

基于无线信号的人体姿态估计综述

陈彦*^① 张锐^① 李亚东^② 宋瑞源^① 耿瑞旭^① 龚汉钦^①
汪斌全^① 张东恒^① 胡洋^①

^①(中国科学技术大学网络空间安全学院 合肥 230026)

^②(华盛顿大学电气与计算机工程学院 西雅图 98195)

摘要: 人体姿态估计在人机交互、动作捕捉和虚拟现实等领域具有广泛的应用前景,一直是人体感知研究的重要方向。然而,基于光学图像的姿态估计方法往往受限于光照条件和隐私问题。因此,利用可在各种光照遮挡下工作,且具有隐私保护性的无线信号进行人体姿态估计获得了更多关注。根据无线信号的工作频率,现有技术可分为高频方法和低频方法,且不同的信号频率对应硬件系统、信号特性、噪声处理和深度学习算法设计等方面均有所不同。该文将以毫米波雷达、穿墙雷达和WiFi信号为代表,回顾其在人体姿态重建研究中的进展和代表性工作,分析各类信号模式的优势与局限,并对潜在研究难点以及未来发展趋势进行了展望。

关键词: 人体姿态估计; 无线感知; 深度学习; 毫米波雷达; 穿墙雷达; WiFi

中图分类号: TN957.51

文献标识码: A

文章编号: 2095-283X(2025)01-0229-19

DOI: 10.12000/JR24189

CSTR: 32380.14.JR24189

引用格式: 陈彦, 张锐, 李亚东, 等. 基于无线信号的人体姿态估计综述[J]. 雷达学报(中英文), 2025, 14(1): 229-247. doi: 10.12000/JR24189.

Reference format: CHEN Yan, ZHANG Rui, LI Yadong, *et al.* An overview of human pose estimation based on wireless signals[J]. *Journal of Radars*, 2025, 14(1): 229-247. doi: 10.12000/JR24189.

An Overview of Human Pose Estimation Based on Wireless Signals

CHEN Yan*^① ZHANG Rui^① LI Yadong^② SONG Ruiyuan^① GENG Ruixu^①
GONG Hanqin^① WANG Binquan^① ZHANG Dongheng^① HU Yang^①

^①(School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230026, China)

^②(Department of Electrical and Computer Engineering, University of Washington, Seattle 98195, United States)

Abstract: Human pose estimation holds tremendous potential in fields such as human-computer interaction, motion capture, and virtual reality, making it a focus in human perception research. However, optical image-based pose estimation methods are often limited by lighting conditions and privacy concerns. Therefore, the use of wireless signals that can operate under various lighting conditions and obstructions while ensuring privacy is gaining increasing attention for human pose estimation. Wireless signal-based pose estimation technologies can be categorized into high-frequency and low-frequency methods. These methods differ in their hardware systems, signal characteristics, noise processing, and deep learning algorithm design based on the signal frequency used. This paper highlights research advancements and notable works in human pose reconstruction using millimeter-wave radar, through-wall radar, and WiFi. It analyzes the advantages and limitations of each signal type and explores potential research challenges and future developments in the field.

收稿日期: 2024-09-16; 改回日期: 2024-11-07; 网络出版: 2024-11-26

*通信作者: 陈彦 eecyan@ustc.edu.cn *Corresponding Author: CHEN Yan, eecyan@ustc.edu.cn

基金项目: 国家自然科学基金(62172381, 62201542)

Foundation Items: The National Natural Science Foundation of China (62172381, 62201542)

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/>)

Key words: Human pose estimation; Wireless sensing; Deep learning; Millimeter-wave radar; Through-Wall Radar (TWR); WiFi

1 引言

人体姿态估计技术在人机交互、动作捕捉、虚拟现实等领域具有广泛的应用前景,在过去数十年间一直是人体感知领域的研究热点。人体姿态估计旨在从输入的传感器数据中分析人体的形态与运动特征,进而估计出人体各个部位的姿态。随着近年来计算机视觉以及深度学习领域的快速发展,基于光学图像的各类数据驱动方法相继涌现,在人体姿态估计、目标检测和活动识别等各类人体感知任务上展现出了优异的性能^[1-4]。然而,由于光学相机采集的图像质量易受光照条件的影响,此类方法在复杂实际环境中的性能表现难以获得保证。此外,当目标被障碍物完全遮挡时,基于光学图像的人体姿态估计系统通常无法有效工作。除了对环境鲁棒性不足外,基于光学图像的感知技术通常会采集过多的人体身份特征以及其他敏感信息,从而引发用户隐私泄露的问题。

相比之下,随着无线技术的快速发展,越来越多的研究人员开始关注如何利用各种无线信号进行人体感知^[5-20]。由于人体活动会改变无线信号的传播方式,通过分析无线信号的变化来估计人体姿态成为一种新的方法^[21,22]。与光学系统(如光学相机)不同,无线信号的传播不受光照条件影响,能够在黑暗环境中正常工作,并且能够穿透非金属墙壁和多种障碍物。此外,无线信号特征的视觉可解释性较低,能够更好地保护用户隐私。因此,基于无线信号的人体感知技术作为光学方法的重要补充和延伸,已经在智能家居^[23-25]、医疗健康^[26]、人机交互^[27,28]等领域取得了显著的研究进展。根据无线信号的工作频率,基于无线信号的人体姿态估计技术可以分为以下两大类:

第1类方法主要在极高频率下工作,例如毫米波雷达^[28]和太赫兹雷达。此类高频设备的波长较短,特别是太赫兹雷达,其波长与人体表面粗糙度相当,因此人体目标可被视为散射体而非反射体。在此情况下,当阵列孔径足够大时,雷达系统可以直接对人体表面进行高精度成像。另外,毫米波雷达通过发射短波长电磁波并接收反射信号,来分析目标的形状、距离和运动状态,进而获得人体姿态信息。近年来,毫米波雷达与深度学习技术的融合,极大地推动了人体姿态估计技术的发展。例如,Sengupta等人^[29]提出了一种双分支的卷积神经网络架构,用于基于毫米波雷达的人体姿态估计,显著

提升了估计的精度和鲁棒性。此外,Xie等人^[18]开发了一套名为“RPM”的人体骨架姿态估计系统,通过结合毫米波雷达和神经网络,实现了高精度的姿态估计。

第2类方法主要在较低频段工作,例如穿墙雷达(Through-Wall Radar, TWR)^[30-32]和WiFi阵列^[33]。由于低频信号的波长较长,这类设备通常具有较低的成像分辨率和更为显著的镜面反射现象,单个时刻的回波所携带的人体信息可能并不完整。尽管如此,相较于毫米波雷达,此类低频设备的主要特点是其发射信号能够穿透非金属墙壁和其他障碍物以感知目标,在复杂环境例如军事、救援以及监控等领域具有广泛应用。例如,Adib等人^[30]展示了如何利用穿墙雷达技术捕捉到墙壁后的人体三维信号,并利用信号处理和深度学习算法实现了人体骨架的精确重建^[34]。此外,WiFi信号作为一种泛在的无线通信技术,也已被应用于人体姿态估计研究。WiFi信号的优势在于其普遍性,有望在结合现有通信设备功能的基础上实现感知与通信的一体化。通过分析WiFi信号的反射和散射特性,研究人员能够推断出人体的位置和姿态。例如,Holl和Reinhard^[33]利用WiFi信号的全息成像技术实现了空间物体的三维重建。Jiang等人^[35]进一步将此技术应用于人体姿态估计,通过构建WiFi信号的三维特征,精确重建了人体姿态。

尽管近年来基于无线信号的姿态估计技术取得了显著的进步,但仍面临着诸多挑战。首先,不同的无线信号在分辨率、穿透能力和环境适应性上存在显著差异,有效选择信号模态并设计适应信号特性的处理算法是实现高精度姿态估计的关键。其次,信号传播特性以及环境噪声对姿态估计的准确性有着显著影响,因此,如何通过信号处理提取更鲁棒的信号特征,仍然是一个亟待深入研究的方向。此外,深度学习模型的设计与优化对姿态估计的性能有着直接的影响,因此结合无线信号特性设计和优化模型也是未来研究的重点之一。如表1^[14,18,19,29,34,36-43]所示,本文将依次介绍毫米波雷达、穿墙雷达以及WiFi信号在人体姿态估计中的应用,并详细回顾各信号模态的技术特点和解决方案。此外,本文总结了两个经典的基于毫米波和穿墙雷达的开源数据集,并对比分析了各类人体感知技术。最后,本文对未来的研究方向和挑战进行了展望并总结全文。

2 人体姿态模型

人体姿态模型是人体姿态估计(Human Pose Estimation, HPE)的核心组成部分, 通常被划分为3大类: 运动学模型、平面模型和体积模型^[2]。运动学模型可用于2D和3D的姿态估计。其通过模拟人体的关节结构和运动学特性, 从而准确捕捉人体的动态变化, 常用于涉及关节运动分析的应用; 平面模型则主要应用于2D HPE, 其采用二维坐标来表示人体关键点, 具有计算量小、效率高的特点, 适合资源受限的场景; 体积模型则专注于3D HPE, 其基于三维体积的表示方式, 能够更全面地反映人体的姿态和形状变化, 广泛用于虚拟环境、运动捕捉及医疗分析等需要精细化姿态描述领域。

2.1 运动学模型

运动学模型通过一组关节位置和肢体方向来表示人体的结构(如图1(a)所示)。一种被广泛使用的图模型是图片结构模型(Picture Structure Model, PSM), 也称为树状结构模型。PSM因其直观且灵活的结构设计, 广泛应用于2D HPE^[44,45]和3D HPE^[46,47]中。该模型将人体关节视为图中的节点, 肢体视为连接的边, 能够有效地表达人体各个部分的相对位置和运动关系, 尤其适用于姿态的动态变化建模。Boulic模型是另一种基于运动学的人体姿态模型, 强调对关节运动和生物力学约束的精确描述。该模型由Boulic等人^[48]提出, 通过引入关节的生理限制和运动学参数, 能够更加真实地模拟人体的自然运

表 1 基于无线信号的人体姿态估计研究现状总结
Tab. 1 Summary of research status on pose estimation based on wireless signals

基于频率的分类	设备	雷达特征信息	代表性工作
基于高频无线信号的人体姿态估计	毫米波雷达 (30~300 GHz)	3D point cloud	mmPose ^[29]
		Heatmap	RPM ^[18]
		Heatmap	RPM 2.0 ^[14]
		Heatmap	MobiRFPose ^[19]
		Heatmap	RF-Pose ^[34]
基于低频无线信号的人体姿态估计	穿墙雷达 (300 MHz~10 GHz)	Heatmap	RF-Pose3D ^[30]
		单帧3D成像体素	MIMDSN ^[37]
		多帧3D成像体素	ST ² W-AP ^[38]
		Heatmap和3D成像体素	Dual-task Net ^[39]
		多帧雷达回波	RadarFormer ^[40]
WiFi (2.400~5.825 GHz)	WiFi (2.400~5.825 GHz)	Channel state information	Person-in-WiFi ^[41]
		Channel state information	Person-in-WiFi 3D ^[42]
		Channel state information	DensePose From WiFi ^[43]

