UWB-HA4D-1.0: 超宽带雷达人体动作四维成像数据集

金　添*　　宋永坤　　戴永鹏　　胡锡坤　　宋勇平　　周小龙　　邱志峰

(国防科技大学电子科学学院　长沙　410073)

摘　要：雷达人体行为感知系统具有穿透探测能力，在安防、救援、医疗等领域具有广泛的应用前景。近年来，深度学习技术的出现促进了雷达传感器在人体行为感知领域的发展，同时对相关数据集的样本规模和丰富性提出了更高的要求。该文公开了一个超宽带雷达人体动作四维成像数据集，该数据集以超宽带多输入多输出雷达为探测传感器来获取了人体目标的距离-方位-高度-时间四维动作数据，共采集了11个人体目标的2757组动作数据，动作类型包含走路、挥手、打拳等10种常见动作，有穿透探测和不穿透探测的实验场景。该文详细介绍了数据集的系统参数、制作流程、数据分布等信息。同时，基于飞桨平台使用计算机视觉领域应用较多的深度学习算法对该数据集进行人体动作识别实验，实验对比结果可以作为参考，为学者使用该数据集提供技术支撑，方便在此基础上进一步探索研究。

关键词：超宽带雷达；四维成像数据集；人体动作识别；深度学习；飞桨平台

中图分类号：TN957　　**文献标识码**：A　　**文章编号**：2095-283X(2022)01-0027-13

DOI: 10.12000/JR22008

引用格式：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.

Reference format: 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.

UWB-HA4D-1.0: An Ultra-wideband Radar Human Activity 4D Imaging Dataset

JIN Tian*　　SONG Yongkun　　DAI Yongpeng　　HU Xikun　　SONG Yongping
ZHOU Xiaolong　　QIU Zhifeng

(*College of Electronic Science and Technology, National University of Defense Technology, Changsha* 410073, *China*)

Abstract: A radar human behavior perception system has penetration detection ability, which gives it a wide application prospect in the fields of security, rescue, medical treatment, and so on. Although the development of deep learning technology has promoted radar sensor research in human behavior perception, it requires more prompted dataset availability. This paper provides a four-dimensional imaging dataset of human activity using ultra-wideband radar, UWB-HA4D, which uses three-dimensional ultra-wideband multiple-input multiple-output radar as the detection sensor to capture the range-azimuth-height-time four-dimensional activity data of a human target. The dataset contains the activity data of 2757 groups for 11 human targets, including 10 common activities, such as walking, waving, and boxing. It also contains penetration and nonpenetration detection experimental scenarios. The radar system parameters, data generation process, data distribution, and other information of the dataset are introduced in detail herein. Meanwhile, several deep learning algorithms that are based on the PaddlePaddle framework and are widely used in the computer version field are applied to this dataset for human activity recognition. The experimental comparison results can be used to provide

收稿日期：2022-01-09；改回日期：2022-02-16；网络出版：2022-02-24
*通信作者：金添 tianjin@nudt.edu.cn　　*Corresponding Author: JIN Tian, tianjin@nudt.edu.cn
基金项目：国家自然科学基金(61971430)
Foundation Item: The National Natural Science Foundation of China (61971430)
责任主编：李廉林　Corresponding Editor: LI Lianlin

references for scholars and facilitate further investigation and research on this basis.

Key words: Ultra-wideband radar; Four-dimensional imaging dataset; Human activity recognition; Deep learning; PaddlePaddle framework

1 引言

人体动作识别技术应用于健康监护、运动分析、智能家居、场景监控等诸多领域，是计算机视觉领域的研究热点[1]。所谓动作识别，即通过对视频或图像序列进行处理分析，构建视频与人体动作之间的映射关系，使计算机能够像人一样去理解视频[2]。随着深度学习的出现和传感器技术的发展，大量的人体动作数据可以使用低廉的摄像头获取，为深度学习的训练提供了数据支撑，以可见光、结构光为探测手段的深度学习动作识别研究取得了较大成效[3]，逐渐取代传统算法。目前，在计算机视觉领域，已有众多基于光学传感器的人体动作识别数据集，如KTH[4], UCF-101[5], HMDB[6], Kinetics[7], NTU RGB+D[8]等，这些数据集中样本的数目由几千到几十万不等，为研究者提供了便利的研究条件，进一步推动了动作识别技术的应用落地。

随着时代的进步，人们对动作识别系统的应用场景和隐私保护等方面提出了更高的要求。传统的光学传感器无法在无光、遮蔽和非视距等复杂环境下工作且存在隐私泄露的风险，无法适应多样性的应用需求。然而，以电磁波信号为信息传递载体的雷达是一种主动探测系统，得益于电磁波的穿透性，雷达系统可以在遮挡环境下稳定工作。同时，雷达系统通过分析人体目标反射回波提取人体动作信息，不直接获取人体面部信息，具有较好的隐私保护性能[9]。因此，相比于被动探测的光学传感器，基于雷达传感器的人体动作识别系统有更广泛的使用场景，具有较大的研究价值。

利用雷达进行人体动作识别的一般步骤是，首先选择合适的人体动作特征，然后从接收的雷达回波中提取有效特征，最后采用合适的分类器实现动作识别。早期的研究大多使用手工提取特征的方法，采用一定的雷达信号处理方法，首先提取雷达信号的幅度、频率、相位等信号波形特性[10]，目标的距离、方位、高度等目标特性，以及微多普勒运动等信息作为特征，再使用支持向量机、贝叶斯分类、决策树等传统的机器学习方法进行分类[11]。然而，此类方法在实现过程中涉及多项操作，需要大量人工干预，且动作识别的精度不高，应用受限。近年来，深度学习技术在光学动作识别领域表现良好，实现了特征设计、提取和分类识别的一体化的设计，动作识别精度均优于传统算法，一些学者开始将此技术应用于雷达人体动作识别的研究中[12]。加州大学的Kim等人[13]最早将雷达回波信号处理得到的微多普勒谱输入3层卷积神经网络中，取得了较好的动作识别准确率。美国约翰霍普金斯大学Craley等人[14]引入了长短时记忆网络(LSTM)对雷达多普勒图像进行分类识别，充分利用了人体动作特征中的时序信息。电子科技大学的Wang等人[15]设计了一种多个LSTM堆叠的网络结构，并在实测数据上进行了验证，动作识别精度优于深度卷积神经网络(DCNN)。北京邮电大学Li等人[16]采用迁移学习方法来解决小样本雷达数据集下的人体行为识别问题，提高了网络对于新场景下人体行为识别的泛化能力。国防科技大学的Du等人[17]提出了基于距离-多普勒-时间的三维点云输入模式，距离信息的引入提高了不同位置肢体的差异，进一步改善了动作识别率。北京大学李廉林等人[18]基于智能电磁感知技术使用低成本的超材料天线实现了个体和肢体动作的智能识别，拓宽了人体感知的研究方向[19]。

目前基于雷达传感器的人体动作识别数据集相对稀缺，在一定程度上制约了深度学习在雷达人体动作识别领域的发展。尽管部分学者公布了一些雷达动作识别领域的数据集[20–22]，但这些数据集大都基于单通道或者较少通道的雷达回波信号的多普勒特征进行动作识别，而多普勒特征包含的信息量远不如光学传感器提供的人体姿态图像特征丰富，使得在人体动作识别领域中雷达传感器的实用化远远滞后于光学传感器。近年来，低频超宽带多输入多输出(Multiple-Input Multiple-Output, MIMO)雷达技术逐渐成熟，在实现较好穿透性的同时，具有距离、方位和高度三维信息感知能力，能够获取与光学传感器类似的人体姿态图像序列[23]。另外，相比于单通道或少通道雷达系统，MIMO雷达具有更好的空间分辨能力，可实现多目标的探测和分离，有较大的研究价值。然而，由于缺少相关公开数据集，严重限制了MIMO雷达传感器在人体动作识别领域的实用化进程。

