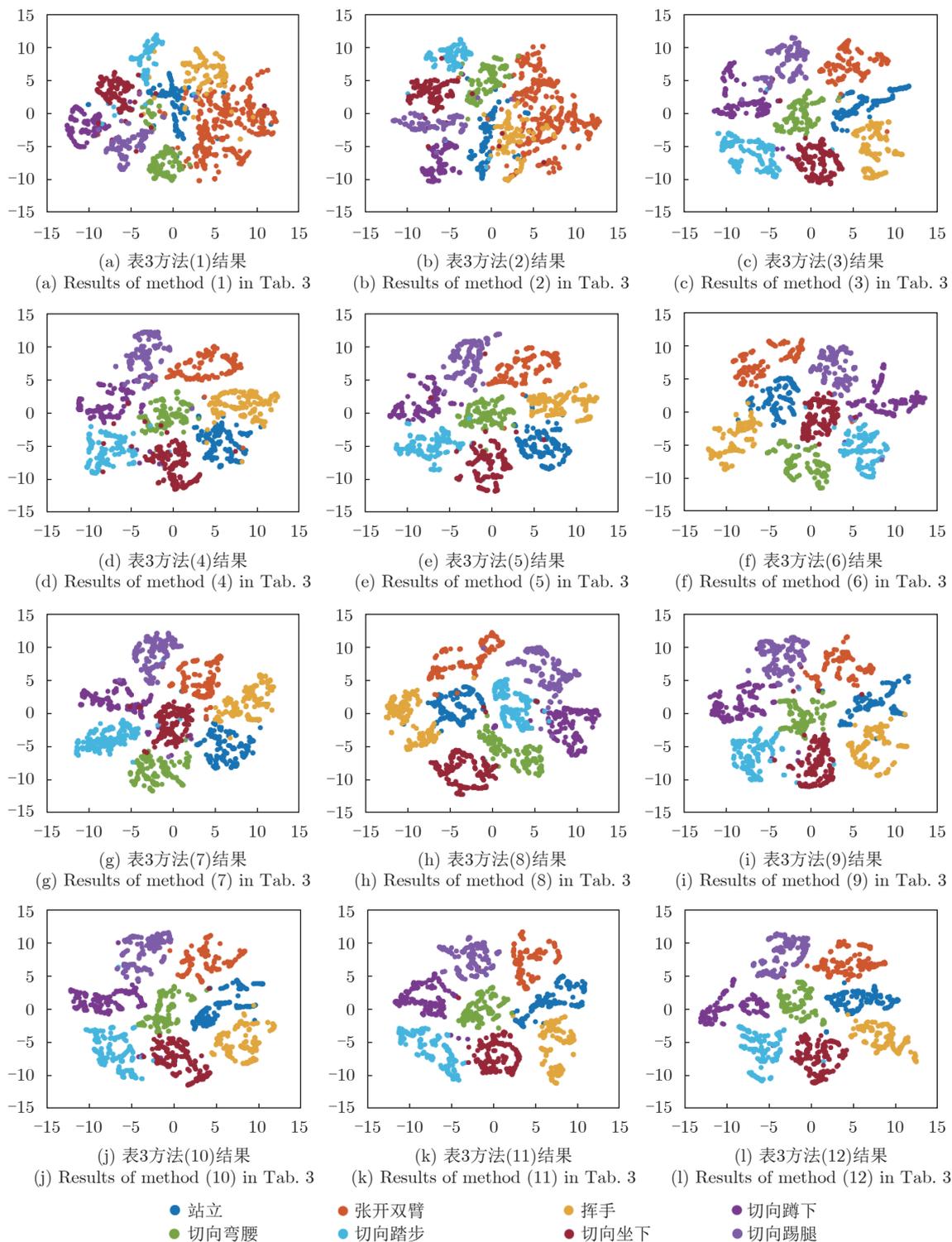


图 16 切向人体姿态识别方法网络特征的t-SNE二维可视化结果

Fig. 16 Visualization results of tangential human postures recognition methods

显著的分类界线。而随着去噪算法的优化,可视化结果中离散的异常点越来越少,这也与表3中对应识别准确率的提高相吻合。

4.3 个体差异鲁棒性

利用已知个体的数据训练出的网络模型来处理

并识别未知个体的数据一直是识别方法走向实际应用不可避免的一项挑战。本文采用留一法探究所提方法在面对个体差异时的鲁棒性能。具体来说,首先,从8个志愿者的数据中随机挑出7个个体的数据作为训练集来训练出网络识别模型,剩下的1个个体的数据作为测试集来得到测试结果。通过遍历