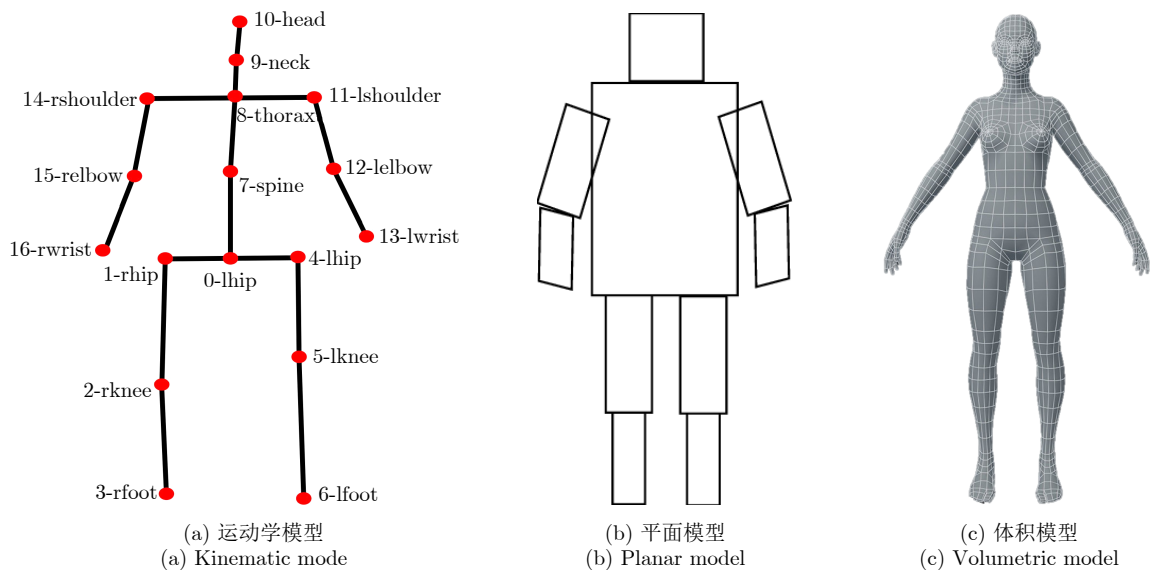


图 1 人体姿态模型

Fig. 1 Human pose models

动。Boulic模型特别关注关节的自由度、运动范围以及关节之间的相互影响,从而在模拟中避免出现不符合生物学规律的姿态。与传统的运动学模型相比,Boulic模型在处理复杂的关节限制和运动协调方面具有优势。在动画制作和虚拟现实等领域,Boulic模型被广泛应用于提高虚拟人角色的动作真实性和自然性。例如,在步态模拟中,Boulic模型可以精确地再现人类行走时姿态变化和重心转移^[49]。此外,该模型也被用于生物力学研究和康复医学中,帮助专业人员分析和评估患者的运动功能。然而,尽管Boulic模型在提高模拟精度方面具有优势,其复杂性也带来了计算量较大的问题。在实时应用中,需要对模型进行适当的简化或优化,以满足性能要求。尽管运动学模型在捕捉人体的关节运动和相对姿态方面较为灵活,特别是在应对复杂的姿态变化时表现出较好的适应性,但其在表示人体表面纹理和形状等细节信息上存在一定局限。因此,运动学模型通常需要与其他模型(如体积模型)相结合,以更全面地刻画人体的形态特征。

2.2 平面模型

与运动学模型不同,平面模型主要用于捕捉人体的形状和外观特征。它通过表示人体各部分之间的几何关系,使用矩形等简单的几何形状来近似人体轮廓(如图1(b)所示)。其中,纸板模型^[50]是一个典型的例子,其通过多个矩形表示人体的四肢和躯干结构,早期的研究^[51]已成功应用该模型进行人体姿态估计。此外,主动形状模型^[52]也是一种常见的平面模型,其通过主成分分析从大量人体图像中提取并捕捉人体的轮廓信息^[53,54]。相比纸板模型,主动形状模型能够更精确地描述人体的形状变化,在细致的姿态重建和外形捕捉任务中表现出色。然而,平面模型在三维信息表达上存在局限,通常需要结合其他方法来提升复杂场景下的姿态估计性能。

2.3 体积模型

随着3D人体重建技术的快速发展,研究人员提出体积模型用于表示多样化的人体形态(如图1(c)所示)。其中,多人物线性蒙皮模型(Skinned Multi-Person Linear model, SMPL)^[55]是3D HPE领域广泛应用的模型之一。它通过基于姿态的自然变形来精确建模软组织的动态变化。研究者利用1786个高分辨率的3D人体扫描样本,使用SMPL模板网格优化了关节权重、姿态依赖的形变、平均模板形状以及从顶点到关节位置的回归器,从而提升了姿态估计的精度。除了SMPL外,还有一些其他著名的体积模型,如PIFu^[56], DYNA^[57], Stitched Puppet^[58],

Frankenstein & Adam^[59]以及GHUM和GHUMLite^[60],这些模型使得姿态估计的准确性和细节表达取得了显著提升,在虚拟现实、动画制作和医学领域中得到了广泛应用。

3 基于毫米波雷达的人体姿态估计

毫米波雷达工作在30~300 GHz的频段,波长介于1~10 mm,凭借其短波长、高频率和高带宽的特性,能够提供高分辨率的空间信息,并且具备一定的穿透能力,能够穿透衣物。这些特性使毫米波雷达在复杂环境中对人体信息进行捕捉时表现出显著优势,特别是在对隐私保护和空间分辨率要求较高的场景中,如医疗监控、智能家居和安防等应用。毫米波雷达通过发射电磁波并接收目标物体反射的信号,结合多普勒频移、距离和角度信息,推断出物体的三维位置和运动状态。与其他类型的雷达(如超声波雷达和微波雷达)相比,毫米波雷达因其较短的波长而更加紧凑且多径效应更弱,能够在狭小空间内实现更高的距离分辨率。此外,毫米波雷达不仅可以获取目标物体的距离,还能同时捕获目标速度和角度信息,从而更精确地反映人体在动态环境中的运动。毫米波雷达在人体姿态估计中的广泛应用得益于其独特的技术优势,主要体现在以下几个方面。

(1) 高分辨率:毫米波雷达能够捕捉到细微的运动变化,适用于需要高精度姿态估计的场景。

(2) 穿透能力:毫米波雷达能够穿透衣物和薄壁,适用于复杂环境中的目标检测场景。

(3) 实时性强:毫米波雷达具有快速的数据采集和处理能力,能够提供实时的姿态估计反馈。

近年来,基于毫米波雷达的人体姿态估计技术快速发展,主要经历了从单目标到多目标场景、从复杂计算到轻量化实时处理的演进过程。本节将详细阐述4种基于毫米波雷达的代表性人体姿态估计方法,涉及硬件架构、数据设计和网络设计等方面。这些典型方法包括mm-Pose^[29], RPM^[18], RPM 2.0^[14]以及MobiRFPose^[19],其中每种方法都在其前身基础上进行创新,共同推动了毫米波雷达在人体姿态估计领域的应用,为复杂环境下的人体运动捕捉提供了不同的解决方案。

3.1 毫米波雷达姿态估计的开创性方法: mm-Pose

Sengupta等人^[29]首次提出了mm-Pose,这是一个开创性的研究,首次利用毫米波雷达进行人体姿态估计。该系统能够通过毫米波雷达精确估计超过15个不同人体关节位置。该系统巧妙地结合了毫米波雷达的高分辨率特性与深度学习技术,实现了

在动态复杂环境中的实时、高精度人体姿态估计，为交通监控和医疗监护等领域提供了一种全新的解决方案。

在硬件配置上，mm-Pose采用了Texas Instruments的AWR 1642毫米波雷达，具备两个发射通道和四个接收通道。两个雷达模块分别以水平和垂直方向安装，确保了对目标深度、方位角和仰角的全方位检测。雷达工作频率为79 GHz，带宽3.072 GHz，数据采集速率为每秒20帧，能够捕捉256个反射点的信息。数据预处理阶段，雷达通过毫米波信号捕捉目标反射点云，生成深度、方位角和仰角的三维信息。为了应对数据稀疏性和降低计算复杂度，mm-Pose将点云数据投影至深度-方位角(XY 平面)和深度-仰角(XZ 平面)，形成16像素 \times 16像素的3通道图像。其中，3个通道分别表示点的 x 坐标、 y 或 z 坐标以及归一化后的反射强度，这种处理方式不仅保留了关键信息，还便于后续的深度学习模型处理。

在mm-Pose的网络设计中，作者提出了一种新颖的分叉式卷积神经网络(Convolutional Neural Network, CNN)架构，利用毫米波雷达捕获的数据来估计关节位置。网络输入为两张3通道图像，每张图像分别通过独立的3层CNN进行处理，采用ReLU作为激活函数，且在每层后使用Dropout来

减少过拟合的风险。在CNN提取出人体关节位置相关的特征后，输出特征会被拼接成一个 $N \times N \times 128$ 的张量，输入全连接层进一步处理。最后，输出层预测25个人体关节的三维位置。整个网络架构的轻量化设计显著减少了计算复杂度，使得毫米波雷达数据的实时处理成为可能。

mm-Pose方法具有实时性强、计算复杂度低的优点，能够在不同环境下通过毫米波雷达有效估计人体骨架姿态，且不易受到光线条件影响。然而，该方法在处理复杂姿态时的表现仍有限，尤其是在捕捉更细小的关节如手腕和手指时，预测精度较低。数据采集和标注过程昂贵，限制了模型的泛化性和在多种姿态下的鲁棒性。

3.2 单人场景下的姿态估计：RPM

人体对于无线信号具有镜面反射特性，这意味着单帧射频信号无法完全捕捉人体。因此，人体反射的无线信号往往是稀疏且不完整的，这使得提取高质量特征以用于三维人体姿态估计变得困难。考虑到这些射频信号的特性，Xie等人^[18]提出了一种全新的框架RPM，该框架能够从射频信号中生成三维骨骼，框架具体结构如图2所示。

硬件部分使用了毫米波雷达和多相机系统。雷达采用了两个TI MMWCAS-RF-EVM雷达，通过

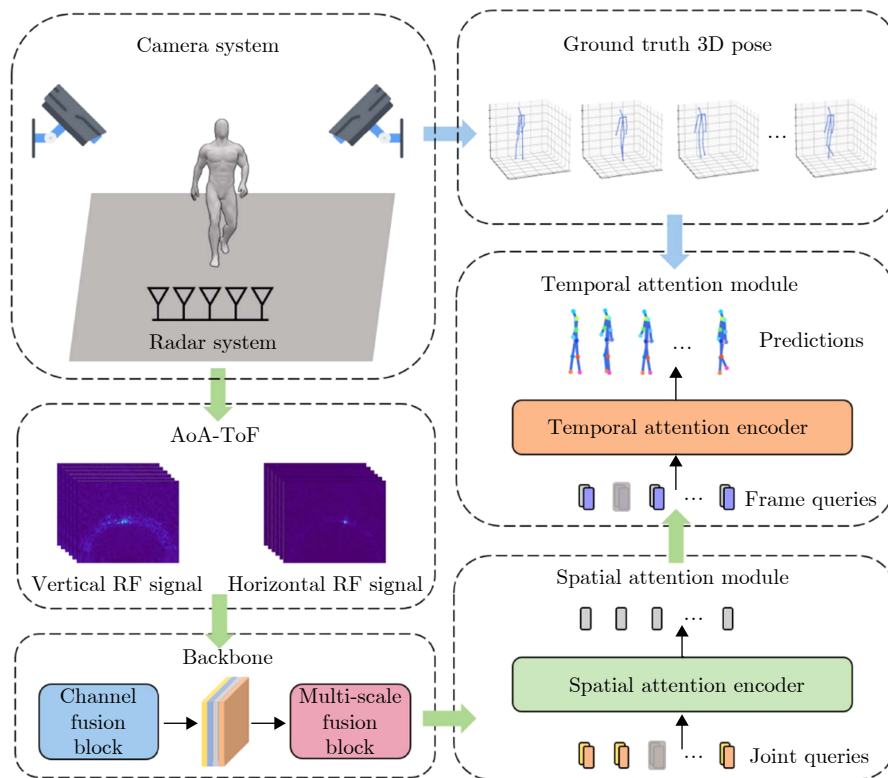


图 2 RPM模型框架图^[18]

Fig. 2 Diagram of the RPM framework^[18]

天线阵列接收反射信号,并将信号解构为水平和竖直两平面的热图。这些热图提供了人体不同部位的位置信息,通过对这些信号的多尺度网络处理,能够提取出人体的三维骨架结构。此外,为了获取准确的3D骨架标签,RPM还构建了一个多相机系统采集真值。这个系统由13个安装在不同角度的Raspberry Pi相机组成,这些摄像头使用广角镜头,并通过局域网内的网络时间协议(Network Time Protocol, NTP)服务器进行时间同步。通过多视角的2D骨架信息,系统可以利用三角测量法生成精确的3D骨架标签。这个多相机系统生成的人体骨架信息与雷达信号同步,作为深度学习模型的监督标签。