为了促进雷达人体动作识别研究的发展，丰富雷达数据集的多样性，本文基于低频超宽带MIMO雷达，构建了超宽带雷达人体动作四维成像数据集(Ultra-Wideband radar Human Activity 4D imaging dataset, UWB-HA4D-1.0)。不同于传统的基于微多普勒谱进行动作识别的数据集，该数据集是

国际首个基于雷达四维成像的人体动作数据集，开辟了人体动作识别领域研究的新路线。本数据集包含人体目标的距离-方位-高度-时间四维信息，共采集了11个不同体型人体目标的10种不同动作，以及3种不同场景的雷达数据。该数据集已可通过《雷达学报》官网的相关链接(https://radars.ac.cn/web/data/getData?dataType=UWB-HA4D)免费下载使用。另外，本文以PaddlePaddle为网络框架，使用了计算机视觉领域几种常用的动作识别深度学习网络对数据集进行训练和验证，为该数据集的使用和开发提供参考，方便其他学者进行更进一步的探索研究。

2　UWB-HA4D-1.0数据集信息

2.1 超宽带MIMO雷达系统

结合雷达人体动作四维数据采集的任务需求，本节对所需雷达系统的参数进行讨论分析。关于雷达系统的工作频段的选择，已知工作于0～3 GHz频段的低频雷达有较好的穿透性，可穿透多种墙体介质对遮挡目标进行探测[24]，适用于多种探测场景。而超宽带雷达相比于窄带雷达具有更优的距离分辨率，可获取目标高精度距离信息[25]。对于雷达系统的信号体制而言，常见的有窄脉冲信号、线性调频信号、步进频信号等，相比于其他两种信号，步进频信号具有高发射功率，频带拓展性好、大时宽、大带宽的特性，有效克服了窄脉冲信号平均功率较低的缺陷，广泛应用于超宽带雷达领域[26]。关于雷达系统的阵列构型，按照天线的排布可分为一维雷达、二维雷达和三维雷达[27]，其中一维雷达采用单发单收的天线形式，仅具有距离分辨能力；二维MIMO雷达的天线采用一维线阵排布，可提供目标的距离、方位二维信息；三维MIMO超宽带雷达的天线采用二维面阵排布，可获取目标的距离、方位、高度三维信息，对人体的肢体轮廓进行描绘。相比于一维和二维雷达，三维雷达可提供更详细的人体目标信息，对判定人体目标的行为状态提供了良好的信息支撑。因此，本文采用二维MIMO雷达阵列发射低频超宽带步进频信号的雷达系统技术方案。

综合考虑多种因素，本文设计了一款三维超宽带MIMO雷达系统，系统参数指标如表1所示。为了获取更好的方位和高度向分辨率，保证雷达成像质量，系统采用了10发10收的大规模MIMO阵列排布形式，阵列尺寸为60 cm×88 cm。由于采用了1.78～2.78 GHz的低频电磁波信号，雷达系统具有较好的穿透性，可穿透幕布、木板、塑料、泡沫、砖墙等常见遮挡物进行目标探测。另外，本系统的信号发射功率仅为20 dBm，不会对人体造成伤害。

本文所设计的三维超宽带MIMO雷达系统样机如图1所示，其中雷达系统进行信号的发射和接收，计算机负责数据的存储和实时处理。二维MIMO阵列的等效图如图2(a)所示，左右两侧的阵列为发射天线，上下两行的阵列为接收天线，10发10收的阵列等效为100个虚拟阵元，本阵列设计将发射天线在高度维不规则排布来降低旁瓣水平。二维MIMO阵列实物图如图2(b)所示，其中天线阵元结构为宽带蝶形阵子天线。

2.2 数据采集与处理

雷达回波信号的采集和处理流程如图3所示，首先MIMO雷达发射电磁波信号，并接收人体目标反射回波，对接收到的100个通道的雷达回波进行动目标显示(Moving Target Indication, MTI)处理[28]，滤除静止杂波。然后进行大范围的方位-距离二维成像，成像区域为设定的系统探测范围，对二维成像结果进行恒虚警率(Constant False Alarm Rate, CFAR)检测[29]和跟踪处理来锁定目标在方位-距离二维平面的位置，最后对目标所在位置的方位向±1 m、距离向±1 m、高度向0～2.5 m范围进行

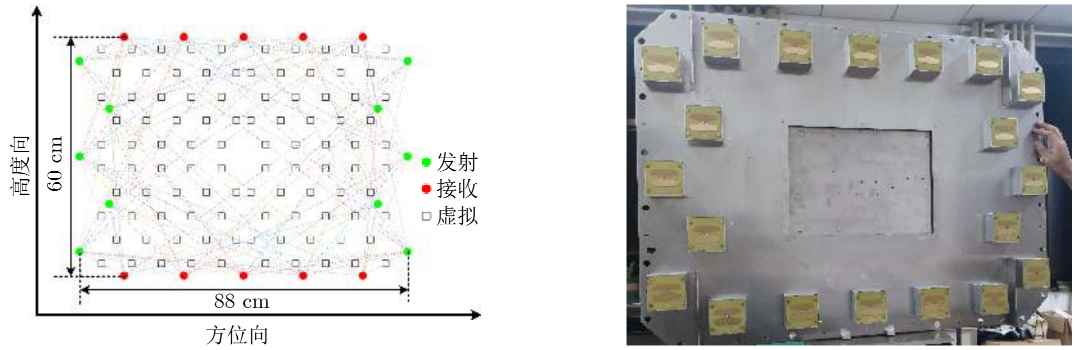
表 1 雷达系统参数

Tab. 1　Radar system parameters

参数	指标
工作频段	1.78～2.78 GHz
信号带宽	1 GHz
信号体制	步进频信号
信号步进带宽	4 MHz
脉冲重复频率	10 Hz
天线阵元数	10发10收(MIMO)
信号发射功率	20 dBm (100 mW)
系统尺寸	60 cm×88 cm
可穿透介质	幕布、木板、塑料、泡沫、砖墙等