8个志愿者分别得到各个个体数据的识别准确率，并求出其平均值作为面对个体差异时方法鲁棒性的衡量标准。如表4所示，本文所提方法面对个体差异时，得到的平均准确率为93.75%，标准差为2.50%。特别地，可以观察到每一次留一法的识别准确率均高于90%，本文的8个志愿者的个体差异对所提方法的识别性能影响不大。这些结果也验证了本文所提方法在面对个体差异时的鲁棒性。

5 结语

针对传统多普勒方法在基于雷达传感器的切向人体姿态识别性能下降的问题，本文提出了一种基于MIMO雷达成像图序列的切向人体姿态识别方法，以高质量成像图序列的形式来表征切向姿态的人体轮廓结构及其动态变化，通过提取图像内的空间特征和图序列间的时序特征，实现对切向人体姿态的准确识别。首先，通过恒虚警检测算法定位人体目标所在距离门，利用慢时滑窗将目标动作按时序进行划分，对每帧数据用二维Capon算法估计出人体姿态的俯仰角度和方位角度，并结合距离信息得到人体姿态成像图，将各帧的成像图按照时序串联起来，构成人体姿态成像图序列。然后，通过改进的多域联合自适应阈值去噪算法，在去除掉背景环境噪声的同时，增强人体轮廓细节和结构信息，改善成像图序列的质量。最后，通过搭建基于空时注意力的卷积长短期记忆网络ST-ConvLSTM，在学习成像图序列多维特征的同时，强调图像内部的空间特性和图序列间的时序特性。实验结果表明，相比于传统方法，本文所提出的方法能够在保留人体轮廓和结构的同时，有效去除干扰噪声，并在8种典型的切向人体姿态的识别中取得了最高96.9%的准确率。在t-SNE的二维可视化结果对比中，所提方法的聚类效果显著，各类切向姿态间的分布边界清晰。通过留一法还研究了所提方法在面对个体差异时的鲁棒性。在后续的工作中，我们将继续扩

充数据集，容纳更多更复杂的切向人体姿态。此外，还将继续探索更加复杂场景下的雷达图像处理和优化方法。

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参考文献

- [1] 金添, 宋永坤, 戴永鹏, 等. UWB-HA4D-1.0: 超宽带雷达人体动作四维成像数据集[J]. 雷达学报, 2022, 11(1): 27–39. doi: [10.12000/JR22008](https://doi.org/10.12000/JR22008).
JIN Tian, SONG Yongkun, DAI Yongpeng, et al. UWB-HA4D-1.0: An ultra-wideband radar human activity 4D imaging dataset[J]. *Journal of Radars*, 2022, 11(1): 27–39. doi: [10.12000/JR22008](https://doi.org/10.12000/JR22008).
- [2] LE KERNEC J, FIORANELLI F, DING Chuanwei, et al. Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms[J]. *IEEE Signal Processing Magazine*, 2019, 36(4): 29–41. doi: [10.1109/MSP.2019.2903715](https://doi.org/10.1109/MSP.2019.2903715).
- [3] QI Fugui, LV Hao, WANG Jianqi, et al. Quantitative evaluation of channel micro-Doppler capacity for MIMO UWB radar human activity signals based on time-frequency signatures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(9): 6138–6151. doi: [10.1109/TGRS.2020.2974749](https://doi.org/10.1109/TGRS.2020.2974749).
- [4] TANG Longzhen, GUO Shisheng, JIAN Qiang, et al. Through-wall human activity recognition with complex-valued range-time-Doppler feature and region-vectorization ConvGRU[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5111014. doi: [10.1109/TGRS.2023.3329561](https://doi.org/10.1109/TGRS.2023.3329561).
- [5] LI Zhi, JIN Tian, LI Lianlin, et al. Spatiotemporal processing for remote sensing of trapped victims using 4-D imaging radar[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5103412. doi: [10.1109/TGRS.2023.3266039](https://doi.org/10.1109/TGRS.2023.3266039).
- [6] YANG Shufan, LE KERNEC J, ROMAIN O, et al. The human activity radar challenge: Benchmarking based on the ‘radar signatures of human activities’ dataset from Glasgow university[J]. *IEEE Journal of Biomedical and Health Informatics*, 2023, 27(4): 1813–1824. doi: [10.1109/JBHI.2023.3240895](https://doi.org/10.1109/JBHI.2023.3240895).
- [7] BAI Xueru, HUI Ye, WANG Li, et al. Radar-based human gait recognition using dual-channel deep convolutional neural network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(12): 9767–9778. doi: [10.1109/TGRS.2019.2929096](https://doi.org/10.1109/TGRS.2019.2929096).