网络部分由特征融合网络和时空注意力网络两大模块组成。特征融合网络主要用于从雷达信号中提取高分辨率的特征。由于雷达信号在水平和垂直方向上存在不同的特性,水平信号用于提供人体位置信息,而垂直信号则反映人体部位的高度和深度信息。为了融合这两种方向的信号,作者设计了一个通道融合模块,该模块使用卷积层和瓶颈块来进行通道级别的特征融合。此外,文章还提出了多尺度融合模块,通过卷积层来提取人体特征,并结合可变形卷积来捕捉形状不同的特征,增强了网络对人体姿态的识别能力。时空注意力网络(Spatial-Temporal Attention Network, STAN)是文章的另一核心部分,用于从稀疏的雷达信号中重建人体的3D骨架。STAN包含空间注意力模块和时间注意力模块。空间注意力模块旨在通过多头自注意力机制来建模关节之间的非局部关系,从而重建雷达信号未能捕捉到的身体部位。具体实现上,网络会随机屏蔽部分输入关节特征,通过预测这些缺失的特征来模拟人体反射信号不完整的情况。时间注意力模块则通过建模帧与帧之间的时间依赖性,进一步优化人体的3D姿态序列。通过随机屏蔽某些帧,网络能够学习如何从时间维度上重建完整的3D骨架序列。

RPM通过特征融合网络与时空注意力网络的协同作用,成功解决了毫米波雷达信号稀疏、人体反射不完全以及时间一致性问题,实现了高精度的3D人体姿态估计。尽管RPM在单目标场景中表现出色,但未考虑多目标场景中的复杂性。

3.3 多人场景下的姿态估计: RPM 2.0

RPM 2.0^[14]是基于毫米波雷达的人体姿态估计技术的进一步发展。相比其前身RPM,RPM 2.0解决了1.0版本中存在的一些关键问题,尤其是有效改善了多目标场景中由于多径效应和信号重叠导

致的不同目标间信号混淆的问题,从而使多目标场景下的姿态估计结果更加稳定。RPM 2.0采用和1.0版本相同的两个正交TI MMWCAS-RF-EVM雷达系统和由13个Raspberry Pi相机组成的多视角相机系统。

RPM 2.0的优势主要体现在网络框架上。RPM 2.0进一步深化了对多人场景下3D姿态估计,通过特征提取网络、多视角融合网络和时空注意力网络3大核心模块的协同作用,显著提升了在复杂环境中的姿态估计精度和鲁棒性。(1)特征提取网络作为RPM 2.0的基础,用于从毫米波雷达信号中提取高质量的特征表示。面对多人场景中信号的稀疏性和重叠挑战,RPM 2.0引入了轻量级的无锚定检测模块,该模块能够直接定位并裁剪雷达信号中的目标区域,生成精确的边界框,克服了传统基于锚点方法的冗余检测问题。这一模块不仅加速了目标检测过程,还确保了在处理小目标时信息的完整性。网络的骨干部分采用多尺度特征建模策略,增强了对不同大小目标的适应能力。(2)多视角融合网络解决了毫米波雷达在不同视角下信号统一的难题。RPM 2.0将水平和垂直方向的雷达信号视为具有独特视角的“RF相机”,通过该网络将这些信号映射至统一的潜在特征空间,再依据视角间的几何关系进行加权融合。这种设计策略不仅提高了特征的表达能力,还确保了从多角度获取的雷达信号能够被有效整合。(3)时空注意力网络则继承了RPM 1.0中的STAN模块。通过上述3大模块的协同工作,RPM 2.0不仅有效解决了毫米波雷达在多人3D姿态估计中面临的信号稀疏、重叠和不完整等挑战,还通过时空注意力机制的引入,显著增强了姿态估计在复杂场景下的精度和鲁棒性,为毫米波雷达在人体姿态估计领域的应用开辟了新的可能。

总体而言,RPM 2.0显著提升了多目标场景下的姿态估计能力,但计算资源开销仍然较高,在实时性、降低计算复杂性和模型轻量化方面还有待进一步优化。

3.4 轻量化实时姿态估计: MobiRFPose

到RPM 2.0为止,已有的基于射频的人体姿态估计方法通常需要大量的计算成本,无法满足移动设备的实时处理和可移植性要求。为了解决这个问题,文献[19]提出了一种轻量级的基于射频的姿态估计模型,即MobiRFPose。

MobiRFPose的网络架构由人体位置估计模块(Human Localization Network, HLN)和姿态估计模块(Pose Estimation Network, PEN)两大核心模块构成。HLN模块首先根据水平RF热图进行人体

位置的初步估计, 通过裁剪出包含人体的局部热图区域, 有效减少了后续处理的计算量, 避免了多人场景下信号重叠带来的复杂性。随后, PEN模块对裁剪后的局部热图进行深度处理, 精确估计出人体的姿态信息。这种“裁剪-估计”的策略不仅简化了计算流程, 还确保了在多人场景中对每个个体姿态的精准识别。

为了满足轻量化和实时性的需求, MobiRFPose在模型设计上进行了深度优化, 大幅减少了网络参数和运算量。与传统的基于RF信号的3D姿态估计模型(如RF-Pose3D)相比, MobiRFPose在保持高精度的同时, 所需参数和计算量分别减少了90%和99%, 实现了显著的性能提升。这种轻量化设计使得MobiRFPose能够在移动设备上实现高效、实时的姿态估计, 测试结果显示其处理速度可达66 帧/s (Frames Per Second, FPS), 模型参数占用仅为268 KB, 展现了卓越的计算效率和资源利用率。

MobiRFPose能够实现高效姿态估计, 且对环境具备优秀的适应性和鲁棒性, 使RF姿态估计能够在便携式设备上实时运行。

3.5 未来展望

未来, 基于毫米波雷达的人体姿态估计技术有望在算法和部署两个层面取得突破性进展。(1)算法层面: 多模态融合和模态对齐。随着多模态传感器融合技术的发展, 结合深度相机等其他传感器数据, 将显著提高姿态估计的精度和鲁棒性。同时, 深度学习和迁移学习的进步将促进更通用、适应性更强的模型出现, 能够快速适应不同环境和任务需求。(2)部署层面: 边缘和移动部署。随着算法优化和硬件进步, 该技术将更好地适应边缘计算设备, 实现更广泛的部署。这些进展将推动毫米波雷达人体姿态估计技术在虚拟现实、康复医疗、智能家居等多个领域的创新应用, 为未来智能系统提供更精确、高效且普适的人体感知能力。

4 基于穿墙雷达的人体姿态重建

随着无线通信和传感技术的迅速发展, 穿墙雷达在人体姿态估计领域的应用逐渐成为研究的热点。与传统的光学传感器相比, 穿墙雷达能够穿透混凝土和砖墙等障碍物, 因此在遮挡、光线不足或隐私敏感的复杂场景中表现尤为出色。本文介绍了3类基于穿墙雷达的人体姿态估计方法。

第1类方法使用成像算法将雷达回波转换为雷达图像, 结合深度学习技术从图像中提取人体信息。如图3所示, 典型的研究工作包括Zhao等人提出的RF-Pose^[34]和RF-Pose3D^[36], Zheng等人^[37]的

分辨率引导姿态估计方法MIMDSN, 张锐等人^[38]的时空融合网络ST²W-AP, 以及Song等人^[39]的双分支网络Dual-task Net。这类方法通过成像算法的干扰抑制和目标聚焦作用, 通常能够提供较高的姿态估计精度。然而, 成像算法的引入扩大了数据维度, 增加了深度学习模型的处理复杂度。因此, 这类方法必须在运算效率和性能之间进行权衡。例如, RF-Pose^[34]和RF-Pose3D^[36]通过将高维雷达图像投影为多个低维图像来提高效率, MIMDSN^[37]采用单帧雷达图像结合定制化训练策略, ST²W-AP^[38]通过优化模型结构, 减少计算开销, 提升模型性能。

第2类方法则跳过了成像步骤, 直接利用深度学习从雷达回波中提取人体信息。Zheng等人^[40]提出的RadarFormer便是这一类的代表, 使用端到端信号处理直接从雷达回波实现人体感知。由于不需要成像过程, RadarFormer具备更高的计算效率和简化的模型设计, 不需要额外模块的支持。然而, 缺少成像算法带来的干扰抑制和目标聚焦功能, 使得从原始回波中提取人体信息难度增加, 导致该模型的训练复杂度较高。

第3类方法是利用微多普勒效应从人体运动中提取关键信息, 进行人体姿态重建。不同于前两类方法, 该方法专注于分析雷达信号中的频率偏移, 特别是由人体不同部位运动带来的微多普勒特征。微多普勒效应反映了人体各个部分(如四肢和躯干)运动的速度变化, 因此, 可以通过频谱分析来捕捉人体姿态的细粒度变化。

4.1 基于2D成像热图的态度估计

4.1.1 穿墙雷达中的2D姿态估计: RF-Pose

RF-Pose是首个将无线信号与深度学习结合, 实现人体姿态估计的开创性工作。Zhao等人^[34]利用WiFi频段的无线信号, 通过与光学信息的跨模态监督, 解决了无线信号无法直接标注的问题, 从而

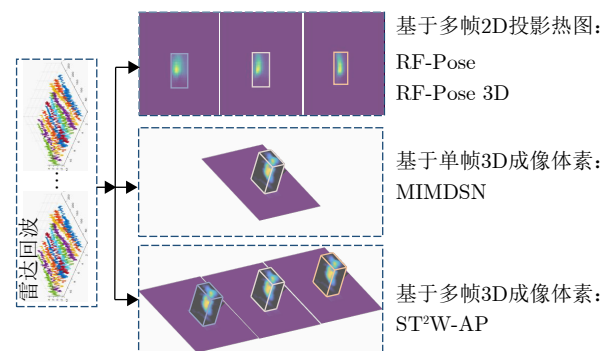


图3 基于成像的人体姿态估计方法

Fig. 3 Radar imaging-based human pose estimation methods

借助深度学习模型重建人体姿态。RF-Pose采用调频连续波雷达,配备水平和垂直双天线阵列,生成对应方向的二维热图。由于单帧射频信号信息稀疏,作者通过融合多帧时域信息,克服这一局限。通过雷达系统与相机的精确同步,平均同步误差小于7 ms,为无线信号提供视觉监督。训练阶段使用视觉信息标注人体关键点位置,监督射频信号。在网络训练完成后,系统能够仅依赖射频信号估计人体姿态,无需视觉输入。

该网络分为教师网络和学生网络两部分,分别处理视觉信号和无线信号。通过教师网络的跨模态监督,解决了数据标签缺失的问题。教师网络采用现有的基于视觉的姿态估计模型,通过相机捕捉图像生成关键点置信度图,监督学生网络的训练。学生网络由编码器和解码器组成,从RF信号中提取人体姿态信息。具体来说,学生网络包含两个独立的RF编码器,分别处理来自垂直和水平天线阵列的热图。每个编码器通过多层卷积神经网络减少空间维度,将热图转化为高维特征表示,涵盖人体整体形状和姿态信息。随后,这些水平和垂直的特征表示在通道维度拼接并传递给解码器。解码器结合时空卷积与反卷积操作,从高维特征中解码出人体姿态。

由于RF信号分辨率低且单帧信息有限,且人体表面会产生镜面反射效应,RF-Pose需依赖多帧数据进行预测,聚合人体运动的动态信息。这一设计使得网络能够捕捉不同时间点的人体各部位,克服射频信号低分辨率和镜面反射问题,生成更准确的关键点置信度图。

总的来说,RF-Pose通过跨模态学习,首次实现了仅凭RF信号预测人体关键点。即使在遮挡或墙壁阻隔的情况下,系统仍能较为准确地估计人体姿态。然而,RF-Pose仅限于二维姿态估计,无法进行三维姿态预测,限制了其适用性。为解决这一问题,团队随后推出了RF-Pose 3D^[36]。

4.1.2 穿墙雷达中的3D姿态估计: RF-Pose 3D

Zhao等人^[36]进一步将二维姿态估计扩展为三维姿态估计,提出了RF-Pose 3D,首次通过无线信号获取3D关节和骨骼信息。该系统采用多天线频率调制连续波雷达,配备水平和垂直放置的天线阵列,工作频段介于5.46~7.24 GHz,较低的频率使其具备较好的穿透障碍物能力。相机系统由12个相机节点组成,负责从不同视角拍摄2D图像,通过OpenPose^[44]生成2D骨架,并利用多视角几何算法将2D骨架转换为3D骨架标签。相机系统仅在训练