图 1 三维超宽带MIMO雷达系统

Fig. 1　Three-dimensional UWB MIMO radar system



(a) 二维MIMO阵列等效图
(a) Equivalent diagram of two-dimensional MIMO array

(b) 二维MIMO阵列实物图
(b) Physical image of two-dimensional MIMO array

图 2 二维MIMO阵列
Fig. 2 Two-dimensional MIMO array

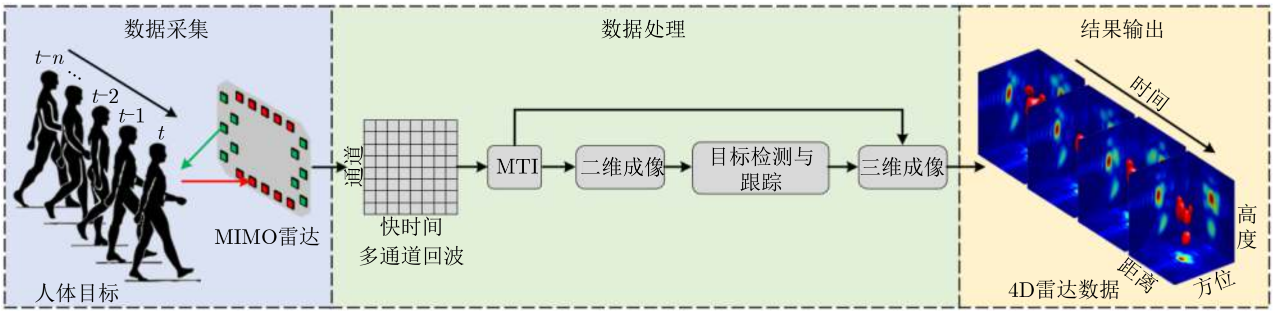


图 3 数据采集与处理流程
Fig. 3 Data collection and processing flow

三维成像(假定目标位于地平面)。最后，联合时间维度信息构成人体动作4D雷达数据。关于人体目标的成像，本文采取的先大范围二维成像再小范围三维成像的成像思路，不仅可以避免无目标区域三维成像造成的运算量浪费，节约成像时间，还可以保证三维成像结果有足够的成像网格密度，兼顾了成像效率和成像质量。

本文选择后向投影(Back Projection, BP)算法[30]作为MIMO雷达成像算法，该算法是一种典型的时域成像算法，对阵元排布没有特殊要求，广泛应用于各种MIMO阵列成像领域。BP算法的基本思想是对成像区域进行网格划分，计算在成像区域中的像素点到天线阵列的距离从而计算出传输的时间延迟，根据这个时间延迟来搜索天线阵列接收到的雷达回波信号，将每个通道的回波信号进行叠加计算[31]。虽然BP成像算法运算量稍大，但配合GPU强大的并行运算能力，BP算法成像耗时通常并不显著高于其他成像算法。因此，本数据集的数据生成阶段采用BP算法进行雷达成像。

2.3 数据采集场景

本数据集的采集场景有3个，分别是无遮挡场景S1, 3 cm塑料板遮挡场景S2，以及27 cm砖墙遮挡场景S3，具体信息及场景照片如表2和图4所示。本数据集中的训练集只是在S1场景下采集，而测试集包括了S1，S2，S3 3个场景的数据，其中场景S2和S3的数据可以用来测试所设计的动作识别方法的环境适应性。需要说明的是，本文在不同的场景都采用直接成像的方式获取四维雷达图像，未根据不同材料墙体遮挡对电磁波信号造成的影响进行补偿。

2.4 人体动作信息

本数据集所采集的10种动作类别示意图如图5所示，10种动作分别为开双臂、打拳、静坐、踢腿、坐下、站立、向前走、向左走、向右走、挥手。相比于使用运动微多普勒信息进行动作识别的数据集仅有运动动作，本数据集包含了目标人体各个身体部位的位置和运动双重信息，可以对运动和静止人体动作进行识别。因此，本数据集中不仅有运动动作还有像静坐、站立这样的静止动作，以及向左走、向右走这类易混淆动作，提供了更加丰富的动作类型。各个动作的组数如表3所示，每个动作的组数在269～278组，其中训练集组数都在149～158组，3个场景的测试集中每个动作为40组，共120组。共采集2757组动作，每组动作40帧三维雷达数据，共110280帧数据。

(a) 无遮挡场景S1　(b) 塑料板遮挡场景S2　(c) 砖墙遮挡场景S3

(a) Unobstructed scene S1　(b) Scene with plastic board blocked S2　(c) Scene with brick wall blocked S3

图 4 数据集采集场景

Fig. 4　Dataset collection scenes

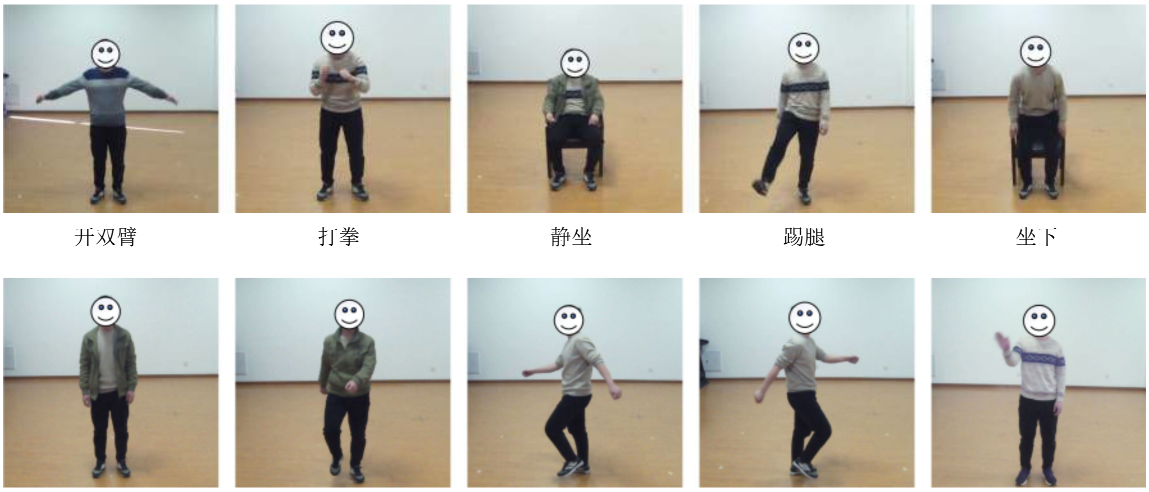
表 2 数据集采集场景信息

Tab. 2　Dataset collection scene information

场景编号	遮挡情况	训练集	测试集
S1	无遮挡	√	√
S2	3 cm塑料板遮挡	×	√
S3	27 cm砖墙遮挡	×	√

注：√表示有，×表示无。

取其中一组开双臂的雷达成像数据做方位-高度向最大值投影进行数据预览，结果如图6所示，分别是第10帧、第24帧、第32帧、第40帧的参考光学图像和雷达图像投影。由图6可知，本雷达系统的成像结果保留了人体目标的身体轮廓和躯干运动信息，可以用于不同动作的识别。



开双臂　打拳　静坐　踢腿　坐下

站立　向前走　向左走　向右走　挥手

图 5 动作类型

Fig. 5　Activity types

表 3 不同动作的数据量(组)

Tab. 3　The amount of data for different actions (groups)

标号	动作	S1场景训练	S1场景测试	S2场景测试	S3场景测试	总数
1	开双臂	149	40	40	40	269
2	打拳	155	40	40	40	275
3	静坐	156	40	40	40	276
4	踢腿	158	40	40	40	278
5	坐下	155	40	40	40	275
6	站立	156	40	40	40	276
7	向前走	157	40	40	40	277
8	向左走	156	40	40	40	276
9	向右走	158	40	40	40	278
10	挥手	157	40	40	40	277