表 4 面对个体差异基于留一法的所提算法鲁棒性结果

Tab. 4 Robustness performance in individual diversity study

志愿者	准确率(留一法)
1	93.8%
2	91.3%
3	95%
4	97.5%
5	96.2%
6	93.8%
7	90%
8	92.5%

- [8] LI Changzhi, PENG Zhengyu, HUANG T Y, *et al.* A review on recent progress of portable short-range noncontact microwave radar systems[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2017, 65(5): 1692–1706. doi: [10.1109/TMTT.2017.2650911](https://doi.org/10.1109/TMTT.2017.2650911).
- [9] KIM Y and LING Hao. Human activity classification based on micro-Doppler signatures using a support vector machine[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(5): 1328–1337. doi: [10.1109/TGRS.2009.2012849](https://doi.org/10.1109/TGRS.2009.2012849).
- [10] DING Chuanwei, HONG Hong, ZOU Yu, *et al.* Continuous human motion recognition with a dynamic range-Doppler trajectory method based on FMCW radar[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 6821–6831. doi: [10.1109/TGRS.2019.2908758](https://doi.org/10.1109/TGRS.2019.2908758).
- [11] EROL B and AMIN M G. Radar data cube processing for human activity recognition using multisubspace learning[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2019, 55(6): 3617–3628. doi: [10.1109/TAES.2019.2910980](https://doi.org/10.1109/TAES.2019.2910980).
- [12] DING Chuanwei, ZHANG Li, CHEN Haoyu, *et al.* Sparsity-Based Human Activity Recognition With PointNet Using a Portable FMCW Radar. *IEEE Internet of Things Journal*, 2023, 10(11): 10024–10037. doi: [10.1109/JIOT.2023.3235808](https://doi.org/10.1109/JIOT.2023.3235808).
- [13] LI Xinyu, HE Yuan, FIORANELLI F, *et al.* Semisupervised human activity recognition with radar micro-Doppler signatures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5103112. doi: [10.1109/TGRS.2021.3090106](https://doi.org/10.1109/TGRS.2021.3090106).
- [14] KIM W Y and SEO D H. Radar-based human activity recognition combining range-time-Doppler maps and range-distributed-convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1002311. doi: [10.1109/TGRS.2022.3162833](https://doi.org/10.1109/TGRS.2022.3162833).
- [15] QIAO Xingshuai, AMIN M G, SHAN Tao, *et al.* Human activity classification based on micro-Doppler signatures separation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5105014. doi: [10.1109/TGRS.2021.3105124](https://doi.org/10.1109/TGRS.2021.3105124).
- [16] LUO Fei, BODANESE E, KHAN S, *et al.* Spectro-temporal modeling for human activity recognition using a radar sensor network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5103913. doi: [10.1109/TGRS.2023.3270365](https://doi.org/10.1109/TGRS.2023.3270365).
- [17] DING Chuanwei, CHAE R, WANG Jing, *et al.* Inattentive driving behavior detection based on portable FMCW radar[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2019, 67(10): 4031–4041. doi: [10.1109/TMTT.2019.2934413](https://doi.org/10.1109/TMTT.2019.2934413).
- [18] WANG Bo, ZHANG Hao, and GUO Yongxin. Radar-based soft fall detection using pattern contour vector[J]. *IEEE Internet of Things Journal*, 2023, 10(3): 2519–2527. doi: [10.1109/JIOT.2022.3213693](https://doi.org/10.1109/JIOT.2022.3213693).
