

## 基于金字塔型空洞卷积残差模型的Wi-Fi室内人体姿态估计方法

刘淼<sup>①</sup> 曾小路\*<sup>①</sup> 杨小鹏<sup>①②</sup> 邢程荐<sup>①</sup> 刘宇<sup>②</sup>

<sup>①</sup>(北京理工大学信息与电子学院 北京 100081)

<sup>②</sup>(北京理工大学长三角研究院(嘉兴) 嘉兴 314000)

**摘要:** 人体姿态估计技术能支撑准确获取人体动作与行为特征,在智能监测、人机交互及健康感知等领域展现出广泛的应用潜力。Wi-Fi感知技术因其普遍性、低成本、非接触感知等优势,成为当前人体姿态非接触式感知技术的研究热点。然而,人体活动具有多尺度、非线性及动态变化复杂等特征,不同肢体部位在时间、空间上运动幅度存在显著差异,对姿态估计算法的多尺度特征建模能力提出了更高要求。现有Wi-Fi人体姿态估计算法普遍存在模型参数量大、特征提取不充分的问题,难以在保证计算效率的同时兼顾估计精度,从而限制了其在复杂场景下的应用潜力。针对上述问题,该文设计并优化了一种基于金字塔型空洞卷积的残差网络架构。针对多尺度人体运动特征设计了金字塔型空洞卷积结构单元,该结构能够在保持空间分辨率的同时显著扩大卷积层的感受野,从而有效捕捉多尺度空间与动态变化信息。同时,空洞卷积结构设计能够在一定程度上减少计算量,提升计算效率。为缓解深层网络训练中的梯度消失与模型退化问题,该文进一步设计了残差结构网络,确保模型在深层结构下的特征表达能力与稳定性。为了验证所提方法的有效性,论文设计搭建了完整的多源数据采集系统,可高效获取Wi-Fi姿态估计数据与对应真值数据。实验结果表明,所提方法在人体姿态估计任务中表现优异,MPCK@0.10指标达到94.96%,优于现有算法,验证了方法的有效性与优越性。

**关键词:** Wi-Fi智能感知; 信道状态信息; 人体姿态估计; 深度学习; 多尺度特征提取

中图分类号: TN957.52

文献标识码: A

文章编号: 2095-283X(2026)x-0001-20

DOI: 10.12000/JR26024

CSTR: 32380.14.J26024

**引用格式:** 刘淼,曾小路,杨小鹏,等. 基于金字塔型空洞卷积残差模型的Wi-Fi室内人体姿态估计方法[J]. 雷达学报(中英文),待出版. doi: 10.12000/JR26024.

**Reference format:** LIU Miao, ZENG Xiaolu, YANG Xiaopeng, *et al.* Wi-Fi-based indoor human pose estimation using a pyramid dilated convolutional residual network[J]. *Journal of Radars*, in press. doi: 10.12000/JR26024.

## Wi-Fi-based Indoor Human Pose Estimation Using a Pyramid Dilated Convolutional Residual Network

LIU Miao<sup>①</sup> ZENG Xiaolu\*<sup>①</sup> YANG Xiaopeng<sup>①②</sup> XING Chengjian<sup>①</sup> LIU Yu<sup>②</sup>

<sup>①</sup>(School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China)

<sup>②</sup>(Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314000, China)

**Abstract:** Human pose estimation allows for precise capture of movement and behavioral traits, holding significant potential for applications such as intelligent surveillance, human-computer interaction, and health monitoring. Among emerging approaches, Wi-Fi sensing has gained increasing research interest for contactless human pose detection because of its widespread availability, affordability, and privacy-preserving qualities.

收稿日期: 2026-01-19; 改回日期: 2026-04-14; 网络出版: 2026-05-12

\*通信作者: 曾小路 [xlzeng09@bit.edu.cn](mailto:xlzeng09@bit.edu.cn) \*Corresponding Author: ZENG Xiaolu, [xlzeng09@bit.edu.cn](mailto:xlzeng09@bit.edu.cn)

基金项目: 国家自然科学基金(62301042), 科技创新领军人才(3050013532502)

Foundation Items: The National Natural Science Foundation of China (62301042), The National Leading Talents in Scientific and Technological Innovation Program (3050013532502)

责任主编: 陈彦 Corresponding Editor: CHEN Yan

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License

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

However, human activities are multiscale, nonlinear, and highly dynamic, with notable spatiotemporal variations in motion amplitude across different body parts. These characteristics pose high demands on the ability of algorithms to model multiscale features effectively. Current Wi-Fi-based techniques often struggle with excessive parameter complexity and limited feature extraction, which makes it hard to balance computational speed with accuracy in complex situations. To address these issues, this paper introduces a pyramid dilated convolution block that expands the receptive field while maintaining spatial resolution, making it possible to capture multiscale spatial and dynamic details efficiently. The dilated design also lessens computational redundancy, improving overall efficiency. Building on this, a residual network is designed to prevent gradient vanishing and model degradation, ensuring solid feature representation in deep networks. To test the proposed method, a comprehensive multisource data system was built to synchronize Wi-Fi pose data with ground-truth labels. Experimental results show the proposed approach's superiority, reaching a mean percentage of correct keypoints (MPCK@0.10) of 94.96%, surpassing current leading algorithms. These results confirm the method's effectiveness for reliable and efficient human pose estimation.

**Key words:** Intelligent wireless sensing; Channel State Information (CSI); Human Pose Estimation (HPE); Deep learning; Multi-scale feature extraction

## 1 引言

人体姿态蕴含着人体活动状态、健康状况等丰富信息, 是人体行为识别与感知的重要信息来源。人体姿态估计(Human Pose Estimation, HPE)因其能从传感器数据中直接获取人体动作信息, 在人机交互、智能健康监护及安防监控等场景中展现出广阔的应用前景。尽管以OpenPose<sup>[1]</sup>为代表的计算机视觉技术在姿态估计领域在技术和精度上都取得了显著突破<sup>[2-4]</sup>, 但光学传感器在非视距、低照度、存在遮挡等场景下依然面临难以获取准确信息挑战, 且存在严重的隐私泄露风险, 限制了其在家庭监控等私密空间的应用。以RF-Pose<sup>[5]</sup>为代表的基于雷达(如毫米波、超宽带(Ultra Wide Band, UWB))的感知方案<sup>[6-9]</sup>凭借其射频频透性规避了上述问题, 但其对专用硬件的依赖造成了一定成本限制, 难以满足泛在计算对低成本、高普适、易部署的要求<sup>[10]</sup>。