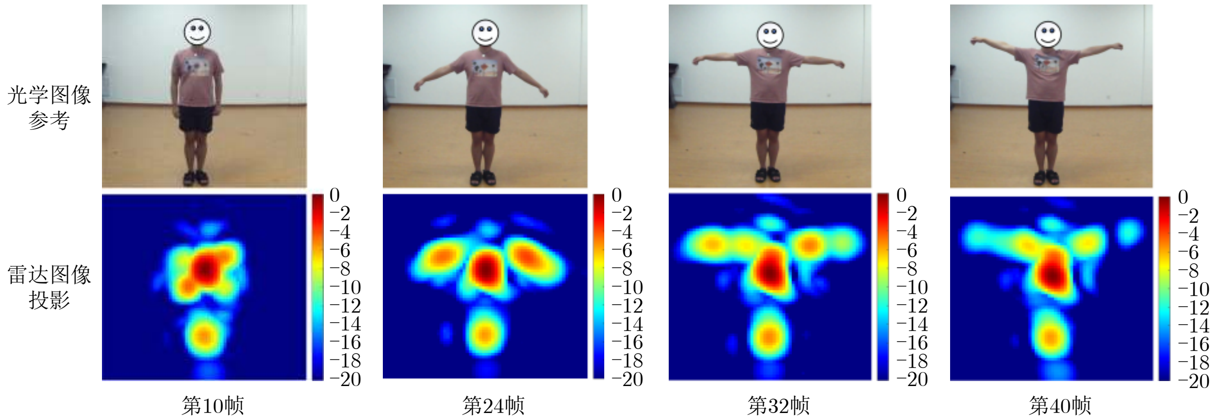


图 6 三维雷达图像投影

Fig. 6 Projection of three-dimensional images

2.5 人体目标信息

为了保证数据集中人体目标的多样性，本文采集了11个不同身高体重的人体目标，具体信息如表4所示，其中身高范围是163～186 cm，体重范围是53～85 kg。本文对人体目标进行编号，分别是H1—H11，根据人体目标的身高体重分布，选择不同身高体重段具有代表性的H6和H8为测试目标。另外，H1—H11只在S1场景录制训练集，而H6和H8两个人体目标在S1, S2, S3 3个场景分别录制测试集。

2.6 数据格式

本数据集对10个不同的动作分别标号是A1—A10，真值标号为0～9，具体如表5所示。数据以mat格式存储，每个文件的命名规则为“Am_Hn_Sp_q.mat”，其中Am为动作编号，m=1, 2, ···, 10，Hn为人体目标编号，n=1, 2, ···, 11，Sp为场景，p=1, 2, 3，q为组号。以“A2_H5_S1_9.mat”为例，该文件名即指在S1场景下H5目标的A2动作的第9组数据。每组数据中，存储雷达四维图像的矩阵名称为“radar_data_sequence”，数据的大小为40×64×64×64，其中40是三维雷达图像的帧数，64×64×64是三维雷达图像的大小，所有三维成像结果均为归一化后的幅度图。同时，本数据集提供了4个标注文件“train_label.txt”、“test_S1_label.txt”、“test_S2_label.txt”、“test_S3_label.txt”，分别对应1个训练集和3个测试集的标签。标签内容为“A2/A2_H5_S1_9.mat 3”，其中A2为动作类型文件夹名称，“A2_H5_S1_9.mat”为四维雷达数据名称，3为动作类型真值标号，与数据名以空格隔开。

表 4 人体目标信息

Tab. 4 Human target information

目标编号	身高(cm)	体重(kg)	S1场景	S2场景	S3场景
H1	175	70	√	×	×
H2	172	72	√	×	×
H3	178	68	√	×	×
H4	182	85	√	×	×
H5	170	75	√	×	×
H6	179	74	√	√	√
H7	165	60	√	×	×
H8	169	65	√	√	√
H9	162	53	√	×	×
H10	186	80	√	×	×
H11	171	67	√	×	×

表 5 人体动作标号

Tab. 5 Human activity labels

动作编号	动作类型	真值标号	动作编号	动作类型	真值标号
A1	开双臂	0	A6	站立	5
A2	打拳	1	A7	向前走	6
A3	静坐	2	A8	向左走	7
A4	踢腿	3	A9	向右走	8
A5	坐下	4	A10	挥手	9

3 动作识别方法

现有的动作识别方法可分为传统动作识别方法和基于深度学习的动作识别方法，其中传统动作识别方法包括基于模板匹配的方法[32]、基于时空兴趣点检测的方法[33]、基于关节点轨迹跟踪的方法[34]。对于基于深度学习方法的动作识别方法，根据网络卷积维度的不同可以分为基于二维卷积神经网络(Two Dimensional Convolutional Neural Networks, 2D CNN)的方法、基于三维卷积神经网络(Three Dimensional Convolutional Neural Networks, 3D CNN)的方法。由于基于深度学习的动作识别方法在实现复杂度和识别精度方面均有优异的表现，逐渐取代了传统动作识别算法，因此，本节只介绍基于深度学习的几种经典动作识别方法。

3.1 基于2D CNN的动作识别方法

基于2D CNN的动作识别方法研究主要有基于双流网络和基于时间特征提取模块的两个研究分支。Simonyan等人[35]最早提出了基于2D CNN的双流网络，该网络包含了两个相互独立的流，分别是空间流网络和时间流网络，其中空间流用于构建外观特征，时间流用来构建运动特征，最后将两个网络流的softmax结果融合，得到预测的动作类型。该网络存在的缺点是仅考虑了相邻帧之间的运动特征，对于长时间运动特征的提取具有一定的局限性。为了解决这个问题，Wang等人[36]在双流网络的基础上提出了一种时域分段网络(Time Segment Network, TSN)，该网络引入了稀疏采样的方法，将输入网络的视频分割成若干个视频片段，再进行时空特征提取，最后对各个片段的特征提取结果进行融合，得到预测结果。该方法具有全局时空特征的提取能力，有效解决了原始双流网络存在的长时间运动特征提取能力差的问题。但是，TSN的稀疏采样无法保证有效动作信息的提取。Lin等人[37]提出了一种时间移位模块(Temporal Shift Module, TSM)用来捕获时间域上的有效特征。该算法的核心思想是将部分信道沿时间维进行移位，便于相邻帧之间的信息交换，扩大了时间感受野。其优点是在完成有效时间信息建模的同时，几乎没有带来额外的计算量，并取得了较好的动作识别精度。

3.2 基于3D CNN的动作识别方法

基于光学传感器的视频序列中的人体动作是方位-高度-时间的三维数据，使用3D CNN可以直接获取人体动作在这3个维度上的特征。Ji等人[38]最早提出采用3D CNN视频时空特征提取架构，该架构从相邻帧数据中提取多通道的信息，分别进行卷积处理，最后综合各个通道的特征预测动作类型。Tran等人[39]在3D CNN的基础上提出了C3D (Convolutional 3D)框架，该框架最终获得了比2D CNN更加高效的特征提取。基于前期研究，Tran等人[40]将C3D架构与Resnet网络相结合，提出了新的Res3D网络，进一步提高了动作识别精度。为了提高动作识别网络对于时空特征变化的适应性，Feichtenhofer等人[41]提出了快慢网络(SlowFast Networks, SFN)，该网络包含慢帧率和快帧率两条通道，其中慢帧率通道用来提取动作的空间语义信息，快帧率通道用来提取精细的动作特征，该网络对变化快的动作的识别具有较大的优势。

4 实验结果及分析

本文选取计算机视觉领域的几种代表性动作识别方法在UWB-HA4D-1.0数据集上进行实验验证，主要有基于2D CNN的TSN, TSM，以及基于3D CNN的Res3D, SFN这几种算法。由于四维雷达图像在网络处理的过程中需要消耗大量的运算单元，所以本文采取将人体目标的三维成像结果做方位-距离、方位-高度、距离-高度3个二维平面上的最大值投影，来实现减少数据量的目的，即将大小为64×64×64的数据转化为大小为3×64×64的数据，以下实验均基于投影后的数据进行处理。