- [19] 元志安, 周笑宇, 刘心溥, 等. 基于RDSNet的毫米波雷达人体跌倒检测方法[J]. *雷达学报*, 2021, 10(4): 656–664. doi: [10.12000/JR21015](https://doi.org/10.12000/JR21015).
- YUAN Zhi'an, ZHOU Xiaoyu, LIU Xinpu, *et al.* Human fall detection method using millimeter-wave radar based on RDSNet[J]. *Journal of Radars*, 2021, 10(4): 656–664. doi: [10.12000/JR21015](https://doi.org/10.12000/JR21015).
- [20] TAHMOUSH D and SILVIOUS J. Radar micro-Doppler for long range front-view gait recognition[C]. 3rd IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), Washington, USA, 2009: 1–6. doi: [10.1109/BTAS.2009.5339049](https://doi.org/10.1109/BTAS.2009.5339049).
- [21] ALNUJAIM I, RAM S S, OH D, *et al.* Synthesis of micro-Doppler signatures of human activities from different aspect angles using generative adversarial networks[J]. *IEEE Access*, 2021, 9: 46422–46429. doi: [10.1109/ACCESS.2021.3068075](https://doi.org/10.1109/ACCESS.2021.3068075).
- [22] QI Fugui, LI Zhao, MA Yangyang, *et al.* Generalization of channel micro-Doppler capacity evaluation for improved finer-grained human activity classification using MIMO UWB radar[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2021, 69(11): 4748–4761. doi: [10.1109/TMTT.2021.3076055](https://doi.org/10.1109/TMTT.2021.3076055).
- [23] FIORANELLI F, PATEL J, GÜRBÜZ S Z, *et al.* Multistatic human micro-Doppler classification with degraded/jammed radar data[C]. 2019 IEEE Radar Conference (RadarConf), Boston, USA, 2019: 1–6. doi: [10.1109/RADAR.2019.8835618](https://doi.org/10.1109/RADAR.2019.8835618).
- [24] THIEL M and SARABANDI K. Ultrawideband multi-static scattering analysis of human movement within buildings for the purpose of stand-off detection and localization[J]. *IEEE Transactions on Antennas and Propagation*, 2011, 59(4): 1261–1268. doi: [10.1109/TAP.2011.2109349](https://doi.org/10.1109/TAP.2011.2109349).
- [25] FIORANELLI F, RITCHIE M, and GRIFFITHS H. Aspect angle dependence and multistatic data fusion for micro-Doppler classification of armed/unarmed personnel[J]. *IET Radar, Sonar & Navigation*, 2015, 9(9): 1231–1239. doi: [10.1049/iet-rsn.2015.0058](https://doi.org/10.1049/iet-rsn.2015.0058).
- [26] FAIRCHILD D P and NARAYANAN R M. Multistatic micro-Doppler radar for determining target orientation and activity classification[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2016, 52(1): 512–521. doi: [10.1109/TAES.2015.130595](https://doi.org/10.1109/TAES.2015.130595).
- [27] FIORANELLI F, RITCHIE M, GÜRBÜZ S Z, *et al.* Feature diversity for optimized human micro-Doppler classification using multistatic radar[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2017, 53(2): 640–654. doi: [10.1109/TAES.2017.2651678](https://doi.org/10.1109/TAES.2017.2651678).
- [28] QIAO Xingshuai, LI Gang, SHAN Tao, *et al.* Human activity classification based on moving orientation determining using multistatic micro-Doppler radar signals[J]. *IEEE Transactions on Geoscience and Remote*