阶段用于生成标签数据,网络训练完成后,系统能够仅依赖无线信号实现3D关节重建。

网络部分包含3个核心模块:特征网络(Feature Network, FN)、区域建议网络(Region Proposal Network, RPN)和姿态估计网络(PEN)。FN负责去除杂波并提取目标信息。雷达信号本质上是一个四维时空张量,由于4D CNN对计算资源的需求较高且复杂,常见的深度学习框架(如PyTorch, TensorFlow)尚不支持4D卷积操作,作者通过将四维时空张量分别投影为两个三维张量,简化为3D卷积操作,显著降低计算成本。RPN模块旨在生成每个个体的候选区域,然后裁剪出个体的特征。PEN模块是系统的核心部分,从RPN生成的候选区域中进一步推断人体的14个关键点(如头部、肩膀、手肘、膝盖等)。PEN通过卷积网络在空间和时间维度上聚合特征,不仅推断出被遮挡的关键点,还使系统能够持续跟踪人体的动态变化,生成连续的3D骨架。

RF-Pose 3D通过将复杂的4D卷积问题简化为3D卷积,并设计了不同的模块,以应对多人和复杂环境下的姿态估计挑战。FN模块负责从原始成像热图中提取高维特征,RPN模块处理多人场景中的信号干扰,PEN模块则专注于最终的姿态估计。通过分层次的处理,RF-Pose 3D成功实现了在复杂环境下的多人3D姿态估计。

4.2 基于3D成像体素的姿态估计

4.2.1 基于单帧3D体素的姿态估计: MIMDSN

由于射频信号的物理特性,如成像分辨率低和镜面效应,基于RF的姿态估计过程具有高度的不稳定性。Zheng等人^[37]设计了一种分辨率引导的姿态估计网络,称为互信息最大化深度监督网络(MIMDSN),用于从雷达3D成像体素中估计人体姿态。与RF-Pose和RF-Pose 3D不同,MIMDSN的跨模态监督不仅在网络输出上进行,还应用于隐藏层的特征提取,缓解了将穿墙雷达(TWR)图像转换为三维人体骨架的不稳定性问题。MIMDSN通过4发8收的二维平面阵列捕获三维空间信息,工作频率中心1 GHz,带宽1 GHz。慢时间维度的扫描速率为3 FPS,快时间维度的采样速率为16 GHz,同时,彩色深度(Red, Green, Blue-Depth, RGB-D)相机提供14个骨架关键点的3D位置信息。

MIMDSN网络设计充分考虑了雷达成像的物理特性和分辨率限制。具体而言,MIMDSN通过堆叠多个卷积层逐步提取雷达图像中的高级语义信

息。网络的卷积层数量、卷积核大小及步幅可以根据3D成像像素的空间分辨率进行调整, 以确保在应对不同目标位置和距离时, 能够提取有效特征。

为提高特征的区分性, MIMDSN引入了基于互信息最大化的深度监督学习范式。这种机制不仅在输出层对预测结果进行监督, 还在隐藏层最大化特征与伪标签之间的互信息, 从而使网络能够在各层级上学习到更加丰富和有意义的特征, 提升整体性能。此外, MIMDSN提出了一种自适应损失重加权策略, 用于动态调整隐藏层的损失权重, 避免网络在训练过程中对某些层过度关注或忽略。这一机制不仅提高了训练的稳定性, 还加快了网络的收敛速度。

通过结合分辨率引导的特征提取、深度监督学习以及自适应损失重加权策略, MIMDSN有效解决了TWR系统中低分辨率、遮挡和多目标等复杂场景中的问题, 显著提升了3D人体姿态估计的准确性和鲁棒性。

4.2.2 基于多帧3D体素的姿态估计: ST²W-AP

由于主流机器学习平台(如PyTorch和TensorFlow)仅支持3D卷积, 现有系统通常将多帧二维投影热图或单帧三维成像像素作为输入。然而, 这种方式无法同时利用在行为识别中至关重要的多帧时域信息, 以及在细粒度姿态估计中关键的三维空间信息。张锐等人^[38]提出了ST²W-AP时空融合网络, 通过充分结合多帧3D体素的时空信息, 在保留三维空间信息的同时有效利用长序列时域信息, 实现了显著的性能提升。

ST²W-AP搭建的多发多收(Multiple Input Multiple Output, MIMO)雷达系统包括一个矢量网络分析仪、12个发射天线和12个接收天线, 发射0.8~2.8 GHz的步进频率连续波, 使雷达系统的距离分辨率达到约75 mm。凭借低频段、宽带宽和高密度虚拟天线阵列, 系统不仅具备较强的穿透能力, 还实现了较高的三维空间分辨率。配备的12个摄像头节点与雷达系统可实现毫秒级同步。

在模型设计上, ST²W-AP采用了先空间后时间的特征提取策略, 设计了一种伪4D残差块, 将4D卷积分解为3D空间卷积(用于编码空间信息)和1D时间卷积(用于提取时域信息), 并将其集成到深度残差网络中, 用于从4D体素中提取通用特征, 支持姿态估计和行为识别任务。如图4所示, 混凝土墙会使得信号强度衰减和传播路径折射, 进而导致穿墙信号延迟和衰减, 进而影响成像质量, 削弱人体感知性能。为此, ST²W-AP引入了深度回波

域补偿器, 加速传统墙体抑制中的补偿过程, 显著降低了时间成本。

ST²W-AP首次在隔墙感知中实现了基于多帧3D体素的人体姿态恢复与行为识别。通过结合三维空间信息与长序列时域信息, 该网络大幅提升了姿态估计和行为识别任务的性能。

4.2.3 基于多帧2D成像热图和单帧3D体素的姿态估计: Dual-task Net

Dual-task Net^[39]是一种结合了姿态估计与行为识别的双分支多阶段的任务框架。作者同时利用多帧2D成像热图和单帧3D体素, 将姿态估计与行为识别两个任务整合处理。在数据处理上, 首先采用移动目标指示技术对接收到的雷达信号进行预处理, 以去除静态背景噪声。接下来, 通过后向投影算法进行2D成像, 再结合恒虚警率检测和跟踪算法锁定人体目标, 生成高分辨率的3D体素, 为后续的姿态重建和动作识别任务提供输入。Dual-task Net网络主要由两个关键模块构成: 姿态重建网络和动作识别网络。姿态重建网络基于3D卷积神经网络, 从输入的单帧3D体素中提取特征, 预测人体姿态的3D关节坐标。通过这种方式, 能够在不同场景下精确地重建人体姿态。动作识别网络则采用双分支结构: 一支分支基于多帧2D成像热图, 另一支分支基于姿态重建网络输出的多帧3D姿态。2D成像热图分支通过连续帧的2D成像数据提取人体运动特征, 而3D姿态分支通过连续帧的姿态矩阵来捕捉人体的动态信息。最后, 两个分支的输出结果融合, 生成最终的动作识别结果。这种双任务网络不仅共享了部分计算资源, 还通过互约束的方式提升了两个任务的精度。姿态重建任务通过端到端

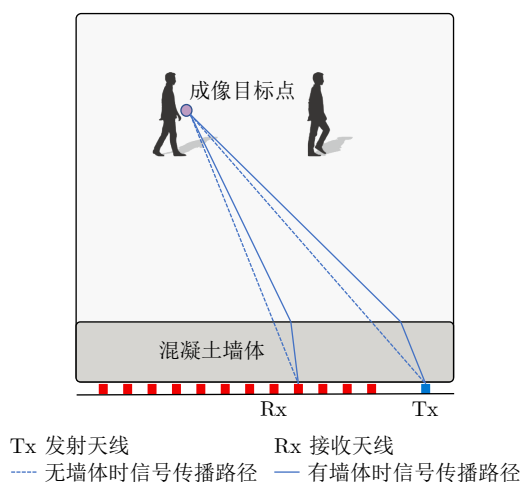


图4 混凝土墙体对于信号传播路径的影响

Fig. 4 The impact of concrete walls on signal propagation paths

的方式精确预测人体姿态,动作识别任务则通过捕捉多帧数据中的时空特征实现高准确率的动作分类。

4.3 基于雷达回波的姿态估计: RadarFormer

传统的基于无线信号的人体姿态估计方法通常先将雷达回波转换为雷达图像,再通过神经网络提取特征以恢复人体姿态信息。Zheng等人^[40]提出了一种新方法RadarFormer,首次引入自注意力机制,直接从雷达回波中实现人体感知,跳过了传统成像算法,实现了端到端的信号处理。

雷达系统采用了由4个发射天线和8个接收天线组成的平面阿基米德螺旋天线阵列,形成了32个等效的发射接收通道。天线阵列设计确保了系统在方位角、高度和对角方向上具有较高的分辨率,能够同时捕捉人体的三维信息。雷达使用伪随机编码的脉冲信号,中心频率为1 GHz,带宽为1 GHz,发射功率为30 dBm,扫描速率为3 帧/s。目标的彩色(Red, Green, Blue, RGB)视频则通过光学相机采集,为定量评估提供标签数据,雷达和相机的时间同步误差小于1 μ s。

RadarFormer基于类似Transformer结构,直接从回波中提取全局特征,网络结构由编码器和解码器组成,编码器部分设计了慢/快时间自注意力机制,分别处理多帧连续回波中的时域信息和单帧回波中的空间信息。慢时间自注意力机制从多帧数据中提取时序特征,而快时间自注意力机制则通过单帧回波捕获空域特征,这一设计增强了网络对目标的全局语义理解。解码器则通过可学习的查询矩阵感知所有可能的目标。结合交叉注意力机制,解码器将查询向量与编码器的特征相匹配,逐步完善每个目标的特征表示。为了处理多目标任务,解码器堆叠了多个自注意力和交叉注意力模块,生成针对不同任务的特征表示。最终,RadarFormer为每个任务(如姿态估计、活动识别和身份重识别)使用专用的多层感知机生成预测结果。

实验表明,RadarFormer在多个任务上的表现优于传统的基于CNN和RNN的雷达感知方法,尤其在计算复杂度和内存占用上具有显著优势。通过直接处理雷达回波信号,RadarFormer克服了传统方法诸多局限,为穿墙雷达感知任务提供了高效、精准的解决方案。

4.4 基于多普勒信息的人体姿态重建: 空时分步增量估计网络

除了基于雷达成像的姿态估计研究,基于运动目标回波产生的微多普勒效应进行姿态估计是一种新颖的、具有广阔前景的方法。人体运动中不同部

位速度不同会带来载频中心附近的多普勒频移,通过瞬时频率分析方法,可以提取躯干和四肢的微多普勒频移,从而进行细粒度感知任务^[61]。李柯蒙等人^[62]提出一种单通道超宽带雷达人体姿态增量估计方案,为采用该技术路线的代表性工作。雷达系统采用了S波段单通道超宽带雷达,基频为2.5 GHz,工作带宽为600 MHz。由于其较低的工作频率,该雷达系统能够穿透墙体等障碍物,因此也可以划归为基于穿墙雷达的人体姿态重建方法。通过构造空时分步增量估计网络,采用空域伪3D卷积和时域膨胀卷积的组合策略,分步提取人体运动的微动特征,并将这些特征映射为人体姿态增量。最终,通过与光学传感器提供的初始姿态结合,实现对人体三维姿态的精确估计。该方法的核心是空时分步增量估计网络,包含空域特征展开和时域特征提取两大模块。空域特征通过伪3D卷积层展开,利用空间卷积捕捉人体在不同方向上的运动信息;时域特征则通过膨胀卷积来提取,以更大范围的感受野捕获时间上的变化,增强网络对长时依赖的建模能力。伪3D卷积的引入解决了普通3D卷积参数量大的问题,降低了计算成本,同时保持了对时空特征的捕捉能力。此外,增量估计模块负责解码网络输出的时空特征,将其转换为人体关节点的姿态增量。通过将多帧雷达回波的微动特征映射为关节点的位移量,并结合光学传感器的初始值,网络可以实现对复杂环境下的连续姿态估计。该方法不仅在准确性上优于对比方法,而且其计算效率显著提升,适用于实时应用场景。