4.1 实验网络设计

本文设计的基于TSN网络结构的雷达图像人体动作识别网络如图7所示，首先将时间为T的一段四维雷达数据分割成N段，$S_1, S_2, \cdots, S_N$，分别从N段数据中取出一帧三维成像结果，进行3个平面

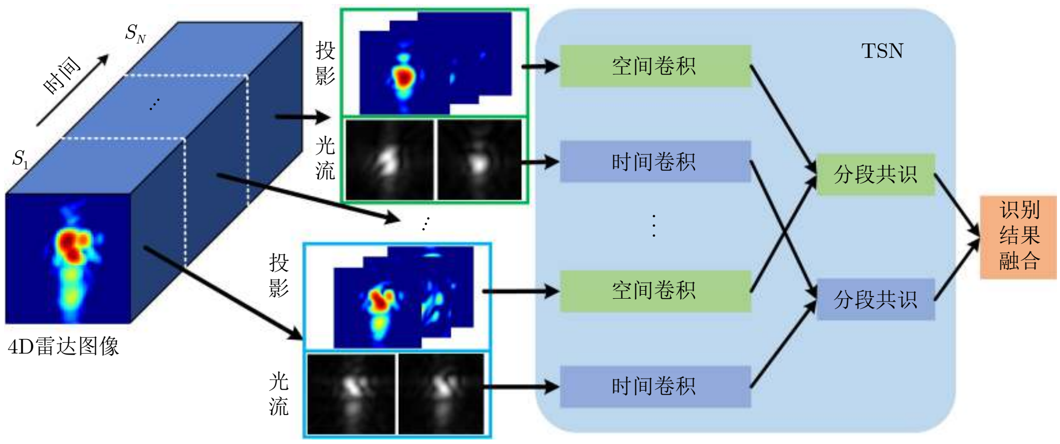


图 7 TSN结构图

Fig. 7 TSN structure

的最大值投影，并提取光流信息；然后分别使用二维空间卷积提取每一帧雷达图像中的空间特征，使用二维时间卷积提取光流图中的时间特征，最后再将时间和空间特征融合，得到最后的动作识别结果。

基于TSM结构的雷达图像人体动作识别网络的预处理与TSN网络一致，首先将四维雷达数据分割成N段，然后分别从每段中取出一帧数据，共N帧数据。然后按照图8的时间移动方法对N帧数据进行处理，将数据分别沿着时间维进行前移和后移，最后使用2D CNN同时提取时间和空间信息，得到动作识别结果。

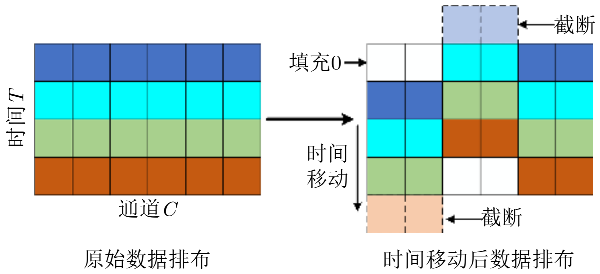


图 8 TSM网络核心结构

Fig. 8 The core structure of TSM network

基于Res3D网络的雷达图像人体动作识别网络的处理思路是直接使用3D CNN同时提取连续帧雷达图像中人体动作的时间和空间特征，网络结构图如图9所示，通过Resnet网络提取特征，最后使用全卷积层实现动作识别。

基于STN网络结构的雷达图像人体动作识别方法的实现框图如图10所示，由图10可知，慢帧率通道相比于快帧率通道的数据采样间隔较大，数据量较少，两个通道分别使用3D CNN提取特征。另外，快速通道的特征通过侧向连接与慢速通道相连，实现特征的融合，最后将两个通道的结果融合得到最终的动作预测结果。

4.2 实验设置与结果

本节使用在S1场景下采集的9人的10个动作作为训练集对4.1节设计的TSN, TSM, Res3D, STN 4个网络进行训练，将其他2人在S1, S2, S3 3个场景下采集的数据作为测试集。在训练的过程中，使用Momentum作为优化算法，设置训练循环周期数为100，网络的初始学习率为0.01，并分别在第25和第60个周期以十分之一递减，batch size设置为8，num_workers设置为4。几种算法的特征提取网络是Resnet网络，网络实现是使用百度公司

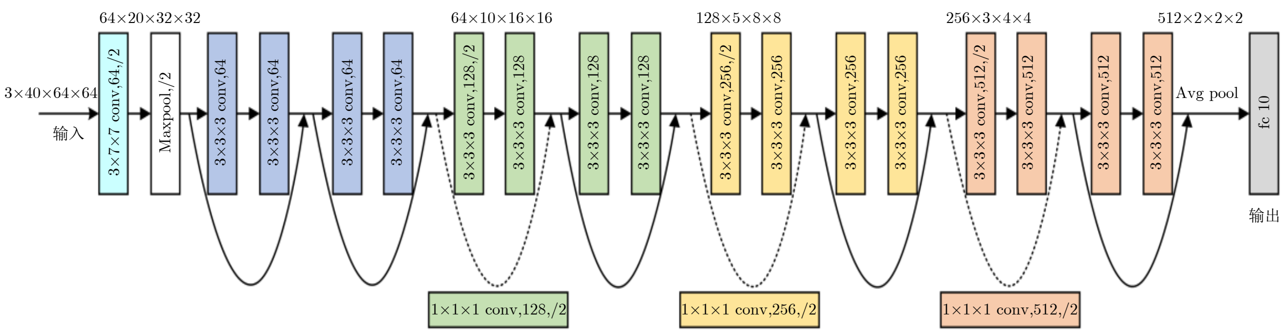


图 9 Res3D网络结构图

Fig. 9 Res3D network structure

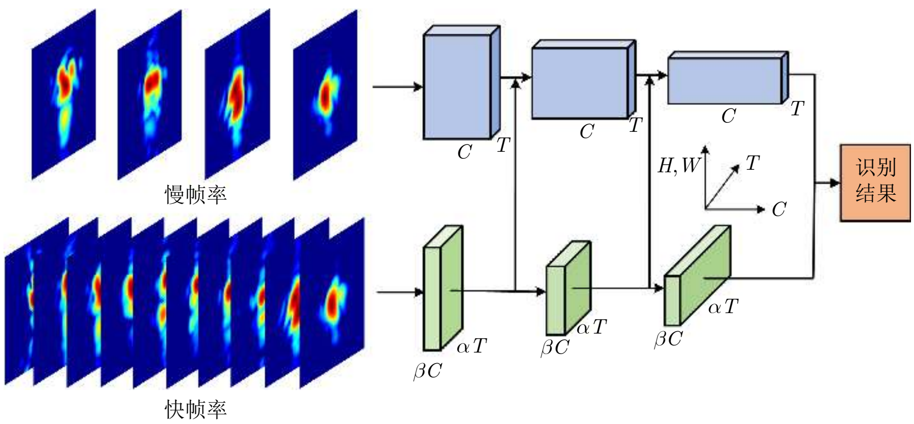


图 10 SFN结构图

Fig. 10 SFN structure

开发的飞桨平台PaddlePaddle框架。网络的训练和测试环境为Ubuntu 20系统，显卡为NVIDIA RXT2070，计算机运行内存为64GB。

对4种网络分别进行训练，并对3个不同场景的数据进行测试，得到结果如表6所示。由表可知，Res3D网络在S1测试集上取得了最佳识别精度，达到了92.25%，优于其他几种方法。对比S1, S2, S3 3个测试场景的动作识别精度可知，S2场景的识别精度与S1场景较为接近。而S3场景由于较厚墙体的遮挡，电磁波能量衰减较大，成像质量与无遮挡的S1场景相比下降较多，所以动作识别精度也大幅度降低。同时也可知，基于Res3D的方法在非同一探测场景下比其他网络取得了更好的识别精度，具有更好的环境适应性，而TSN方法的环境适应性最差。