相比之下, 基于商用Wi-Fi设备的无线感知技术凭借其普适性、低成本及非侵入式隐私保护特性, 逐渐成为泛在感知领域的研究热点<sup>[11]</sup>。与早期利用粗粒度接收信号强度(Received Signal Strength Indicator, RSSI)<sup>[12]</sup>的方法不同, 现代Wi-Fi感知主要依赖物理层的信道状态信息(Channel State Information, CSI)。CSI能够以子载波为粒度, 精细刻画无线信号在多径传播环境下的幅度衰减与相位偏移<sup>[13]</sup>, 从而为捕捉微小的人体肢体动作提供了丰富的物理特征支撑。

为了弥合无线信号与视觉语义之间的模态差异, 早期研究主要致力于将一维CSI信号映射为二维图像特征, 进而迁移计算机视觉领域的成熟架构。例如, Wang等人<sup>[14]</sup>在Person-in-WiFi中率先提出将CSI幅度谱映射为二维图像, 并利用Mask

R-CNN<sup>[15]</sup>和U-Net<sup>[16]</sup>进行人体分割与姿态回归。尽管此类方法验证了跨模态感知的可行性, 但其本质上过度依赖参数量巨大的视觉骨干网络, 推理延迟显著, 难以在资源受限的IoT边缘设备上部署。

为了摆脱对视觉模型的依赖并提升回归精度, 后续研究转向利用深度学习架构显式建模CSI的时空依赖性。Yang等人<sup>[17]</sup>基于跨模态CNN训练实现姿态估计点的端到端回归。Zhou等人<sup>[18]</sup>引入深度信号通道注意力机制, 提出了PerUnet架构, 有效增强了模型对CSI关键特征的感知能力。Deng等人<sup>[19]</sup>提出了CSI-信道空间分解策略(Channel Spatial Decomposition Strategy, CSDS)策略, 利用空间方向与通道敏感度的双视点解耦机制, 进一步提升了模型对姿态动作的辨识度。随着Transformer在序列建模领域的统治级表现, Zhou等人<sup>[20]</sup>提出的MetaFi++引入自注意力机制捕捉长距离时空依赖。然而, 尽管堆叠注意力模块显著提升了特征提取能力, 但Transformer的计算复杂度随序列长度呈二次方增长, 且需要海量数据进行预训练以避免过拟合。对于仅需输出2D坐标的实时应用而言, 这种模型策略在边缘设备上的部署与处理具备挑战性。

为解决上述效率问题, Jiang等人<sup>[21]</sup>尝试结构精简策略, 摒弃了雷达感知中常见的3D高维体素卷积, 利用轻量级CNN提取空间特征并结合长短期记忆网络(Long Short-Term Memory, LSTM)处理时序依赖, 在商用Wi-Fi设备上实现了快速端到端估计。Gian等人<sup>[22,23]</sup>提出的WiLHPE与HPE-Li++利用自适应核选择机制和双重选择性卷积, 使网络能够根据输入信号的频率与通道特征动态调整感受野。为进一步挖掘轻量级模型的性能极限并提升抗噪能力, Nguyen等人<sup>[24]</sup>创新性地设计了

SDy-CNN, 通过动态权重机制精准聚焦于高信息量子载波, 引入贝叶斯优化策略, 降低模型计算量的同时实现了模型配置的最佳协同。然而, 尽管上述方法在参数效率上取得了显著突破, 但其本质仍主要依赖数据驱动的特征拟合, 缺乏针对无线信号物理特性的适用性设计, 此类通用性轻量化方法的模型在复杂动态环境下, 往往难以有效解耦躯干大动作与肢体细粒度微动, 导致姿态估计的细粒度特征丢失, 限制了系统在真实场景下的最终精度与泛化能力。

尽管Wi-Fi人体姿态估计取得了显著进展, 但在实际边缘侧部署中, 如何平衡复杂时空特征的有效提取与计算资源的严格限制仍是该领域面临的核心难题。CSI人体行为信号在时频域上表现出显著的多尺度特性, 大幅位移表现为低频高能分量, 而肢体的细微动作则表现为高频瞬态纹理。为了应对这一挑战, 近期的研究已开始探索复杂的时空融合机制。部分高精度方法引入了Transformer的自注意力机制或双流网络架构, 试图通过显式的全局建模来捕捉长距离的时空依赖。尽管这些策略显著提升了特征表达能力, 但其代价是计算复杂度的非线性增长与巨大的参数冗余。与之相反, 为了适应边缘计算环境, 现有的通用轻量化方法往往倾向于简化特征提取结构, 采用单一尺度的卷积操作或通用的轻量化骨干。然而, 这种简化由于缺乏针对无线信号物理特性的归纳偏置, 通用结构难以在多径干扰严重的复杂环境下有效解耦躯干低频信号与肢体高频微动。这种特征提取能力的缺失, 往往导致全局结构的一致性与局部细节的保真度无法两全, 进而引起关键点预测精度降低。

实际上, 针对时序依赖的建模方式在边缘侧场景下具备一定低效特性。在Wi-Fi人体姿态估计任务中, CSI信号输入通常被切分为极短的时间窗口, 在该微观窗口内, 人体骨架结构变化极微, 不具备追踪长时序的状态演变的需求。然而, 部分方法往往忽略了这一物理特性, 习惯于引入LSTM或

门控循环单元(Gate Recurrent Unit, GRU)等显式时序模块进行状态追踪。这种做法不仅在一定程度上引入了计算冗余, 更导致了难以并行的串行时间延迟瓶颈。因此, 如何在摒弃显式循环单元及庞大注意力机制的前提下, 设计一种能够隐式聚合短时空窗口内多尺度时空信息的高效机制, 是平衡估计精度与时间效率矛盾的关键难点。