4.5 未来展望

短期而言,基于成像热图的方法仍然是主要发展方向。具体而言,基于成像热图由于成像算法的引入,而为系统带来了较强的可解释性,便于模型性能的改进和优化。相比之下,基于雷达回波的姿态估计尽管在设计上更简洁,但性能的提升依赖于大量实验和调试,这使得其迭代和改进速度相对较慢。基于多普勒信息的方法尽管可以通过雷达信号中的一维微多普勒特征推断出三维姿态,在硬件复杂度和功耗方面具有较低的要求,适用于需要高效计算、低功耗的应用场景。长期而言,未来的研究方向将围绕提高精度、增强鲁棒性、扩展应用场景以及解决实际部署中的挑战等方面展开,具体包括:(1)多模态融合与协同感知:更加注重多种传感技术的融合,如结合毫米波雷达、WiFi信号、超声波等多种无线感知技术,以及与光学传感器的协同。这种多模态融合不仅能够提高姿态估计的精度,还能增强系统在复杂环境下的鲁棒性。(2)深

度学习算法的创新: 更加先进的深度学习算法。例如, 自监督学习和半监督学习技术的应用可以减少对大量标注数据的依赖, 使系统能够更好地适应新环境。此外, 神经架构搜索(NAS)、Mamba等技术用于自动优化网络结构, 以平衡计算效率和性能; (3) 硬件-算法端到端协同设计: 根据算法的需求来定制雷达系统的参数, 或者设计专门的硬件加速器来提高特定算法的处理速度, 以最大限度地发挥硬件和软件的潜力, 实现性能的整体优化。

5 基于WiFi的人体姿态重建

相比于传统的光学或雷达传感技术, WiFi信号不仅能在隐私保护场景中提供非侵入式感知, 还因其广泛的普及性和低成本, 成为一种高效且实用的姿态感知手段。此外, WiFi信号具备更强的穿透墙壁和其他障碍物的能力, 使其在遮挡严重的环境中更具优势。基于此, 研究人员已经开发出多种基于WiFi信号的姿态估计方法。具体而言, WiFi信号在人体周围传播时会受到人体动作的干扰, 信号的幅度、相位等特征都会发生变化。通过捕捉这些变化, 并结合深度学习模型, 研究人员能够从WiFi信号中提取关键特征并进行姿态估计, 从而实现对人体运动的精准感知。

本节介绍了3种代表性工作, 包括分别重建人体2D和3D骨骼的Person-in-WiFi^[41]与Person-in-WiFi 3D^[42], 以及3D人体网格重建的DensePose From WiFi^[43]。这些方法之间相互存在联系而又互有区别。具体而言, 上述3种代表性工作在信号处理方面均选择信道状态信息作为信号特征, 这是因为信道状态信息包含丰富的由人体运动产生的信道变化特征。此外, 当前方法均采用信号处理与深度学习相结合的框架, 以及采用视觉信号为无线信号提供跨模态监督。上述方法的不同之处主要在于神经网络结构设计, 网络输出形式, 以及设备部署方式。例如, Person-in-WiFi的网络结构主要基于UNet, 而Person-in-WiFi 3D则基于Transformer进行网络结构设计。此外, Person-in-WiFi 3D采用利用一台设备作为发射器, 三台作为接收器, 部署于房间的4个角落。而DensePose From WiFi则采用了3个WiFi发射器与3个对应的接收器, 以检测并复原人体姿态的密集对应关系。

该类方法的发展趋势整体呈现从低维到高维的趋势。Person-in-WiFi旨在输出二维关键点坐标, Person-in-WiFi 3D旨在输出三维关键点坐标, 而DensePose From WiFi的输出为高密度的人体表面网格分布。与此相对应, 不同方法的适用场景与优劣势也有所不同。例如, 2D输出的Person-in-

WiFi模型效率更高, 但同时对人体姿态表示能力有限, 而输出人体表面网格分布的DensePose From WiFi能够更加精准对人体姿态进行建模, 但同时也对模型的特征提取能力具有更高的要求。

5.1 2D姿态估计: Person-in-WiFi

Wang等人^[41]首次采用WiFi设备实现人体姿态估计。WiFi天线只能接收电磁波振幅的信号, 即接收到的振幅是三维空间的一维投影。因此, WiFi姿态估计的最大挑战在于从一维投影中重建细粒度的空间信息。

作者采用了两组市售的WiFi设备, 分别作为发射和接收端。每组设备包含3个天线, 工作于2.4 GHz频段, 排列方式类似于常见的家用WiFi路由器。系统采集通道状态信息(Channel-State-Information, CSI)数据, 同时与RGB视频流同步, 视频数据用于提供人体关节坐标和分割掩码的标注, 即结合OpenPose生成的关节热力图(Joint Heat-Maps, JHMs)和身体部位亲和场(Part Affinity Fields, PAFs), 为深度学习模型的训练提供监督信息。为增强数据集的多样性和泛化能力, 数据采集在16个不同室内场景中进行, 涵盖了单人及多人活动的复杂场景。

网络设计上, 输入为多对天线组合下不同时间序列采集的CSI数据构成的张量。首先, 对输入数据进行上采样处理, 以增强空间特征的捕捉能力。随后, 通过残差卷积块和U-Net结构进行深度特征提取。网络的输出包括3个主要部分: 人体分割掩码(Segmentation Mask, SM)、关节热力图和身体部位亲和场。人体分割掩码用于识别和分割人体轮廓, 关节热力图则定位人体的关键关节, 而亲和场用于表示关节之间的连接方向和长度。通过多任务学习框架, 模型可以同时输出人体分割和姿态估计的结果。整个网络使用多重损失函数进行训练, 包括二元交叉熵损失(针对SM)和加权的L2损失(针对JHMs和PAFs)。此外, Person-in-WiFi引入了Matthew Weight加权机制, 针对JHMs和PAFs的稀疏性问题进行了优化。Matthew Weight机制根据关节热力图中像素值的大小动态调整优化权重, 确保网络更关注关节区域。

5.2 3D姿态估计: Person-in-WiFi 3D

Person-in-WiFi 3D^[42]是Person-in-WiFi的改进版本, 实现人体3D姿态的估计。然而, 简单地将Person-in-WiFi网络中的2D操作替换为3D操作, 会导致网络的收敛性问题。此外, 替换为3D操作的网络还存在一系列问题, 包括参数量过大、训练

速度缓慢、后处理延迟高、空间分辨率低等,尤其是缺乏端到端的处理流程,导致误差累积。因此,作者提出了一种基于Transformer和DETR的改进版Person-in-WiFi 3D网络,旨在有效解决这些问题,提升网络的整体性能和精度,模型整体框架如图5所示。

在数据采集方面,Person-in-WiFi 3D采用了与前代不同的设备配置,利用一台设备作为发射器,三台作为接收器,部署于房间的4个角落,形成全方位的信号覆盖。为应对WiFi信号传输过程中的噪声、时间延迟和相位漂移等干扰,作者开发了一种名为PhaseFi的相位去噪方法,通过线性变换有效消除了CSI相位中的干扰,显著提升了数据的准确性和后续3D姿态估计的可靠性。

Person-in-WiFi 3D网络主要由3个关键模块构成:WiFi编码器、姿态解码器和精细解码器,旨在解决人体3D姿态估计中由WiFi信号带来的挑战,特别针对传统3D网络在参数量、训练速度、空间分辨率等方面的不足进行了优化设计。首先,WiFi编码器的主要任务是从经过分块处理的WiFi信号中提取全局上下文信息。WiFi信号被离散化为若干token,通过编码器的处理,可从这些token中提取到丰富的时空信息。在姿态解码器中,多个初始化的随机查询与WiFi编码器提取的全局上下文信息进行交互,最终实现人体姿态的预测。这一方法的优势在于,姿态解码器不再受限于逐层卷积提取的局部特征,而是通过全局信息的推断来预测姿态关节点的位置,此外,多查询机制的引入有助于捕捉姿态的多样性和不确定性,增强了网络对复杂姿态的适应能力。精细解码器则专注于进一步优化姿态解码器的输出预测结果。尽管姿态解码器能够初步生成人体姿态,但鉴于WiFi信号的复杂性和环境噪声的干扰,初步预测的姿态往往存在一定的误差。精细解码器通过更精细的注意力机制,有效修

正姿态解码器可能存在的偏差,进一步提升最终姿态估计的精度。

综上所述,Person-in-WiFi 3D通过WiFi编码器、姿态解码器和精细解码器3个模块,形成了一个完整的3D姿态估计流程。WiFi编码器提取全局的时空信息,姿态解码器通过全局交互预测关节点位置,而精细解码器则进一步优化和修正这些预测。这种基于Transformer自注意力机制的架构有效解决了传统3D网络的收敛性、训练效率和空间分辨率问题。

5.3 3D网络估计: DensePose From WiFi

Geng等人^[43]在深化WiFi信号与深度学习架构的融合应用方面,设计了一种基于WiFi信号的3D人体网格估计网络架构,能够利用WiFi信号生成人体表面的UV坐标,从而实现对人体表面的高密度网格重建。在复杂多变的多人场景中,作者采用了3个WiFi发射器与3个对应的接收器,以检测并复原人体姿态的密集对应关系。

在信号处理层面,接收原始CSI信号后,对振幅和相位信号进行了噪声过滤与拟合处理,旨在提高信号的纯净度和信息提取的准确性。随后,处理过的CSI数据通过模态转换网络,从CSI域转换至空间域,这一过程涉及对振幅和相位信号的编码与融合,为后续的深度学习处理提供了结构化的数据输入。通过编码器-解码器网络,这些特征被进一步转换至图像域,为后续的深度学习模型提供了丰富的视觉信息。在图像域中,作者采用了DensePose-RCNN网络架构,该架构使用ResNet-FPN作为主干网络,从图像特征图中提取空间特征,随后将输出送入区域提议网络。对于每一个对齐的提议特征,通过关键点检测头和密集姿态头分别提取关键点特征和人体表面特征,以实现对人体姿态的精准定位与重建。为了更好地整合不同来源的互补信

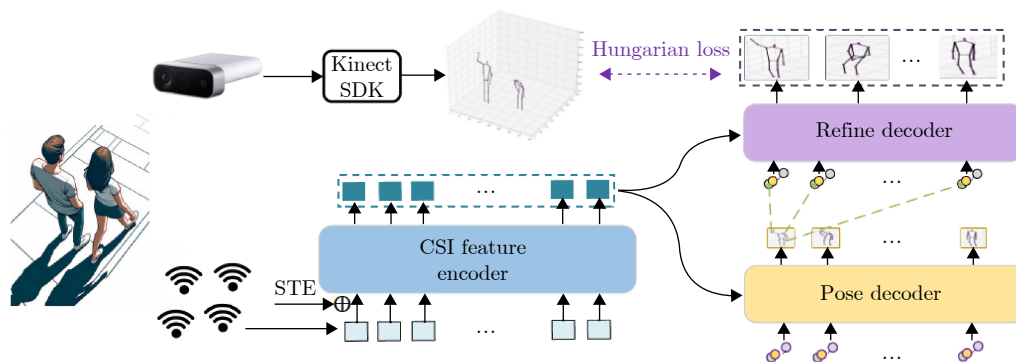


图5 Person-in-WiFi 3D模型框架^[42]

Fig. 5 The framework of Person-in-WiFi 3D^[42]

息, 作者通过refinement单元对两个特征进行融合与重构, 进一步提升了模型的预测精度和鲁棒性。

为了解决训练效率问题和应对标签数据的稀缺性, 作者采取了一种迁移学习策略, 将基于图像的DensePose网络迁移到基于WiFi的网络中。首先, 训练了一个基于图像的DensePose-RCNN模型作为教师网络, 随后, 学生网络由模态转换网络和WiFi-DensePose RCNN组成, 通过最小化学生模型与教师模型生成的多层特征图之间的差异, 实现了基于WiFi信号的DensePose网络的高效训练与性能提升, 缓解了标签数据不足的问题。

5.4 未来展望

尽管当前基于WiFi的人体姿态重建技术已经取得了一定的进展, 该技术的大规模部署仍然需要解决一系列实际问题。

感知与通信的有机结合: 不同于雷达, WiFi原本具备的主要功能为通信。因此, 在使用WiFi进行感知时, 需要确保感知功能不会影响设备原有的通信性能。如何将通信感知有机结合, 在保持通信性能的前提下进行人体姿态重建值得进一步研究。

不同场景的泛化能力: 尽管目前方法在进行实验评估时, 均尝试从多个不同环境中进行数据采集和验证。然而在实际应用中, 可能存在部署的场景千变万化, 包括房间的大小、结构、家具种类、数量、摆放位置等, 均可能对WiFi信号产生影响。因此, 如何更充分有力地验证WiFi人体姿态估计方法在不同场景下的泛化能力仍需进一步探索。