本节以TSM网络为例展开分析，由4.1节可知，基于2D CNN方法的TSN和TSM网络需要对数据进行分段处理，然后从每段数据中取出一帧输入网络，本文对网络分段数与最终动作识别的精度之间的关系进行探究。以TSM网络为例，网络的输入数据的长度为40帧，为了减少非等间距采样对结果的影响，本文将输入数据分为可以被40整除的段数，即为4段、5段、8段、10段，以及20段。分别使用几种分段数对网络进行训练，并使用S1数据进行验证，得到100个训练周期对应的测试结果，具体如图11所示，其中图11(a)为不同分割段数的TSM网络测试在不同训练周期得到的动作识别测试精度，图11(b)是不同分割段数的最佳测试精度。由图11可知，随着分段数目的增多，网络的数据量增大，获取到更多的目标运动信息，所以得到了较好的动作识别精度。

为了直观了解不同动作的识别率，本节对网络在不同场景的测试结果进行对比分析，以基于3D CNN的Res3D网络为例，得到3个不同场景下的测试结果如表7所示。由表可知，S1和S2场景中的动作识别率较高，而S3场景由于电磁波能量减弱，对挥手、踢腿等轻微动作的识别率较低。同时，该网络对坐下和静坐两个动作的识别率都较高。

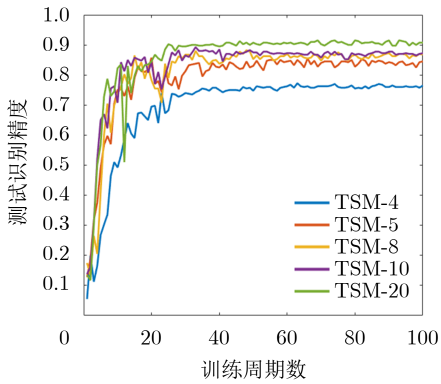
4.3 实验讨论

本节借鉴计算机视觉领域的动作识别算法实现了4种基于雷达四维成像数据集的动作识别网络，

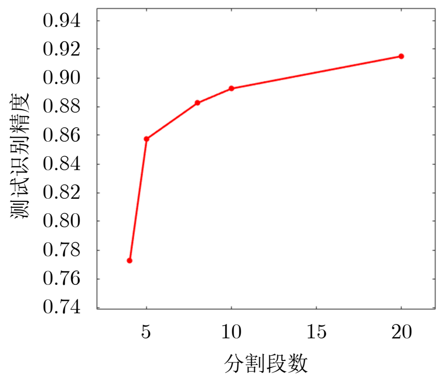
表 6 实验结果对比表

Tab. 6 Experimental results comparison table

识别方法	网络框架	S1识别精度	S2识别精度	S3识别精度
2D CNN	TSN	85.75%	83.5%	60.75%
	TSM	91.50%	88.0%	73.75%
3D CNN	SFN	88.00%	80.5%	70.25%
	Res3D	**92.25%**	**90.0%**	**77.00%**



(a) 不同分割段数的TSM网络测试曲线
(a) TSM network test curves with different segment numbers



(b) 不同段数TSM网络最优动作识别精度
(b) The best activity recognition accuracy of TSM network with different segment numbers

图 11 TSM网络测试结果

Fig. 11 TSM network test results

表 7 Res3D网络在不同场景下的动作识别精度(%)

Tab. 7 Human activity recognition accuracy of Res3D networks in different scenes (%)

探测场景	张开双臂	打拳	静坐	踢腿	坐下	站立	向前走	向左走	向右走	挥手	平均
S1场景	90	90.0	97.5	82.5	100	85.0	97.5	100	100	80	**92.25**
S2场景	85	92.5	100.0	85.0	100	82.5	85.0	100	100	70	90.00
S3场景	90	82.5	100.0	42.5	100	65.0	50.0	70	100	70	77.00

分别是TSN, TSM, Res3D和SFN，并对4种网络的算法框架进行了介绍。为了便于网络处理，本文对四维数据进行了3个方向的最大值投影，然后分别使用S1场景下9人的数据进行训练，使用其他2人在3个场景下的数据分别进行测试。对测试结果进行分析可知，Res3D网络在同一场景表现良好，测试识别精度达到了92.25%，同时Res3D网络对非训练场景的动作识别精度最高，网络的鲁棒性更强。与其他几种方法相比，Res3D网络没有对数据进行离散采样，保留了所有帧的雷达数据，所以取得了较高的动作识别精度，但也消耗了较多的运算量。

需要说明的是，本文实验中所采取的投影方式实现了数据的压缩，同时也造成了数据的损失，该方法并非最优，仅供数据库使用者参考，实际应用中也可以采取其他数据降维策略。另外，本文所使用的动作识别网络未根据雷达图像属性进行相应修改，识别率仍有较大的提升空间。因此，本文建议该数据集的未来研究可以从以下两个方面入手。

(1) 数据降维预处理。四维雷达图像区别于三维光学视频数据，多了一维距离信息，现有网络无法直接处理，因此如何设计方法对四维数据进行降维，且最大限度保留人体运动信息具有一定的研究价值。

(2) 网络结构设计。雷达系统与光学系统的成像机理存在差异，人体位置和运动特征分布略有不同。因此，可结合雷达系统提取雷达图像特有属性，如多普勒特征，合理设计网络结构和损失函数，进一步提高动作识别精度。

5 总结

针对基于雷达传感器的人体行为感知领域公开数据集缺乏的问题，本文公开了一种基于超宽带雷达四维成像的人体动作数据集，称为UWB-HA4D-1.0。该数据集以具有距离-方位-高度三维空间分辨能力的超宽带MIMO雷达为数据采集系统，通过MIMO雷达成像方法获取人体目标的三维成像结果，联合时间信息构成包含人体动作信息的四维成像。数据集有11个身高体重不同的人体目标、10种常见的动作类型、3种不同的测试场景，共计2757组人体动作数据，其中训练集1557组，3个测试集共1200组数据。本文对数据集的采集和制作、人体目标和动作信息，以及系统标注做了详细介绍。同时，分析了当前几种主流的动作识别方法。并使用了部分动作识别网络在本数据集上进行了实验测试，对实验结果进行分析和讨论，旨在为数据集使用者提供网络设计和动作识别精度参考。

本数据集是首个基于雷达四维成像的人体动作数据集，填补了此领域的空白。然而，本数据集仍存在一些问题待解决，比如系统成像帧率偏低、人体目标位置相对单一、动作类型丰富性不够、未对动作的起始帧和结束帧进行标注等问题。针对现存问题，下一步的任务有以下几项。

(1) 优化系统参数，提高信号帧率。较高的信号帧率可以更好地捕捉人体目标的动作信息，可基于高帧率雷达信号提取人体目标的微多普勒信息，与雷达四维成像结合进一步提高动作识别精度。