针对上述挑战, 本文提出Wi-Fi人体姿态估计模型PyDNet, 主要包含金字塔型空洞卷积模块与残差特征增强模块, 以在多尺度特征建模与网络稳定性之间实现高效平衡。一方面, 本文提出了金字塔型空洞卷积结构, 将传统卷积与空洞卷积相结合, 通过多尺度卷积核在信号域中自适应地捕捉不同层级的细节信息。该设计不仅显著提升了对手部、关节等细微运动的感知能力, 同时在保持全局结构建模能力的前提下, 有效降低了计算量与参数规模, 实现了多尺度特征提取与轻量化设计的统一。网络设计融合了残差块结构以增强深层特征提取能力, 从而在复杂信号环境下保持模型训练的稳定性与特征表达的鲁棒性。通过上述改进, 本文提出的模型在保证姿态估计精度的同时, 显著提升了运算效率, 展现出在实时感知与资源受限应用场景中的部署潜力。本文所提Wi-Fi人体姿态估计框架如图1所示。

## 2 CSI信号模型

在典型的室内传播环境中, Wi-Fi信号不仅包含视距传播路径(主要传播路径), 还受墙壁、家具等静态障碍物影响产生显著的多径效应(如图2所示)。当人体处于信号覆盖区域时, 身体各躯干充当了动态散射体, 引入了随时间变化的额外传播路径。这些由人体运动诱发的动态路径会对信号产生反射、衍射及衰减作用, 从而直接调制接收端的信道状态信息。因此, 接收到的总CSI信息可被建模为由环境决定的静态分量与由人体行为决定的动态分量的线性叠加。基于上述物理模型, CSI可表示为

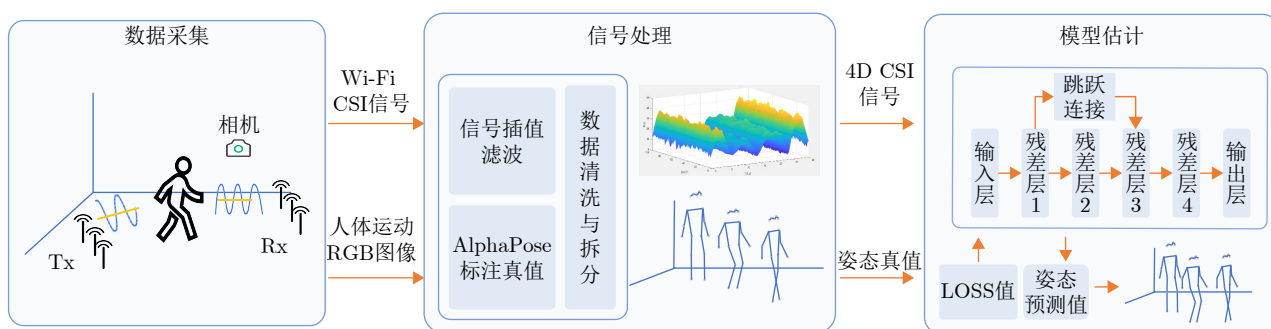


图 1 Wi-Fi人体姿态估计框架

Fig. 1 Framework of the proposed Wi-Fi-based human pose estimation method

$$\mathbf{H}(i, j, k) = \sum_{l=1}^L \mathbf{H}_{s,l}(i, j, k) + \sum_{d=1}^D \mathbf{H}_{m,d}(i, j, k) \quad (1)$$

其中,  $i, j, k$  分别表示发射天线、接收天线和子载波维度索引。 $\mathbf{H}_{s,l}$  表示由室内环境引起的静态信号传播路径的信道特征, 下标  $l$  表示静态信号传播路径索引,  $L$  为静态信号传播路径总数;  $\mathbf{H}_{m,d}$  表示由运动人体引入的动态信号传播路径的信道特征, 下标  $d$  表示动态信号传播路径索引,  $D$  为动态信号传播路径数。 $\mathbf{H}_{s,l}$  为CSI静态分量,  $\mathbf{H}_{m,d}$  包含人体运动信息, 随时间变化, 为动态分量。以跌倒过程为例, 如图3所示, 人体运动会显著改变无线信号的多径传播特性。

图4为实测CSI信号幅度热力图, 图4(a)空场景静态环境下, CSI信号表现出平稳性, 幅值波动仅受微弱的环境噪声影响, 整体均匀且能量分布比较均匀。图4(b)人体跌倒场景中, 当监测区域内发生跌倒行为时, CSI信号产生了剧烈的扰动。特别是人体跌倒时间窗口内, 人体快速下坠与肢体幅度的剧烈变化导致信号发生了显著的反射与散射。这种物理运动在热力图中映射为高能量的幅值状态, 并伴随跨越多个子载波频段的频率选择性衰落纹理。这种显著的信号差异可以证明CSI不仅记录了静态