统一的评估标准和数据集: 当前不同方法均只在各方法对应的数据集上进行评估。尽管各方法均能在自己的数据集上取得较好的性能, 如何客观公正地, 在一个统一的数据集和评估标准下, 对各方法的有效性进行比较也是未来的研究方向之一。

6 人体姿态重建数据集

目前, 基于无线信号的人体感知开源数据集较

为稀缺。如表2所示, 本文重点介绍两类代表性的开源数据集: 基于穿墙雷达的人体感知数据集和基于毫米波雷达的人体姿态估计数据集。其中, 基于穿墙人体感知的数据集以国防科技大学的UWB-HA4D-1.0为代表^[63], 利用低频的步进频信号结合三维超宽带MIMO雷达系统采集了无遮挡自由空间、塑料板遮挡、砖墙遮挡下3种复杂场景下不同动作类别的数据。基于毫米波雷达的人体姿态估计则主要包含HIBER^[16], RT-Pose^[64], mRI^[65], mmBody^[66], HuPR^[66] 5个数据集, 利用高频毫米波雷达结合RGB摄像头、深度摄像头、激光雷达、惯性测量单元(Inertial Measurement Unit, IMU)等设备实现多模态数据集采集。

6.1 基于穿墙雷达的人体感知数据集

UWB-HA4D-1.0是国防科技大学金添等人^[63]开源的首个超宽带雷达人体动作四维成像数据集。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, 共计2757组人体动作数据。数据集采集设备为一款三维超宽带MIMO雷达系统, 工作频段为1.78~2.78 GHz的步进频信号, 天线阵列为10发10收, 阵列尺寸为60 cm × 88 cm。由于工作频段较低, 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。由后向投影(Back Projection, BP)算法作为MIMO雷达成像算法。值得注意的是数据集中均为直接成像的方式获取的四维雷达图像, 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。数据采集场景有3个, 分别是无遮挡场景、3 cm塑料板遮挡场景、以及27 cm砖墙遮挡场景。数据集所采集的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。UWB-HA4D-1.0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, 该数据集仍存在

表 2 基于无线信号的人体姿态估计数据集对比

Tab. 2 Summary of dataset on pose estimation based on wireless signals

数据集	无线设备	真值采集设备	场景数量	行为种类	用户数量	总样本数(帧)
UWB-HA4D-1.0	穿墙雷达	RGB	3	10	11	110280
HIBER	毫米波雷达	RGB	10	4	10	402380
RT-Pose	毫米波雷达	RGB LiDAR	40	6	10	72000
mRI	毫米波雷达	RGB-D IMU	1	12	20	160000
mmBody	毫米波雷达	RGB	100	7	20	>20万
HuPR	毫米波雷达	RGB	1	3	6	141000

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以及未对动作的起始帧和结束帧进行标注等问题。

6.2 基于毫米波雷达的人体姿态估计数据集

HIBER (Human Indoor Behavior Exclusive RF Dataset)是一个开源的多模态数据集^[6],用于研究人类行为对射频信号的影响。可以用于人姿势估计、轮廓分割、动作识别和位置跟踪等人体感知任务。数据集总大小为400 GB,包括射频热图、RGB图像、2D/3D人体骨架、边界框以及人体轮廓的真实标注。数据集的组成细节如下。

(1) 10种环境:在同一房间内的10个不同位置放置FMCW雷达。

(2) 10名志愿者:涉及10名年龄、身高和体重各异的用户。

(3) 4种动作:站立、行走、坐下、蹲下。

(4) 3种遮挡:泡沫板遮挡、纸箱遮挡和瑜伽垫遮挡。

(5) 3种标签:2D/3D人体姿态、人体位置的边界框以及人体轮廓掩码。

采用了两台TI MMwave MMWCAS-RF-EVM FMCW雷达,分别水平和竖直放置。每台雷达包含12发16收的MIMO天线阵列,第1台雷达的扫描频率设置为77.00~78.23 GHz,而第2台雷达的扫描频率设置为79.00~80.23 GHz,两者的带宽均为1.23 GHz。多摄像头系统则由13个摄像头节点组成。多摄像头系统主要用于计算人体的三维关键点。检测到三维关键点后,可以通过计算同一人体三维关键点的最小边界立方体,从而确定人体目标的边界框。在遮挡场景中无法获得轮廓掩码真值时,可以通过将三维关节点重投影回被遮挡的成像平面来计算真值,从而得到真实的人体位置和动作。这套摄像头系统还可以提供更多形式的注释,这使得HIBER数据集适用于许多其他无线传感任务。轮廓掩码由Mask R-CNN生成,采用OpenPose生成2D人体关键点,并由三角测量生成相应的3D骨架。雷达和多摄像机系统使用NTP进行同步,可以达到毫秒级同步误差。

RT-Pose (Radar Tensor-Based Human Pose Dataset)^[64]是一个开源的多模态数据集,用于研究人体运动和姿态对毫米波信号的影响。数据集中包含了10名参与者来执行6种不同类型的动作、5种不同的环境和4种不同的遮挡环境:

(1) 6种不同的动作主要分为两类:第1类是参与者进行站立时的动作,包括挥手、抬腿和随机姿势(伸展、弯曲、扭转),每个动作持续3~5 s,来

增加动作的复杂性;第2类是行走过程中的动作,包含行走、行走后坐下、行走时挥手,每个动作持续30 s。

(2) 5种不同的环境包括:杂乱的室内、整齐的室内、下雨天户外、低光照户外、正常户外。

(3) 4种不同的遮挡环境:无遮挡、纸板遮挡、布遮挡、塑料遮挡。

数据采集设备包含两个RGB相机、一个激光雷达和一个毫米波雷达。毫米波雷达配备12发16收的天线阵列,按照MIMO方式以10帧/s的采样率运行,信号频段为77~81 GHz,可提供4.5 cm的距离分辨率,1.4°的方位角分辨率和18°的俯仰角分辨率,每一帧由64个线性调频脉冲组成,提供3.9 cm/s的速度分辨率。其中,RGB相机和激光雷达用来准确标记3D人体骨骼,用来作为网络训练的真实值。RT-Pose数据集是人体姿态估计领域第一个集成3种模态的多模态数据集,并且保证了环境的多样性。

mRI (a Multi-Modal 3D Human Pose Estimation Dataset with mmWave, RGB-D, and Inertial Sensors)^[65]是一个包含毫米波3D点云、RGB帧和深度图、可穿戴IMU传感器的关节旋转和加速度以及对应的2D关键点、3D关节和12个动作标签的开源多模态数据集,用于研究人体运动和姿态对毫米波信号的影响。数据集中包含了13名男性和7名女性来执行12种不同类型的康复动作:左上肢伸展、右上肢伸展、双上肢伸展、左前弓步、右前弓步、下蹲、左侧弓步、右侧弓步、左肢伸展,右肢伸展、自由伸展和放松、直线行走。数据采集设备包含两个Microsoft Kinect传感器用于捕捉RGB和深度帧、一个10 Hz采样率的IWR1443 Boost毫米波雷达用于生成3D毫米波点云、6个可穿戴的IMU用于捕获人体关节的旋转和加速度,分别紧贴在受试者的左手腕、右手腕、左膝盖、右膝盖、头部和骨盆上,以捕捉完整的人体信息。

mmBody (mmWave 3D Human Body Dataset)^[66]包含7个不同的场景,由20名志愿者采集的100个动作,用于研究如何利用毫米波信号重建3D人体模型。数据集包含3部分:毫米波雷达用于捕捉和记录身体动态、MoCap系统用于标记人体关节位置和全身网格、相机系统用于采集RGB-D图像。其中,毫米波雷达采用Arbe Robotics生产的Phoenix型号,配备48个发射通道和48个接收通道,实现0.4 m的距离分辨率和2.0°的角度分辨率,工作速率为14 FPS,并通过板载处理器将原始雷达信号转换为点云,每一个点都包含其3D位置、距离速度、幅度

以及场景中相应点的反射信号能量。相机采集系统由两个Azure Kinect组成,以30 FPS的速度提供彩色图像和深度图像。MoCap系统由8个摄像头和附着在人体上的标记组成,用来生成3D身体骨架和全身的网格。mmBody数据集涵盖20名志愿者在7个不同环境中的100个动作。在这100个动作中,有16个静态姿势、9个躯干动作、20个腿部动作、25个手臂动作、3个颈部动作、14个运动动作、7个日常室内动作和6个厨房动作。20名志愿者中,女性10名,男性10名,体重范围42~75 kg,身高范围159~183 cm。

HuPR (Human Pose with Millimeter Wave Radar Dataset)^[67]包含两个毫米波雷达以及一台RGB摄像头来生成2D人体姿势关键点的数据集。其中,两块TI IWR1843BOOST毫米波雷达垂直放置,并通过配置1个虚拟发射天线和8个虚拟接收天线来获得15°较高分辨率的俯仰角和方位角。此外,RGB摄像头将采集的目标帧基于图像的2D人体姿势估计网络HRNet来生成2D人体关键点作为后期训练的真值。数据集为在室内环境中采集,每个视频帧包含RGB相机帧、水平雷达帧和垂直雷达帧,采样率均为10 FPS。目标场景中只包含一个人,执行3种类型的运动:静止动作、站立挥手和挥手行走。数据集包含6名测试者。

7 总结与展望

尽管基于无线信号的人体姿态估计技术已经取得了显著的进步,但相关技术距离大规模实际应用与推广仍存在一定差距。以下是具体的技术挑战与未来研究方向。

7.1 技术挑战

(1) 如何进一步增强系统泛化性: 尽管当前方法均尝试在不同场景下采集数据并进行评估,由于不同房间大小、结构、家具均有所不同,其场景数量仍远远小于实际可能部署的场景数量。因此,当前技术面临的主要技术挑战之一便是如何进一步增强系统泛化性。此外,毫米波雷达、穿墙雷达以及WiFi设备的不同信号特性,导致其在提升泛化性上面临的挑战也有所不同。毫米波信号较短的波长通常意味着成像分辨率较高,因此能够较好地区分不同场景下的不同结构,从而可采用一系列空域滤波方法减弱环境的影响。低频段的穿墙雷达成像能力有所下降,但仍可以通过大带宽多天线的设备部署方式进行弥补。而对于低频段、小带宽、少天线的WiFi设备而言,其较低的物理分辨率导致难以区分环境中的不同目标,从而在提升不同场景泛化性方面具

有更大的挑战。因此,在系统设计时需要针对不同信号特性来设计信号处理和深度学习算法,从而增强系统在不同场景下的泛化性能。

(2) 如何提升信号的物理分辨率: 如上所述,信号的物理分辨率对于系统的成像能力和泛化能力均有至关重要的影响。尽管现有方法大多采用深度学习算法在一定程度上弥补信号分辨率不足的缺点,如何从本质上提升信号的物理分辨率仍值得关注且具备较大的技术难度。对于雷达设备而言,通过增加带宽和天线数量能够提升距离和角度分辨率。但更多的天线往往也意味着更高的系统成本以及更大的开发难度。对于WiFi而言,其带宽和天线数量往往已经由通信功能所决定,意味着在设计感知功能时也需要考虑到原有通信功能的需求。

(3) 如何统一评估各方法有效性: 基于图像的人体姿态估计已经建立起了数个大型的开源数据集,其中不少已经成为该任务的标准评估数据集,以供各个方法进行公正的性能评估。然而,在无线人体姿态估计领域,各个方法目前仍只关注在该研究本身采集的数据集上进行验证。因此,如何建立一个公开的大型标准无线数据集,从而统一地评估各方法的有效性也是尚未解决的技术挑战之一。