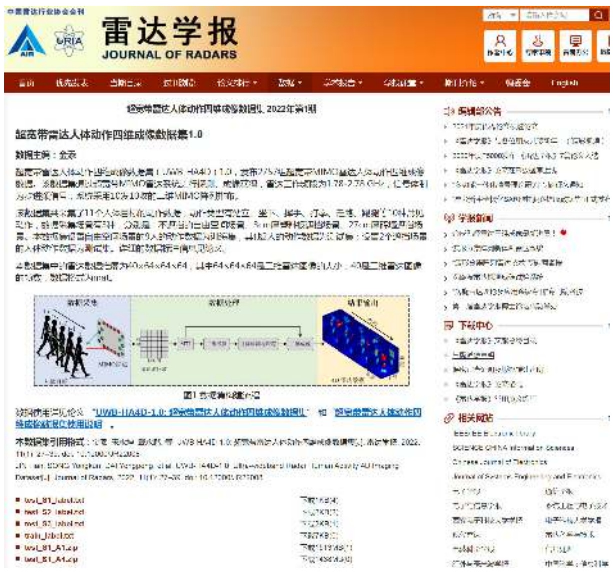
(2) 增加多位置、多人数据，丰富人体动作类型。三维超宽带雷达具有三维空间分辨能力，增加多人场景下的数据，可进行多人动作同时识别的方法研究，具有较大的实用意义。

(3) 对人体动作信息做更加详细的标注。对人体运动的起始帧和结束帧进行详细标注有助于人体目标动作进行定位和识别的多任务研究，加速推动雷达动作识别系统的实用化进程。

(4) 多传感器融合。单一传感器存在探测局限，未来可结合光学传感器、分布式雷达传感器等多源传感器实现对人体目标的全方位、多角度探测，提高人体行为感知能力。

附录

超宽带雷达人体动作四维成像数据集-1.0 (UWB-HA4D-1.0)依托《雷达学报》官方网站发布，数据于每次更新后上传至学报网站“超宽带雷达人体动作四维成像数据集-1.0”页面(附图1)，网址为：https://radars.ac.cn/web/data/getData?dataType=UWB-HA4D。由于网站存储空间限制，本次只上传了部分数据，完整数据的获取可以联系编辑部或作者。



附图 1 超宽带雷达人体动作四维成像数据集1.0发布网页

App. Fig. 1 Release webpage of ultra-wideband radar human activity 4D imaging dataset

参 考 文 献

[1] KUMAR P. Human activity recognition with deep learning: Overview, challenges & possibilities[J]. *CCF Transactions on Pervasive Computing and Interaction*, 2021, 339(3): 1–29. doi: 10.20944/preprints202102.0349.v1..

[2] 黄晴晴, 周风余, 刘美珍. 基于视频的人体动作识别算法综述[J]. 计算机应用研究, 2020, 37(11): 3213–3219. doi: 10.19734/j.issn.1001-3695.2019.08.0253.
HUANG Qingqing, ZHOU Fengyu, and LIU Meizhen. Survey of human action recognition algorithms based on video[J]. *Application Research of Computers*, 2020, 37(11): 3213–3219. doi: 10.19734/j.issn.1001-3695.2019.08.0253.

[3] 钱慧芳, 易剑平, 付云虎. 基于深度学习的人体动作识别综述[J]. 计算机科学与探索, 2021, 15(3): 438–455. doi: 10.3778/j.issn.1673-9418.2009095.
QIAN Huifang, YI Jianping, and FU Yunhu. Review of human action recognition based on deep learning[J]. *Journal of Frontiers of Computer Science & Technology*, 2021, 15(3): 438–455. doi: 10.3778/j.issn.1673-9418.2009095.

[4] SCHULDT C, LAPTEV I, and CAPUTO B. Recognizing human actions: A local SVM approach[C]. 2004 IEEE International Conference on Pattern Recognition, Cambridge, UK, 2004: 32–36. doi: 10.1109/ICPR.2004.1334462.

[5] SOOMRO K, ZAMIR A R, and SHAH M. UCF101: A dataset of 101 human actions classes from videos in the wild[EB/OL]. https://arxiv.org/abs/1212.0402, 2012.

[6] KUEHNE H, JHUANG H, GARROTE E, *et al.* HMDB: A large video database for human motion recognition[C]. 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 2011: 2556–2563. doi: 10.1109/ICCV.2011.6126543.

[7] KAY W, CARREIRA J, SIMONYAN K, *et al.* The kinetics human action video dataset[EB/OL]. https://arxiv.org/abs/1705.06950, 2017.

[8] SHAHROUDY A, LIU Jun, NG T T, *et al.* NTU RGB+D: A large scale dataset for 3D human activity analysis[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1010–1019. doi: 10.1109/CVPR.2016.115.

[9] 杜浩. 基于深度学习的超宽带雷达人体行为辨识研究[D]. [博士论文], 国防科技大学, 2020: 1–5.
DU Hao. Research on deep learning-based human behavior recognition in ultra-wideband radar[D]. [Ph. D. dissertation], National University of Defense Technology, 2020: 1–5.

[10] PAULI M, GOTTEL B, SCHERR S, *et al.* Miniaturized millimeter-wave radar sensor for high-accuracy applications[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2017, 65(5): 1707–1715. doi: 10.1109/TMTT.2017.2677910.

[11] 刘熠辰, 徐丰. 基于雷达技术的手势识别[J]. 中国电子科学研究院学报, 2016, 11(6): 609–613. doi: 10.3969/j.issn.1673-5692.2016.06.009.
LIU Yichen and XU Feng. Gesture recognition based on radar technology[J]. *Journal of China Academy of Electronics and Information Technology*, 2016, 11(6): 609–613. doi: 10.3969/j.issn.1673-5692.2016.06.009.

[12] DING Chuanwei, ZHANG Li, GU Chen, *et al.* Non-contact human motion recognition based on UWB radar[J]. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2018, 8(2): 306–315. doi: 10.1109/JETCAS.2018.2797313.

[13] 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.

[14] CRALEY J, MURRAY T S, MENDAT D R, *et al.* Action recognition using micro-Doppler signatures and a recurrent neural network[C]. 2017 51st Annual Conference on Information Sciences and Systems, Baltimore, USA, 2017: 1–5. doi: 10.1109/CISS.2017.7926181.

[15] WANG Mingyang, ZHANG Y D, and CUI Guolong. Human motion recognition exploiting radar with stacked recurrent neural network[J]. *Digital Signal Processing*, 2019, 87: 125–131. doi: 10.1016/j.dsp.2019.01.013.

[16] 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.

[17] DU Hao, JIN Tian, SONG Yongping, *et al.* A three-dimensional deep learning framework for human behavior analysis using range-Doppler time points[J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(4): 611–615. doi: 10.1109/LGRS.2019.2930636.

[18] 李廉林, 崔铁军. 智能电磁感知的若干进展[J]. 雷达学报, 2021, 10(2): 183–190. doi: 10.12000/JR21049.
LI Lianlin and CUI Tiejun. Recent progress in intelligent electromagnetic sensing[J]. *Journal of Radars*, 2021, 10(2): 183–190. doi: 10.12000/JR21049.

[19] LI Lianlin, SHUANG Ya, MA Qian, *et al.* Intelligent metasurface imager and recognizer[J]. *Light:Science & Applications*, 2019, 8(1): 97. doi: 10.1038/s41377-019-0209-z.