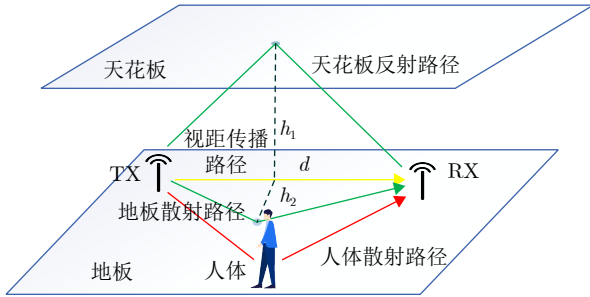


图2 室内空间无线传播路径

Fig. 2 Wireless propagation paths in indoor environment

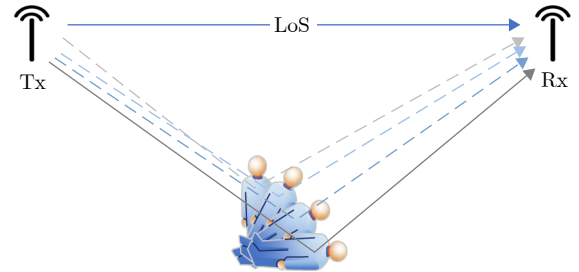


图3 人体运动导致的散射路径变化

Fig. 3 Variation of scattering paths caused by human motion

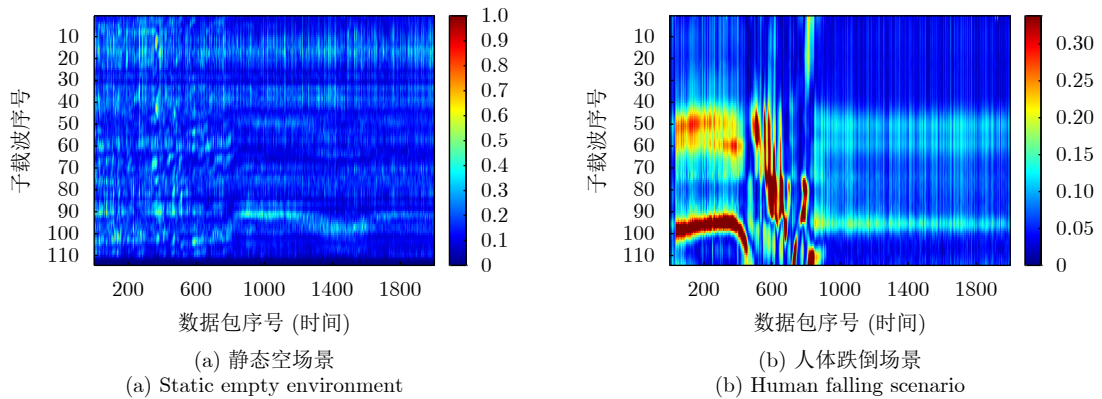


图4 CSI幅值热力图对比

Fig. 4 Comparison of CSI amplitude heatmaps

环境的信道特征, 更敏锐捕捉到人体动态行为引起的多维信号变化。通过解析这些蕴含在幅度和相位中的时频特征能够反演人体的运动状态, 从而实现高精度的姿态估计。

### 3 方法设计

#### 3.1 PyDNet总体架构

针对Wi-Fi信号非线性、非平稳及包含多尺度时空特性的特性, 本文提出了一种基于金字塔空卷积的残差网络PyDNet。其整体架构如图5所示, 主要包含“初始卷积层-残差骨干层-特征输出层”3级级联结构, 旨在从高维CSI时空信号中逐级解耦并提取人体运动特征。网络各阶段的具体参数配置详见表1。

数据流处理首先从时空特征的输入与初始卷积层开始。为了在2D架构中提取时间维度的特征, 网络预先将连续的CSI数据帧序列(本实验中为连续10帧, 对应0.1 s的短时窗口)沿通道维度进行了融合拼接。因此, 网络的输入为预处理后的 $300 \times 136 \times 136$ 二维时空张量, 使得CSI信号随时间演变的动态序列信息被物理内嵌于特征通道中。为了在保留原始信号物理结构的同时降低计算维度, 该输入数据

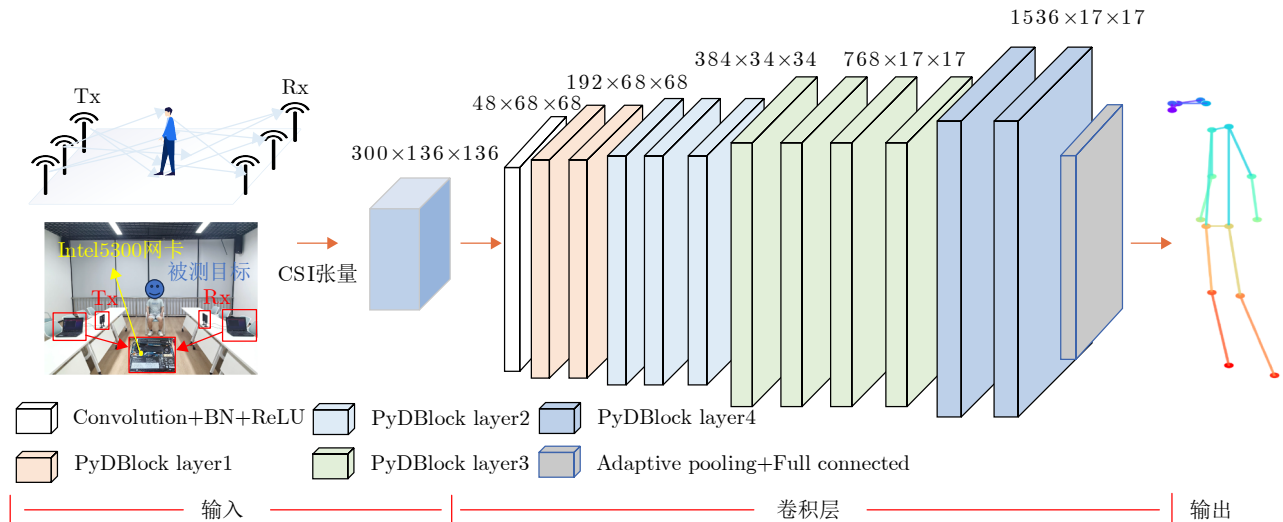


图 5 PyDNet网络结构  
Fig. 5 Architecture of the PyDNet network

表 1 网络结构参数

Tab. 1 Parameters of the network architecture

网络层	组成	输入/输出尺寸( $C \times H \times W$ )	参数说明
初始卷积层	Conv7×7 + BN + ReLU	300×136×136 → 48×68×68	stride=2, padding=3
残差层	Layer1 (2×PyDBlock)	48×68×68 → 192×68×68	multi-scale dilations [1,2,2,3]; grouped conv $G=[3,6,6,12]$
	Layer2 (3×PyDBlock)	192×68×68 → 384×34×34	dilations [2,3]; grouped conv $G=[12,16]$ ; stride=2
	Layer3 (4×PyDBlock)	384×34×34 → 768×17×17	dilations [1,2,3]; grouped conv $G=[6,12,12]$ ; stride=2
	Layer4 (2×PyDBlock)	768×17×17 → 1536×17×17	dilation=3; grouped conv $G=16$
特征融合层	Skip connection (Layer1+Layer3)	768×17×17	1×1 Conv + BN + ReLU; adaptive pooling to 17×17
输出层	Pooling + Fully connected	1536×17×17 → 2×17	FC; output 17 keypoint coordinates

首先经过一个7×7的大卷积核(配置步长为2, 填充为3)进行底层映射, 将特征尺寸下采样至48×68×68。当该2D卷积核在空间维度执行滑动操作时, 其感受野不仅覆盖了空间平面, 更同步跨越了包含时序信息的所有输入通道。通过跨通道的点积求和, 该层在单次计算中便实现了时间动态与空间多径分布的联合特征映射, 有效滤除高频噪声的同时, 提取了反映人体运动模式的基础局部动态分量。随后, 信号进入由Layer1至Layer4构成的残差骨干层, 这是网络提取多尺度特征的核心区域。如图5中部的结构所示, 该阶段堆叠了多个PyDBlock模块, 通道数随着网络深度的增加由192逐级扩展至1536, 以丰富特征的语义表达。Layer1与Layer2主要利用较小的空洞率捕捉短时微动特征, Layer3与Layer4则通过增大空洞率并结合步长为2的下采样操作,