7.2 未来展望

未来的研究除了在解决上述技术挑战以外,还应当感知框架和场景上持续创新,重点关注例如多传感器融合感知、多任务协同感知,以及多场景通用感知等。

(1) 多传感器融合感知: 未来的研究方向不仅应集中于单一技术的优化,还应探索多传感器融合感知的可能性。通过结合雷达信号和其他感知媒介(例如视觉、惯性测量单元以及声学)实现更加准确和鲁棒的融合感知。一方面,其他介质的感知技术可以分担雷达感知的部分功能,从而提升融合感知的效率和能力。例如,文献[68]利用毫米波雷达和相机多模态传感系统来跟踪箱子的位置以及箱子中易碎产品的相对位置。其采用基于视觉的感知算法沿输送带对箱体进行定位,使毫米波雷达能够专注于计算产品的位置。另一方面,通过与其他介质的感知技术相互校正,能够实现更高感知精度的融合感知。例如,文献[69]采用了一种基于毫米波雷达的同时定位和制图解决方案,并辅以IMU。其利用IMU数据将连续雷达扫描点云组合成“多扫描”,以获得准确、鲁棒的SLAM结果。

(2) 多任务协同感知: 多任务学习技术^[70]在过去十年中取得了一定的进步,因此未来人体姿态估计也可与其他任务相结合,实现多任务协同感知。

例如,姿态估计和生命体征都是智能家居健康监测的重要组成部分。如果可用一个毫米波雷达同时实现这两项感知任务,则可简化智能家居的部署和实施。然而,多任务感知并不是多个单任务感知的简单叠加。人类的姿态和生命信号是密切相关的,如何描述和分析反射信号,利用多个感知任务之间的关系存在挑战。

(3) 多场景通用感知:随着无线通信技术的不断革新,如WiFi 6、毫米波通信和更先进的雷达技术将进一步提升人体姿态估计的精度和适应性。同时,随着计算能力的提升和深度学习算法的不断优化,人体姿态估计技术有望在实时性和大规模应用方面取得更大的突破,从而推动其在智能城市、医疗健康、虚拟现实等领域的广泛应用。例如,在智能城市中,相关技术可以用于智慧交通管理和公共安全监控;在医疗领域,基于无线信号的姿态估计可提供非接触式的患者监控和健康管理;在虚拟现实,无线姿态估计将带来更自然的动作捕捉和交互体验,提升用户的沉浸感。

声明:陈彦和张锐对该文具有相同的贡献。

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] ZHAO Zhongqiu, ZHENG Peng, XU Shoutao, *et al.* Object detection with deep learning: A review[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(11): 3212–3232. doi: [10.1109/TNNLS.2018.2876865](https://doi.org/10.1109/TNNLS.2018.2876865).
- [2] CHEN Yucheng, TIAN Yingli, and HE Mingyi. Monocular human pose estimation: A survey of deep learning-based methods[J]. *Computer Vision and Image Understanding*, 2020, 192: 102897. doi: [10.1016/j.cviu.2019.102897](https://doi.org/10.1016/j.cviu.2019.102897).
- [3] MUNEA T L, JEMBRE Y Z, WELDEGEBRIEL H T, *et al.* The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation[J]. *IEEE Access*, 2020, 8: 133330–133348. doi: [10.1109/ACCESS.2020.3010248](https://doi.org/10.1109/ACCESS.2020.3010248).
- [4] JIAO Licheng, ZHANG Ruohan, LIU Fang, *et al.* New generation deep learning for video object detection: A survey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(8): 3195–3215. doi: [10.1109/TNNLS.2021.3053249](https://doi.org/10.1109/TNNLS.2021.3053249).
- [5] 杨小鹏, 高炜程, 渠晓东. 基于微多普勒角点特征与Non-Local机制的穿墙雷达人体步态异常终止行为辨识技术[J]. *雷达学报(中英文)*, 2024, 13(1): 68–86. doi: [10.12000/JR23181](https://doi.org/10.12000/JR23181).
- [6] YANG Xiaopeng, GAO Weicheng, and QU Xiaodong. Human anomalous gait termination recognition via through-the-wall radar based on micro-Doppler corner features and Non-Local mechanism[J]. *Journal of Radars*, 2024, 13(1): 68–86. doi: [10.12000/JR23181](https://doi.org/10.12000/JR23181).
- [6] 金添, 宋勇平, 崔国龙, 等. 低频电磁波建筑物内部结构透视技术研究进展[J]. *雷达学报*, 2020, 10(3): 342–359. doi: [10.12000/JR20119](https://doi.org/10.12000/JR20119).
- [7] JIN Tian, SONG Yongping, CUI Guolong, *et al.* Advances on penetrating imaging of building layout technique using low frequency radio waves[J]. *Journal of Radars*, 2021, 10(3): 342–359. doi: [10.12000/JR20119](https://doi.org/10.12000/JR20119).
- [7] 崔国龙, 余显祥, 魏文强, 等. 认知智能雷达抗干扰技术综述与展望[J]. *雷达学报*, 2022, 11(6): 974–1002. doi: [10.12000/JR22191](https://doi.org/10.12000/JR22191).
- [8] CUI Guolong, YU Xianxiang, WEI Wenqiang, *et al.* An overview of antijamming methods and future works on cognitive intelligent radar[J]. *Journal of Radars*, 2022, 11(6): 974–1002. doi: [10.12000/JR22191](https://doi.org/10.12000/JR22191).
- [8] 夏正欢, 张群英, 叶盛波, 等. 一种便携式伪随机编码超宽带人体感知雷达设计[J]. *雷达学报*, 2015, 4(5): 527–537. doi: [10.12000/JR15027](https://doi.org/10.12000/JR15027).
- [9] XIA Zhenghuan, ZHANG Qunying, YE Shengbo, *et al.* Design of a handheld pseudo random coded UWB radar for human sensing[J]. *Journal of Radars*, 2015, 4(5): 527–537. doi: [10.12000/JR15027](https://doi.org/10.12000/JR15027).
- [9] ZHANG Dongheng, HU Yang, and CHEN Yan. MTrack: Tracking multiperson moving trajectories and vital signs with radio signals[J]. *IEEE Internet of Things Journal*, 2021, 8(5): 3904–3914. doi: [10.1109/JIOT.2020.3025820](https://doi.org/10.1109/JIOT.2020.3025820).
- [10] LI Yadong, ZHANG Dongheng, CHEN Jinbo, *et al.* Towards domain-independent and real-time gesture recognition using mmWave signal[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(12): 7355–7369. doi: [10.1109/TMC.2022.3207570](https://doi.org/10.1109/TMC.2022.3207570).
- [11] ZHANG Binbin, ZHANG Dongheng, LI Yadong, *et al.* Unsupervised domain adaptation for RF-based gesture recognition[J]. *IEEE Internet of Things Journal*, 2023, 10(23): 21026–21038. doi: [10.1109/JIOT.2023.3284496](https://doi.org/10.1109/JIOT.2023.3284496).
- [12] SONG Ruiyuan, ZHANG Dongheng, WU Zhi, *et al.* RF-URL: Unsupervised representation learning for RF sensing[C]. The 28th Annual International Conference on Mobile Computing and Networking, Sydney, Australia, 2022: 282–295. doi: [10.1145/3495243.3560529](https://doi.org/10.1145/3495243.3560529).
- [13] GONG Hanqin, ZHANG Dongheng, CHEN Jinbo, *et al.* Enabling orientation-free mmwave-based vital sign sensing with multi-domain signal analysis[C]. 2024 IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, Korea, Republic of, 2024: 8751–8755. doi: [10.1109/ICASSP48605.2024.8751](https://doi.org/10.1109/ICASSP48605.2024.8751).

- 10.1109/ICASSP48485.2024.10448323.
- [14] XIE Chunyang, ZHANG Dongheng, WU Zhi, *et al.* RPM 2.0: RF-based pose machines for multi-person 3D pose estimation[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(1): 490–503. doi: [10.1109/TCSVT.2023.3287329](https://doi.org/10.1109/TCSVT.2023.3287329).
- [15] YANG Shuai, ZHANG Dongheng, SONG Ruiyuan, *et al.* Multiple WiFi access points co-localization through joint AoA estimation[J]. *IEEE Transactions on Mobile Computing*, 2024, 23(2): 1488–1502. doi: [10.1109/TMC.2023.3239377](https://doi.org/10.1109/TMC.2023.3239377).
- [16] WU Zhi, ZHANG Dongheng, XIE Chunyang, *et al.* RFMask: A simple baseline for human silhouette segmentation with radio signals[J]. *IEEE Transactions on Multimedia*, 2023, 25: 4730–4741. doi: [10.1109/TMM.2022.3181455](https://doi.org/10.1109/TMM.2022.3181455).
- [17] GENG Ruixu, HU Yang, LU Zhi, *et al.* Passive non-line-of-sight imaging using optimal transport[J]. *IEEE Transactions on Image Processing*, 2022, 31: 110–124. doi: [10.1109/TIP.2021.3128312](https://doi.org/10.1109/TIP.2021.3128312).
- [18] XIE Chunyang, ZHANG Dongheng, WU Zhi, *et al.* RPM: RF-based pose machines[J]. *IEEE Transactions on Multimedia*, 2024, 26: 637–649. doi: [10.1109/TMM.2023.3268376](https://doi.org/10.1109/TMM.2023.3268376).
- [19] YU Cong, ZHANG Dongheng, WU Zhi, *et al.* MobiRFPose: Portable RF-based 3D human pose camera[J]. *IEEE Transactions on Multimedia*, 2024, 26: 3715–3727. doi: [10.1109/TMM.2023.3314979](https://doi.org/10.1109/TMM.2023.3314979).
- [20] YU Cong, ZHANG Dongheng, WU Zhi, *et al.* Fast 3D human pose estimation using RF signals[C]. 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, Rhodes Island, Greece, 2023: 1–5. doi: [10.1109/ICASSP49357.2023.10094778](https://doi.org/10.1109/ICASSP49357.2023.10094778).
- [21] MU Kangle, LUAN T H, ZHU Lina, *et al.* A survey of handy see-through wall technology[J]. *IEEE Access*, 2020, 8: 82951–82971. doi: [10.1109/ACCESS.2020.2991201](https://doi.org/10.1109/ACCESS.2020.2991201).
- [22] 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](https://doi.org/10.3390/rs13020241).
- [23] VASISHT D, JAIN A, HSU C Y, *et al.* Duet: Estimating user position and identity in smart homes using intermittent and incomplete RF-data[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 2(2): 84. doi: [10.1145/3214287](https://doi.org/10.1145/3214287).
- [24] HSU C Y, HRISTOV R, LEE G H, *et al.* Enabling identification and behavioral sensing in homes using radio reflections[C]. 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland UK, 2019: 548. doi: [10.1145/3290605.3300778](https://doi.org/10.1145/3290605.3300778).
- [25] FAN Lijie, LI Tianhong, YUAN Yuan, *et al.* In-home daily-life captioning using radio signals[C]. The 16th European Conference on Computer Vision, Glasgow, UK, 2020: 105–123. doi: [10.1007/978-3-030-58536-5_7](https://doi.org/10.1007/978-3-030-58536-5_7).
- [26] TIAN Yonglong, LEE G H, HE Hao, *et al.* RF-based fall monitoring using convolutional neural networks[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 2(3): 137. doi: [10.1145/3264947](https://doi.org/10.1145/3264947).
- [27] AYYALASOMAYAJULA R, ARUN A, WU Chenfeng, *et al.* Deep learning based wireless localization for indoor navigation[C]. The 26th Annual International Conference on Mobile Computing and Networking, London, United Kingdom, 2020: 17. doi: [10.1145/3372224.3380894](https://doi.org/10.1145/3372224.3380894).
- [28] CAO Zhongping, DING Wen, CHEN Rihui, *et al.* A joint global-local network for human pose estimation with millimeter wave radar[J]. *IEEE Internet of Things Journal*, 2023, 10(1): 434–446. doi: [10.1109/JIOT.2022.3201005](https://doi.org/10.1109/JIOT.2022.3201005).
- [29] SENGUPTA A, JIN Feng, ZHANG Renyuan, *et al.* mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs[J]. *IEEE Sensors Journal*, 2020, 20(17): 10032–10044. doi: [10.1109/JSEN.2020.2991741](https://doi.org/10.1109/JSEN.2020.2991741).
- [30] ADIB F, HSU C Y, MAO Hongzi, *et al.* Capturing the human figure through a wall[J]. *ACM Transactions on Graphics*, 2015, 34(6): 219. doi: [10.1145/2816795.2818072](https://doi.org/10.1145/2816795.2818072).
- [31] AHMAD F, ZHANG Yimin, and AMIN M G. Three-dimensional wideband beamforming for imaging through a single wall[J]. *IEEE Geoscience and Remote Sensing Letters*, 2008, 5(2): 176–179. doi: [10.1109/LGRS.2008.915742](https://doi.org/10.1109/LGRS.2008.915742).
- [32] KONG Lingjiang, CUI Guolong, YANG Xiaobo, *et al.* Three-dimensional human imaging for through-the-wall radar[C]. 2009 IEEE Radar Conference, Pasadena, USA, 2009: 1–4. doi: [10.1109/RADAR.2009.4976932](https://doi.org/10.1109/RADAR.2009.4976932).
- [33] HOLL P M and REINHARD F. Holography of Wi-Fi radiation[J]. *Physical Review Letters*, 2017, 118(18): 183901. doi: [10.1103/PhysRevLett.118.183901](https://doi.org/10.1103/PhysRevLett.118.183901).
- [34] ZHAO Mingmin, LI Tianhong, ABU ALSHEIKH M, *et al.* Through-wall human pose estimation using radio signals[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7356–7365. doi: [10.1109/CVPR.2018.00768](https://doi.org/10.1109/CVPR.2018.00768).
- [35] JIANG Wenjun, XUE Hongfei, MIAO Chenglin, *et al.* Towards 3D human pose construction using WiFi[C]. The 26th Annual International Conference on Mobile Computing and Networking, London, United Kingdom, 2020: 23. doi: [10.1145/3372224.3380900](https://doi.org/10.1145/3372224.3380900).
- [36] ZHAO Mingmin, TIAN Yonglong, ZHAO Hang, *et al.* RF-