- Sensing*, 2022, 60: 5104415. doi: [10.1109/TGRS.2021.3100482](https://doi.org/10.1109/TGRS.2021.3100482).
- [29] YU J T, YEN Li, and TSENG P H. mmWave radar-based hand gesture recognition using range-angle image[C]. 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 2020: 1–5. doi: [10.1109/VTC2020-Spring48590.2020.9128573](https://doi.org/10.1109/VTC2020-Spring48590.2020.9128573).
- [30] ZHAO Yubin, YAROVY A, and FIORANELLI F. Angle-insensitive human motion and posture recognition based on 4D imaging radar and deep learning classifiers[J]. *IEEE Sensors Journal*, 2022, 22(12): 12173–12182. doi: [10.1109/JSEN.2022.3175618](https://doi.org/10.1109/JSEN.2022.3175618).
- [31] YANG Yang, HOU Chunping, LANG Yue, *et al.* Omnidirectional motion classification with monostatic radar system using micro-Doppler signatures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(5): 3574–3587. doi: [10.1109/TGRS.2019.2958178](https://doi.org/10.1109/TGRS.2019.2958178).
- [32] YANG Yang, ZHANG Yutong, SONG Chunying, *et al.* Omnidirectional spectrogram generation for radar-based omnidirectional human activity recognition[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5105513. doi: [10.1109/TGRS.2023.3278409](https://doi.org/10.1109/TGRS.2023.3278409).
- [33] YANG Yang, LI Junhan, LI Beichen, *et al.* Few-shot omnidirectional human motion recognition using monostatic radar system[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 2531414. doi: [10.1109/TIM.2023.3328079](https://doi.org/10.1109/TIM.2023.3328079).
- [34] 张哲熙. 基于BM3D的图像去噪算法研究[D]. [硕士论文], 西安电子科技大学, 2017.
ZHANG Zhexi. Research on image denoising algorithm based on BM3D[D]. [Master dissertation], Xidian University, 2017.
- [35] DABOV K, FOI A, KATKOVNIK V, *et al.* Image denoising with block-matching and 3D filtering[C]. *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, San Jose, USA, 2006: 606414. doi: [10.1117/12.643267](https://doi.org/10.1117/12.643267).
- [36] WANG Xinheng, ISTEPANIAN R S H, and SONG Yonghua. Microarray image enhancement by denoising using stationary wavelet transform[J]. *IEEE Transactions on Nanobioscience*, 2003, 2(4): 184–189. doi: [10.1109/TNB.2003.816225](https://doi.org/10.1109/TNB.2003.816225).
- [37] 王安义, 战金龙, 卢建军. 一种新的二维Capon算法的研究[J]. 西安科技学院学报, 2003, 23(4): 437–440. doi: [10.3969/j.issn.1672-9315.2003.04.023](https://doi.org/10.3969/j.issn.1672-9315.2003.04.023).
- WANG Anyi, ZHAN Jinlong, and LU Jianjun. A new 2-D Capon algorithm[J]. *Journal of Xi'an University of Science and Technology*, 2003, 23(4): 437–440. doi: [10.3969/j.issn.1672-9315.2003.04.023](https://doi.org/10.3969/j.issn.1672-9315.2003.04.023).
- [38] RAHMAN S A and ADJEROH D A. Deep learning using convolutional LSTM estimates biological age from physical activity[J]. *Scientific Reports*, 2019, 9(1): 11425. doi: [10.1038/s41598-019-46850-0](https://doi.org/10.1038/s41598-019-46850-0).
- [39] SUDHAKARAN S and LANZ O. Convolutional long short-term memory networks for recognizing first person interactions[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 2017: 2339–2346. doi: [10.1109/ICCVW.2017.276](https://doi.org/10.1109/ICCVW.2017.276).
- [40] KIM Y and MOON T. Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks[J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(1): 8–12. doi: [10.1109/LGRS.2015.2491329](https://doi.org/10.1109/LGRS.2015.2491329).
- [41] LI Wenxuan, ZHANG Dongheng, LI Yadong, *et al.* Real-time fall detection using mmWave radar[C]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 2022: 16–20. doi: [10.1109/ICASSP43922.2022.9747153](https://doi.org/10.1109/ICASSP43922.2022.9747153).
- [42] KIM Y, ALNUJAIM I, and OH D. Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks[J]. *IEEE Sensors Journal*, 2021, 21(12): 13522–13529. doi: [10.1109/JSEN.2021.3068388](https://doi.org/10.1109/JSEN.2021.3068388).

作者简介

丁传威, 博士, 副教授, 主要研究方向为生物医学传感和雷达信号处理等。

刘芷麟, 硕士生, 主要研究方向为基于雷达传感器的人体动作识别。

张 力, 博士, 工程师, 主要研究方向为基于雷达信号处理的杂波抑制、非接触式生命体征探测以及人体动作识别和目标识别。

赵 恒, 博士, 讲师, 主要研究方向为生物医学传感、非接触式生命体征探测以及雷达信号处理等。

周 庆, 博士, 讲师, 主要研究方向为电离层电波传播、无线电海洋遥感以及高频雷达抗干扰等。

洪 弘, 博士, 教授, 主要研究方向为生物医学传感, 语音信号处理以及雷达信号处理等。

朱晓华, 博士, 教授, 主要研究方向为雷达系统, 雷达信号理论以及数字信号处理等。

(责任编辑: 于青)