逐步建立长时序依赖关系与全局姿态语义。这种多层级的金字塔结构设计, 使得PyDNet能够在复杂多径环境下, 以较低的计算代价隐式聚合短时窗口内的时空特征, 并行建模信号中蕴含的瞬态细节与长时依赖关系, 显著增强了对姿态相关信号的敏感性。模型在特征输出前引入了跨层特征融合机制, 将Layer1的小尺度高频特征与Layer3的全局深层特征进行拼接, 以补偿深层网络在多次下采样中可能丢失的肢体末端细节。最后, 经过骨干网络处理的高层语义特征进入输出层进行姿态回归。为了恢复空间关联性并生成最终的预测结果, 网络末端采用1×1卷积进行通道整合与降维, 配合自适应池化将特征统一为17×17维度。最终通过全连接层将提取的深层特征映射为17个人体关键点的二维坐标向量, 完成从无线信号到人体骨架的端到端估计。

### 3.2 金字塔型空洞卷积残差块

为了兼顾人体姿态估计中细粒度局部特征与全局时序依赖的协同建模需求,同时解决传统卷积网络在处理CSI信号时面临的行为多尺度特征耦合严重问题,本文结合CSI物理特性设计了网络核心组件金字塔空洞残差块PyDBlock,其结构如图6所示。PyDBlock在传统残差单元的基础上,创新性地融合了金字塔空洞卷积理论与残差学习机制,利用由2~4个并行卷积分支组成的复合金字塔结构替换了传统单一卷积。通过配置差异化的空洞率与分

组数,各分支能够自适应地覆盖从局部( $3\times 3$ 感受野)到全局( $7\times 7$ 感受野)的多尺度范围,从而实现了针对CSI信号中细粒度肢体微动与全局躯干位移特征的高效解耦与协同建模。

#### 3.2.1 基于金字塔模型的CSI多尺度特征解耦

CSI信号可表现为“时间-子载波”图谱,其呈现出显著的非线性与时变特性,人体行为运动特征在CSI时频域具备显著的多尺度特征差异。如图7实测信号CSI时频图所示,躯干的大幅度位移表现为跨越宽频带的高能量波动,而肢体末端的微动则表