[20] FIORANELLI F, SHAH S A, LI Haobo, *et al.* Radar sensing for healthcare[J]. *Electronics Letters*, 2019, 55(19):

1022–1024. doi: 10.1049/el.2019.2378.

[21] MENG Zhen, FU Song, YAN Jie, *et al*. Gait recognition for co-existing multiple people using millimeter wave sensing[C]. The AAAI Conference on Artificial Intelligence, New York, USA, 2020: 849–856.

[22] ZHU Zhengliang, YANG Degui, ZHANG Junchao, *et al*. Dataset of human motion status using IR-UWB through-wall radar[J]. *Journal of Systems Engineering and Electronics*, 2021, 32(5): 1083–1096. doi: 10.23919/JSEE.2021.000093.

[23] SONG Yongkun, JIN Tian, DAI Yongpeng, *et al*. Through-wall human pose reconstruction via UWB MIMO radar and 3D CNN[J]. *Remote Sensing*, 2021, 13(2): 241. doi: 10.3390/rs13020241.

[24] AMIN M G, 朱国富, 陆必应, 金添, 等译. 穿墙雷达成像[M]. 北京: 电子工业出版社, 2014: 22–25.
AMIN M G, ZHU Guofu, LU Biying, JIN Tian, *et al*. translation. Through-The-Wall Radar Imaging[M]. Beijing, China: Publishing House of Electronic Industry, 2014: 22–25.

[25] 詹姆斯 D. 泰勒, 胡春明, 王建明, 孙俊, 等译. 超宽带雷达应用与设计[M]. 北京: 电子工业出版社, 2017: 54–55.
TAYLOR J D, HU Chunming, WANG Jianming, SUN Jun, *et al*. translation. Ultrawideband Radar: Applications and Design[M]. Beijing, China: Publishing House of Electronic Industry, 2017: 54–55.

[26] 孙鑫. 超宽带穿墙雷达成像方法与技术研究[D]. [博士论文], 国防科学技术大学, 2015: 16–17.
SUN Xin. Research on method and technique of ultra-wideband through-the-wall radar imaging[D]. [Ph. D. dissertation], National University of Defense Technology, 2015: 16–17.

[27] 金添, 宋勇平. 穿墙雷达人体目标探测技术综述[J]. 电波科学学报, 2020, 35(4): 486–495. doi: 10.13443/j.cjors.2020040804.
JIN Tian and SONG Yongping. Review on human target detection using through-wall radar[J]. *Chinese Journal of Radio Science*, 2020, 35(4): 486–495. doi: 10.13443/j.cjors.2020040804.

[28] ASH M, RITCHIE M, and CHETTY K. On the application of digital moving target indication techniques to short-range FMCW radar data[J]. *IEEE Sensors Journal*, 2018, 18(10): 4167–4175. doi: 10.1109/JSEN.2018.2823588.

[29] SONG Yongping, LOU Jun, and TIAN Jin. A novel II-CFAR detector for ROI extraction in SAR image[C]. 2013 IEEE International Conference on Signal Processing, Communication and Computing, Kunming, China, 2013: 1–4. doi: 10.1109/ICSPCC.2013.6663879.

[30] NORTON-WAYNE L. Image reconstruction from projections[J]. *Optica Acta:International Journal of Optics*, 1980, 27(3): 281–282. doi: 10.1080/713820221.

[31] MCCORKLE J W. Focusing of synthetic aperture ultra wideband data[C]. 1991 IEEE International Conference on Systems Engineering, Dayton, USA, 1991: 1–5. doi: 10.1109/ICSYSE.1991.161068.

[32] BOBICK A F and DAVIS J W. The recognition of human movement using temporal templates[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(3): 257–267. doi: 10.1109/34.910878.

[33] DAS DAWN D and SHAIKH S H. A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector[J]. *The Visual Computer*, 2016, 32(3): 289–306. doi: 10.1007/s00371-015-1066-2.

[34] WANG Heng, KLÄSER A, SCHMID C, *et al*. Dense trajectories and motion boundary descriptors for action recognition[J]. *International Journal of Computer Vision*, 2013, 103(1): 60–79. doi: 10.1007/s11263-012-0594-8.

[35] SIMONYAN K and ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]. The 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 568–576. doi: 10.5555/2968826.2968890.

[36] WANG Limin, XIONG Yuanjun, WANG Zhe, *et al*. Temporal segment networks: Towards good practices for deep action recognition[C]. 2016 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 20–36. doi: 10.1007/978-3-319-46484-8_2.

[37] LIN Ji, GAN Chuang, and HAN Song. TSM: Temporal shift module for efficient video understanding[C]. 2019 IEEE/CVF IEEE International Conference on Computer Vision, Seoul, Korea, 2019: 7083–7093. doi: 10.1109/ICCV.2019.00718.

[38] JI Shuiwang, XU Wei, YANG Ming, *et al*. 3D convolutional neural networks for human action recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221–231. doi: 10.1109/TPAMI.2012.59.

[39] TRAN D, BOURDEV L, FERGUS R, *et al*. Learning spatiotemporal features with 3D convolutional networks[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 4489–4497. doi: 10.1109/ICCV.2015.510.

[40] TRAN D, RAY J, SHOU Zheng, *et al*. ConvNet architecture search for spatiotemporal feature learning[EB/OL]. https://arxiv.org/abs/1708.05038, 2017.

[41] FEICHTENHOFER C, FAN Haoqi, MALIK J, *et al*. SlowFast networks for video recognition[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 6202–6210. doi: 10.1109/ICCV.2019.00630.

作 者 简 介

金　添(1980–)，男，湖北人，国防科技大学教授、博士生导师。主要研究方向为新体制雷达系统、智能感知与处理。全国百篇优秀博士论文获得者，国际无线电科学联盟青年科学家奖，入选教育部新世纪优秀人才支持计划，中国电子学会优秀科技工作者。“信号处理与系统”国家精品课程和资源共享课主讲教师，信号处理系列课程国家级教学团队主要成员。出版专著4部、译著1部、教材1部，发表论文百余篇，授权国家发明专利10余项。获省部级科技进步一等奖1项、二等奖2项，电子学会自然科学二等奖1项。中国电子学会雷达分会委员、信号处理分会委员，《雷达学报》、《信号处理》、《雷达科学与技术》、《现代雷达》等期刊编委。多次担任APSAR国际会议、CIE国际雷达会议、IET国际雷达会议等TPC委员或分会主席。

宋永坤(1993–)，男，河南人，国防科技大学信息与通信工程专业博士研究生。主要研究方向为超宽带雷达信号处理及深度学习。

戴永鹏(1992–)，男，山东人，国防科技大学电子科学学院讲师，博士。主要研究方向为MIMO阵列雷达成像与图像增强。

胡锡坤(1994–)，男，湖北人，国防科技大学信息与通信工程专业博士研究生。主要研究方向为遥感图像处理和深度学习。

宋勇平(1989–)，男，四川人，国防科技大学电子科学学院助理研究员，博士。主要研究方向为穿墙探测、MIMO雷达成像、微弱目标检测。

周小龙(1992–)，男，江西人。国防科技大学信息与通信工程专业博士研究生。主要研究方向为雷达信号处理、人体行为识别。

邱志峰(1999–)，男，江西人，国防科技大学电子科学学院硕士研究生。主要研究方向为雷达信号处理与深度学习。