- based 3D skeletons[C]. 2018 Conference of the ACM Special Interest Group on Data Communication, Budapest, Hungary, 2018: 267–281. doi: [10.1145/3230543.3230579](https://doi.org/10.1145/3230543.3230579).
- [37] ZHENG Zhijie, PAN Jun, ZHANG Diankun, *et al.* Through-wall human pose estimation by mutual information maximizing deeply supervised nets[J]. *IEEE Internet of Things Journal*, 2024, 11(2): 3190–3205. doi: [10.1109/JIOT.2023.3294955](https://doi.org/10.1109/JIOT.2023.3294955).
- [38] 张锐, 龚汉钦, 宋瑞源, 等. 基于4D成像雷达的隔墙人体姿态重建与行为识别研究[J]. 雷达学报(中英文), 2025, 14(1): 44–61. doi: [10.12000/JR24132](https://doi.org/10.12000/JR24132).
ZHANG Rui, GONG Hanqin, SONG Ruiyuan, *et al.* Through-wall human pose reconstruction and action recognition using four-dimensional imaging radar[J]. *Journal of Radars*, 2025, 14(1): 44–61. doi: [10.12000/JR24132](https://doi.org/10.12000/JR24132).
- [39] SONG Yongkun, DAI Yongpeng, JIN Tian, *et al.* Dual-task human activity sensing for pose reconstruction and action recognition using 4-D imaging radar[J]. *IEEE Sensors Journal*, 2023, 23(19): 23927–23940. doi: [10.1109/JSEN.2023.3308788](https://doi.org/10.1109/JSEN.2023.3308788).
- [40] ZHENG Zhijie, ZHANG Diankun, LIANG Xiao, *et al.* RadarFormer: End-to-end human perception with through-wall radar and transformers[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 35(10): 4319–4332. doi: [10.1109/TNNLS.2023.3314031](https://doi.org/10.1109/TNNLS.2023.3314031).
- [41] WANG Fei, ZHOU Sanping, PANEV S, *et al.* Person-in-WiFi: Fine-grained person perception using WiFi[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 5452–5461. doi: [10.1109/ICCV.2019.00555](https://doi.org/10.1109/ICCV.2019.00555).
- [42] YAN Kangwei, WANG Fei, QIAN Bo, *et al.* Person-in-WiFi 3D: End-to-end multi-person 3D pose estimation with WiFi[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2024: 969–978. doi: [10.1109/CVPR52733.2024.00098](https://doi.org/10.1109/CVPR52733.2024.00098).
- [43] GENG Jiaqi, HUANG Dong, and DE LA TORRE F. DensePose from WiFi[OL]. <https://arxiv.org/abs/2301.00250>. 2022.
- [44] CAO Zhe, SIMON T, WEI S E, *et al.* Realtime multi-person 2D pose estimation using part affinity fields[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7291–7299. doi: [10.1109/CVPR.2017.143](https://doi.org/10.1109/CVPR.2017.143).
- [45] JOHNSON S and EVERINGHAM M. Clustered pose and nonlinear appearance models for human pose estimation[C]. 2010 British Machine Vision Conference, Aberystwyth, UK, 2010: 1–11.
- [46] CHEN Xianjie and YUILLE A. Parsing occluded people by flexible compositions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3945–3954. doi: [10.1109/CVPR.2015.7299020](https://doi.org/10.1109/CVPR.2015.7299020).
- [47] LI Wenbo, WANG Zhicheng, YIN Binyi, *et al.* Rethinking on multi-stage networks for human pose estimation[OL]. <https://arxiv.org/abs/1901.00148>. 2019.
- [48] BOULIC R, THALMANN N M, and THALMANN D. A global human walking model with real-time kinematic personification[J]. *The Visual Computer*, 1990, 6(6): 344–358. doi: [10.1007/BF01901021](https://doi.org/10.1007/BF01901021).
- [49] BOULIC R, REZZONICO S, and THALMANN D. Multi-finger manipulation of virtual objects[C]. ACM Symposium on Virtual Reality Software and Technology, Hong Kong, China, 1996: 67–74. doi: [10.1145/3304181.3304195](https://doi.org/10.1145/3304181.3304195).
- [50] JU S X, BLACK M J, and YACOOB Y. Cardboard people: A parameterized model of articulated image motion[C]. The 2nd International Conference on Automatic Face and Gesture Recognition, Killington, USA, 1996: 38–44. doi: [10.1109/AFGR.1996.557241](https://doi.org/10.1109/AFGR.1996.557241).
- [51] JIANG Hao. Finding human poses in videos using concurrent matching and segmentation[C]. The 10th Asian Conference on Computer Vision, Queenstown, New Zealand, 2011: 228–243. doi: [10.1007/978-3-642-19315-6_18](https://doi.org/10.1007/978-3-642-19315-6_18).
- [52] COOTES T F, TAYLOR C J, COOPER D H, *et al.* Active shape models-their training and application[J]. *Computer Vision and Image Understanding*, 1995, 61(1): 38–59. doi: [10.1006/cviu.1995.1004](https://doi.org/10.1006/cviu.1995.1004).
- [53] FREIFELD O, WEISS A, ZUFFI S, *et al.* Contour people: A parameterized model of 2D articulated human shape[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 639–646. doi: [10.1109/CVPR.2010.5540154](https://doi.org/10.1109/CVPR.2010.5540154).
- [54] URTASUN R and FUA P. 3D human body tracking using deterministic temporal motion models[C]. The 8th European Conference on Computer Vision, Prague, Czech Republic, 2004: 92–106. doi: [10.1007/978-3-540-24672-5_8](https://doi.org/10.1007/978-3-540-24672-5_8).
- [55] LOPER M, MAHMOOD N, ROMERO J, *et al.* SMPL: A skinned multi-person linear model[J]. *Seminal Graphics Papers: Pushing the Boundaries*, 2023, 2: 88. doi: [10.1145/3596711.3596800](https://doi.org/10.1145/3596711.3596800).
- [56] SAITO Shunsuke, HUANG Zeng, NATSUME Ryota, *et al.* PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 2304–2314. doi: [10.1109/ICCV.2019.00239](https://doi.org/10.1109/ICCV.2019.00239).
- [57] PONS-MOLL G, ROMERO J, MAHMOOD N, *et al.* Dyna: A model of dynamic human shape in motion[J]. *ACM Transactions on Graphics*, 2015, 34(4): 120. doi: [10.1145/2766993](https://doi.org/10.1145/2766993).
- [58] ZUFFI S and BLACK M J. The stitched puppet: A

- graphical model of 3D human shape and pose[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3537–3546. doi: [10.1109/CVPR.2015.7298976](https://doi.org/10.1109/CVPR.2015.7298976).
- [59] JOO H, SIMON T, and SHEIKH Y. Total capture: A 3D deformation model for tracking faces, hands, and bodies[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8320–8329. doi: [10.1109/CVPR.2018.00868](https://doi.org/10.1109/CVPR.2018.00868).
- [60] XU Hongyi, BAZAVAN E G, ZANFIR A, *et al.* GHUM & GHUML: Generative 3D human shape and articulated pose models[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 6184–6193. doi: [10.1109/CVPR42600.2020.00622](https://doi.org/10.1109/CVPR42600.2020.00622).
- [61] CHEN V C, LI Fayin, HO S S, *et al.* Micro-Doppler effect in radar: Phenomenon, model, and simulation study[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2006, 42(1): 2–21. doi: [10.1109/TAES.2006.1603402](https://doi.org/10.1109/TAES.2006.1603402).
- [62] 李柯蒙, 戴永鹏, 宋勇平, 等. 单通道超宽带雷达人体姿态增量估计技术[J]. 雷达学报(中英文), 2025, 14(1): 16–27. doi: [10.12000/JR24109](https://doi.org/10.12000/JR24109).
LI Kemeng, DAI Yongpeng, SONG Yongping, *et al.* Single-channel ultrawideband radar human pose-incremental estimation technology[J]. *Journal of Radars*, 2025, 14(1): 16–27. doi: [10.12000/JR24109](https://doi.org/10.12000/JR24109).
- [63] 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).
- [64] HO Y H, CHENG J H, KUAN Shengyao, *et al.* RT-Pose: A 4D radar tensor-based 3D human pose estimation and localization benchmark[OL]. <https://arxiv.org/abs/2407.13930>. 2024.
- [65] AN Sizhe, LI Yin, and OGRAS U. *mRI*: Multi-modal 3D human pose estimation dataset using mmwave, RGB-D, and inertial sensors[C]. 36th International Conference on Neural Information Processing Systems, New Orleans, USA, 2022: 1988.
- [66] CHEN Anjun, WANG Xiangyu, ZHU Shaohao, *et al.* mmBody benchmark: 3D body reconstruction dataset and analysis for millimeter wave radar[C]. 30th ACM International Conference on Multimedia, Lisboa, Portugal, 2022: 3501–3510. doi: [10.1145/3503161.3548262](https://doi.org/10.1145/3503161.3548262).
- [67] LEE S P, KINI N P, PENG W H, *et al.* HuPR: A benchmark for human pose estimation using millimeter wave radar[C]. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, USA, 2023: 5715–5724. doi: [10.1109/WACV56688.2023.00567](https://doi.org/10.1109/WACV56688.2023.00567).
- [68] GADRE A, VASISHT D, RAGHUVANSHI N, *et al.* MiLTON: Sensing product integrity without opening the box using non-invasive acoustic vibrometry[C]. 21st ACM/IEEE International Conference on Information Processing in Sensor Networks, Milano, Italy, 2022: 390–402. doi: [10.1109/IPSNS54338.2022.00038](https://doi.org/10.1109/IPSNS54338.2022.00038).
- [69] LI Yang, LIU Yutong, WANG Yanping, *et al.* The millimeter-wave radar SLAM assisted by the RCS feature of the target and IMU[J]. *Sensors*, 2020, 20(18): 5421. doi: [10.3390/s20185421](https://doi.org/10.3390/s20185421).
- [70] CARUANA R. Multitask learning[J]. *Machine Learning*, 1997, 28(1): 41–75. doi: [10.1023/A:1007379606734](https://doi.org/10.1023/A:1007379606734).

作者简介

陈彦, 博士, 教授, 主要研究方向为多模态感知、多媒体信号处理和数字健康。

张锐, 博士生, 主要研究方向为多模态感知、视频图像去噪。

李亚东, 博士生, 主要研究方向为毫米波雷达成像。

宋瑞源, 博士生, 主要研究方向为多模态机器学习。

耿瑞旭, 博士生, 主要研究方向为毫米波雷达成像。

龚汉钦, 博士生, 主要研究方向为无线感知。

汪斌全, 博士后, 主要研究方向为无线感知。

张东恒, 博士, 副研究员, 主要研究方向为无线感知。

胡洋, 博士, 副教授, 主要研究方向为计算机视觉、多媒体信号处理和多模态感知。

(责任编辑: 于青)