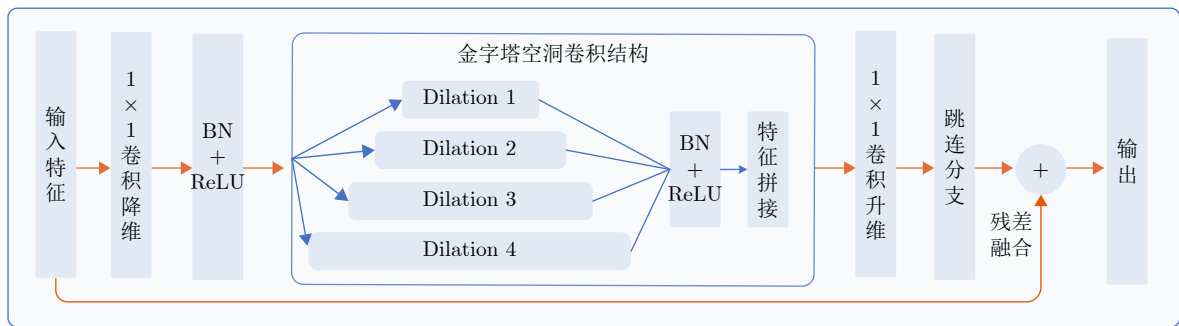


图6 PyDBlock模块结构

Fig. 6 Structure of the PyDBlock module

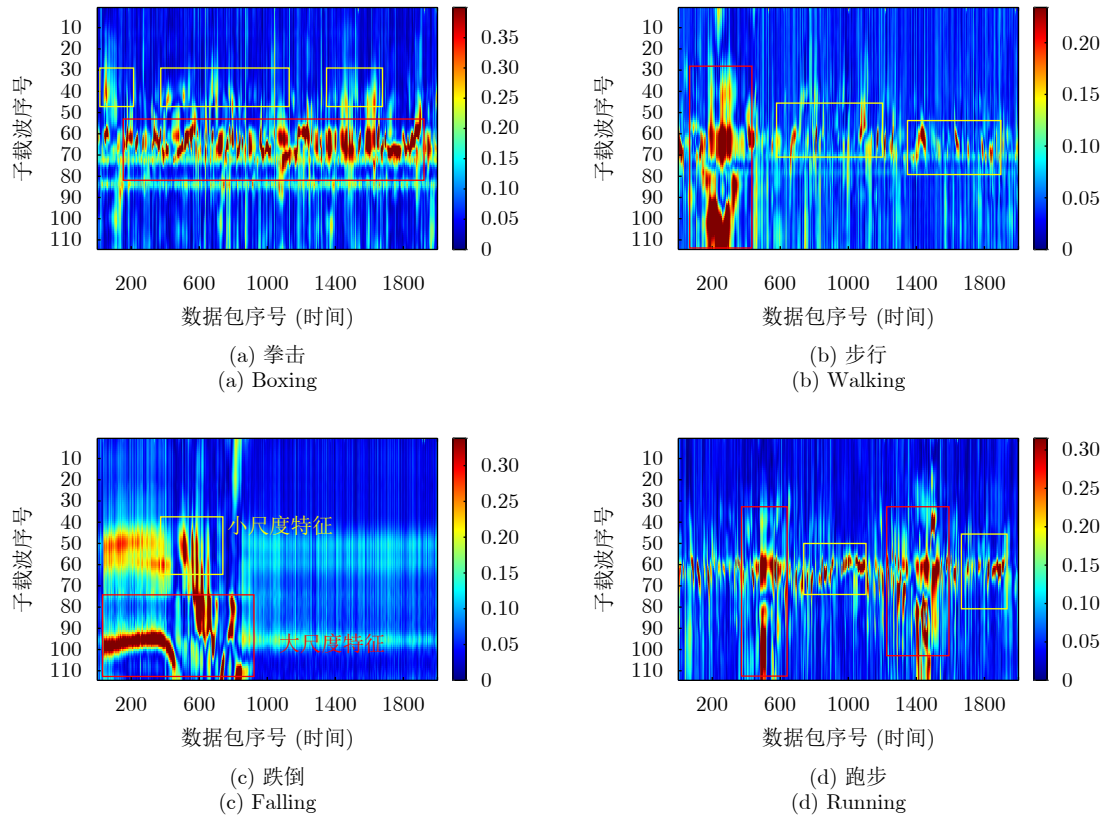


图7 不同人体活动下的CSI幅值热力图

Fig. 7 CSI amplitude heatmaps of different human activities

现为局部子载波的高频特征纹理。这种“全局+局部特征”共存的信号特性，要求特征提取器既捕捉全局躯干特征以聚合上下文信息，也提取局部肢体细节以避免信息模糊。为使网络能够自适应地在多个尺度上捕获从局部微动到全局依赖的完整特征谱，PyDBlock 采用并行金字塔结构，如图8所示，通过配置不同的空洞率来解耦上述特征：小空洞率分支聚焦局部邻域，用于提取肢体末端的瞬态微动特征，避免细节信息丢失；大空洞率分支则利用扩展的感受野，在不降低分辨率的前提下聚合躯干运动的长时序与全频段信息。这种结合金字塔结构的空洞卷积模型设计，使得模型能够并行处理局部细节与全局语义。

### 3.2.2 基于感受野扩展的隐式时序建模

针对长时序依赖建模与模型参数量之间的矛盾，结合Wi-Fi姿态估计任务的物理特性，输入的CSI时空图通常被切分为极短的时间窗口(以本文为例，单个样本覆盖  $T=0.1$  s)，本研究引入“短时准平稳假设”作为网络设计的物理依据。根据人体运动学规律，在这一微观时间单元内，尽管CSI信号因多普勒效应呈现动态波动，但人体骨架的物理构型变化极微，可视为准静态。基于此假设，传统循环神经网络所强调的逐步追踪状态演变在短时段窗口内显得计算冗余且低效。

因此，PyDNet摒弃了显式时序模块，不再将时间维度视为需要递归更新的状态序列，而是将其重新定义为一种特征增强通道。网络的设计目标从“时序预测”转变为“时空聚合”：旨在通过全局感受野并行聚合该窗口内的信号上下文，利用时间维度的相关性来抑制瞬态环境噪声，并增强由肢体微动引起的频率纹理特征。

为了在消除循环结构串行瓶颈的同时有效覆盖短序列依赖，PyDNet采用了全卷积隐式时序建模策略。受时域卷积网络启发，本研究利用空洞卷积的指数级膨胀特性来构建时间维度的全局感知能

力，不同空洞率下的卷积核示意如图9所示。对于卷积核大小为  $K \times K$  且空洞率为  $D$  的空洞卷积，其多层累计等效感受野RF可表示为

$$RF_{out} = RF_{in} + (K - 1)D \quad (2)$$

如图6所示，通过在骨干网络中堆叠多层金字塔空洞卷积，模型深层神经元的有效感受野得以显著扩展。在多层堆叠效应下，网络在时间维度上的累积感受野能够完整覆盖输入的短序列长度。这一设计赋予了模型在结构层面上处理短序列依赖的全局感知能力，全卷积结构能够整体感知整个时间窗口内的动态变化。此外，该设计利用金字塔卷积的并行性取代了循环网络的串行迭代，消除了各个时间步长之间的计算依赖，降低推理延迟，使其更适配于资源受限的边缘计算场景。

### 3.2.3 基于残差学习的梯度稳定机制

在多层金字塔空洞卷积的特征聚合过程中，高频信号容易伴随噪声被过度平滑，在深度网络训练中易出现模型梯度衰减问题，因此引入恒等映射机制：

$$\mathbf{y} = F(\mathbf{x}, \mathbf{W}) + \mathbf{x} \quad (3)$$

其中， $\mathbf{x}$  表示输入特征向量， $F(\mathbf{x}, \mathbf{W})$  表示由卷积、归一化与非线性激活构成的非线性映射函数， $\mathbf{y}$  为残差单元的输出。单纯从残差结构设计来看，PyDBlock中  $F(\mathbf{x})$  的映射机制与标准ResNet及Res2Net等现有变体存在差异。区别于标准ResNet的单分支结构，单一  $3 \times 3$  卷积导致感受野固定，PyDBlock将其重构为并行的多空洞率卷积阵列；区别于Res2Net的分组串行级联，前一分组特征相加至后一分组，易导致不同尺度的特征发生“尺度混叠”，PyDBlock采用并行的独立拓扑，各空洞分支互不干扰地独立提取特定尺度的特征并最终拼接。

该设计构建了一条信息保真通道，允许网络在保留输入信号物理结构(如环境静态反射)的基础上，专注于学习人体动作引起的动态扰动(残差部

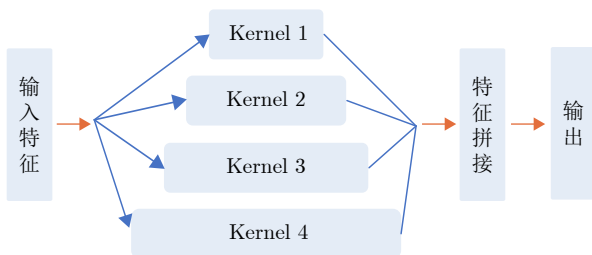


图 8 金字塔卷积结构

Fig. 8 Pyramid convolution structure

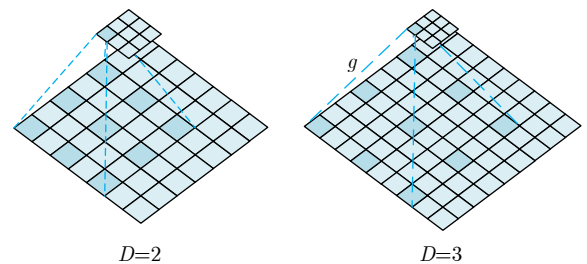


图 9 不同空洞率的空洞卷积示意图

Fig. 9 Illustration of dilated convolutions with different dilation rates

分)。这保障了深层网络的梯度传播,提升了模型在噪声环境下的鲁棒性。

### 3.2.4 CSI信号特性与PyDBlock结构的适配性分析

金字塔空洞卷积结构在视觉图像的语义分割与目标检测等领域已得到广泛应用,其核心优势在于应对二维图像中物理对象像素尺度的剧烈变化。然而,若将该类结构直接移植至Wi-Fi姿态估计任务中进行多尺度特征提取,则面临着显著的跨域物理不匹配问题。首先,CSI信号本质上是受非相干多径干扰的时空频混合序列,空间分辨率较低,常规视觉金字塔结构难以直接处理这种复杂的频率畸变;其次,视觉金字塔模型缺乏对连续信号短时动态演变的建模能力;最后,空洞卷积在离散子载波上的稀疏跳跃采样极易破坏无线信号原本连续的频域纹理,引发严重的网格效应,导致特征保真度丧失。为解决上述模态差异与特征提取瓶颈,本文立足于CSI信号的物理特性与时空演化规律,设计了金字塔空洞残差块PyDBlock,在网络结构与参数化构建上进行了针对性的重构与优化,具体体现在以下3个维度。

(1) 面向多径干扰的频域特征物理层解耦:针对CSI信号中大尺度躯干位移与小尺度肢体微动相互混叠的问题,PyDBlock基于多分支并行结构的空洞率分布,利用较小的空洞率精准捕获由肢体末端产生的高频微动纹理,利用较大空洞率聚合大尺度位移的低频包络,实现多尺度动作特征的自适应提取。

(2) 面向“短时准平稳假设”的时空隐式聚合:为弥补视觉金字塔结构在时序建模上的先天不足,PyDBlock结合人体运动学特征,利用空洞卷积的指数级感受野扩张,巧妙替代了高开销的显式循环时序模块。通过使累积感受野精准覆盖短时准平稳窗口,实现了从常规“时空特征提取”向“短时期序上下文聚合”的结构延展,以高并行度完成了时空动态特征的捕获。

(3) 动态行为特征保真:克服空洞卷积跳跃采样导致的信号非连续性丢失,PyDBlock采用了差异化的空洞率组合以实现特征的密集覆盖,弥补采样空间断层。同时,残差恒等映射机制的引入构筑了底层信号保真通道,使得深层网络能够完整保留环境静态反射等基底信息的同时,金字塔非线性分支能够聚焦于学习由人体动作引发的动态扰动残差。

### 3.3 轻量化参数设计

为了在提升特征表达能力的同时优化模型的参数效率与计算复杂度,PyDNet在PyDBlock模块内

部使用了卷积通道稀疏化与空间参数恒定的双重轻量化策略。首先,在结构设计层面,模块集成了瓶颈架构与稀疏分组卷积机制。如图6所示,输入特征首先经过 $1 \times 1$ 卷积进行通道降维与特征压缩,在低维空间完成多尺度特征提取后,再通过 $1 \times 1$ 卷积恢复通道维度。同时,如表1所示,不同层级采用了差异化的分组策略(如Layer1中 $G=[3, 6, 6, 12]$ ),强制卷积核在特定通道组内运算,从而阻断了通道间的冗余全连接。上述设计的参数效率优势可通过理论分析进行量化论证。对于标准的卷积操作,假设输入通道数为 $C_{in}$ ,输出通道数为 $C_{out}$ ,卷积核尺寸为 $K$ ,其参数量 $P_{std}$ 定义为

$$P_{std} = C_{in} \times C_{out} \times K^2 \quad (4)$$

引入分组卷积(分组数为 $G$ )与瓶颈结构(中间层压缩比为 $r$ ,且 $r < 1$ )后,PyDBlock中核心卷积层的参数量 $P_{ours}$ 显著降低:

$$P_{ours} \approx \frac{1}{G} (r \times C_{in}) \times (r \times C_{out}) \times K^2 \quad (5)$$

由式(4)与式(5)可知,该设计在通道维度上实现了参数的稀疏化,其压缩率主要由分组数 $G$ 与压缩因子 $r$ 决定。

在空间维度上,PyDNet利用空洞卷积打破了传统卷积算子中感受野与参数量之间的二次方依赖关系。若采用标准卷积来获得等效感受野 $R_{dilated}$ ,其所需参数量 $P_{conv}$ 将随感受野面积呈二次方增长,表现为强相关性:

$$P_{conv} \propto R_{dilated}^2 \quad (6)$$

相比之下,PyDNet通过调节空洞率来扩展感受野,在保持物理卷积核尺寸( $K=3$ )恒定的前提下,其实际参数量不由感受野尺度而决定。

$$P_{dilated} \propto K^2 = \text{const} \quad (7)$$

PyDNet通过通道稀疏化设计策略和空间感受野维度的参数恒定特性,消除了模型复杂度对感受野扩展的固有依赖。这使得模型在保持高分辨率特征提取能力的同时,无需依赖过多的下采样操作即可捕获大范围上下文信息,从而在保证性能的前提下显著降低了冗余参数与计算负担。

### 3.4 损失函数

为精确度量模型预测的人体姿态关键点与真实姿态之间的差异,并引导网络在训练过程中实现高效优化,本文采用加权均方误差作为损失函数。损失函数通过计算人体各关键点预测与真实坐标之间的欧氏距离平方,并结合位置置信度进行加权累积。其表达式为

$$\text{Loss} = \frac{1}{P} \sum_{p=1}^P \sum_{i=1}^{17} C_i \|\hat{\mathbf{y}}_{p,i} - \mathbf{y}_{p,i}\|_2^2 \quad (8)$$

其中,  $C_i$ 表示第*i*个人体关键点的置信度权重,  $\hat{\mathbf{y}}_{p,i}$ 为模型预测的第*p*个样本的第*i*个关键点坐标,  $\mathbf{y}_{p,i}$ 为对应的真实关键点坐标,  $\|\cdot\|_2^2$ 用于表示欧氏距离的计算。通过引入置信度加权机制, 模型在训练过程中能够更加关注高置信度区域, 提升人体姿态预测的整体精度与稳定性。

#### 4 实验结果与分析

为验证所提方法的有效性, 本节具体介绍实验数据采集、信号预处理和算法估计结果。数据采集部分涵盖Wi-Fi CSI信号与人体运动RGB图像的多模态数据采集。信号预处理包含信号插值滤波、数据预处理等操作, 以及基于AlphaPose<sup>[25]</sup>的标注真值获取、数据清洗与拆分和多模态数据对齐方法, 为后续模型训练提供高质量数据。本文同时从性能、参数量等维度与现有方法进行了对比, 验证了所提方法的有效性。

##### 4.1 数据采集

本文数据采集包含人体运动CSI信号及其对应的RGB图像关键点标签两部分, 其中RGB图像主要为网络训练阶段提供真值。CSI数据采集系统配备有3根5G天线的Intel5300无线网卡和基于Ubuntu14.04 LTS操作系统的Linux 802.11n CSI Tool (Monitor模式), 如图10所示。RGB数据采集系统采用海康威视DS-E12a自编码摄像头, 配合OpenCv-python工具包来进行人体姿态视频录制及相关RGB图像数据的处理。

为获取高精度姿态真值, 本文利用AlphaPose框架提取RGB图像的二维人体关键点。针对复杂室内场景(如遮挡、低照度)易引发的关键点漂移问

题, 本文在预处理阶段引入置信度截断策略, 以量化并控制真值的标注误差上界。AlphaPose的输出置信度与其空间定位的欧氏误差高度相关, 其在COCO验证集上可达76.8 mAP的高精度基准。据此, 本实验将置信度筛选阈值设定为0.5, 该操作在数学上近似等效于伪真值, 限定了最大容许像素误差上界。置信度低于0.5的关键点将被系统判定为越界噪声并予以剔除。该约束策略有效阻断了视觉特征的漂移误差向Wi-Fi感知网络跨模态传播, 确保了深度网络对人体真实空间姿态的精准收敛。

实验选取典型室内会议室场景, 平台室内布局 and 实景测量如图10所示。受试者群体涵盖了不同体型特征, 身高跨度为153~174 cm, 体型涵盖了偏瘦、匀称到偏胖等多种类型。实验中实验者保持自然运动状态, 保证数据集的真实与完备性。每组数据时长30 s, 采样频率为100 Hz。人体运动过程中, 使用摄像头同步采集人体姿态运动视频, 确保通过数据对齐处理后CSI信号与人体姿态的同步性。实验共采集5人, 总计36000组有效样本。其中数据被划分为训练集(60%)、验证集(20%)和测试集(20%)。训练集用于模型的学习与优化, 验证集用于辅助动态调整超参数, 独立的测试集则用于客观评估模型的性能与泛化能力。

##### 4.2 数据预处理

数据预处理包含CSI数据清洗、RGB图像处理与人体姿态真值标注、CSI与RGB图像时间对齐3大核心步骤。为了确保系统拥有充足的Wi-Fi信道状态信息数据用于匹配人体姿态估计的每组关键点信息, CSI数据采集频率设为100 Hz, 处理后RGB图像为每秒10张。因此, CSI数据与人体姿态关键点数据量为10:1, 即每组数据包含10帧CSI数据和一组真值(17个人体关键点信息)。由于室内环



(a) 实验系统组成  
(a) Composition of the experimental system

(b) 实验场景部署  
(b) Deployment of the experimental scenario

图 10 实验系统与场景示意

Fig. 10 Experimental system and scenario illustration

境多径效应、硬件自身噪声干扰等因素,原始CSI信号中包含异常值和高频噪声,且存在部分数据包丢失的情况。为提升数据质量,本方案对原始数据进行预处理,处理流程如图11所示。首先使用最近邻插值填补因丢包产生的时间序列空缺,然后采用卡尔曼滤波对CSI子载波数据进行平滑处理,以在保留人体运动特征的同时降低环境噪声干扰,最后通过双线性插值统一数据维度,使其适配后续模型的输入要求。信号插值和滤波前后对比的波形分别如图12(a)和图12(b)所示,相较于原始信号,其毛刺和突变明显减少,更清晰地反映了人体动作的变化趋势。

为获取与CSI信号对应的人体姿态真值标签,本方案基于同步采集的视频数据进行构建。首先调用OpenCV-Python库对视频流进行分帧处理,并将帧率调整为10 FPS。随后,使用开源工具AlphaPose算法提取RGB图像中的人体形态特征,输出包含17个关键点的二维坐标及置信度。最后,将提取的坐标序列保存为JSON文件。该方法实现了数据的自动化标注,能够有效获取连续的人体姿态坐标数据。

在多模态数据采集过程中,视觉视频流与CSI序列分别受控于上位机操作系统内核与Wi-Fi网卡底层硬件时钟。为确保多模态监督信号在时间维度上的严密对应,本实验依次进行了绝对时间轴校准操作。由于硬件时钟存在固有的起振相位差异,两类模态数据在物理时间轴上存在初始偏差。实验采用基准偏移的线性时间映射来弥补该初始相位偏

差。设采集启动时首个CSI数据包的硬件时间戳为 $T_{hd}^{start}$ ,同时记录的上位机系统时间戳为 $T_{sys}^{start}$ ,由此提取异构时钟间的初始固定偏差 $\Delta T$ 。

$$\Delta T = T_{sys}^{start} - T_{hd}^{start} \quad (9)$$

后续以系统时间为绝对基准,对于任意第 $i$ 个CSI数据包(其硬件时间戳记为 $T_{hw}^i$ ),其映射至系统时间轴的校准时间戳 $\hat{T}_{sys}^i$ 可表示为

$$\hat{T}_{sys}^i = T_{hw}^i + \Delta T \quad (10)$$

通过该线性补偿操作,离散的CSI数据包序列被对齐至视频帧所在的统一绝对时间坐标系下,有效消除了时钟起振误差。

### 4.3 评估准则

为评估所提出人体姿态估计算法的性能,本文采用每关节位置误差(Per Joint Position Error, PJPE)和正确关键点百分比(Percentage of Correct Keypoints, PCK)作为核心评价指标,分别从“绝对定位精度”和“相对准确率”两个维度对模型性能进行综合量化评估。

每关节位置误差(PJPE)通过计算预测关键点与真实关键点之间的像素级欧氏距离,定量反映关键点预测位置偏差,能有效评估模型的关键点定位能力,定义如下:

$$PJPE_i = \frac{1}{P} \sum_{p=1}^P \|\hat{y}_{p,i} - y_{p,i}\|_2 \quad (11)$$

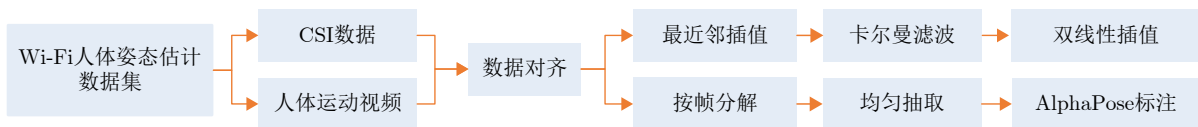
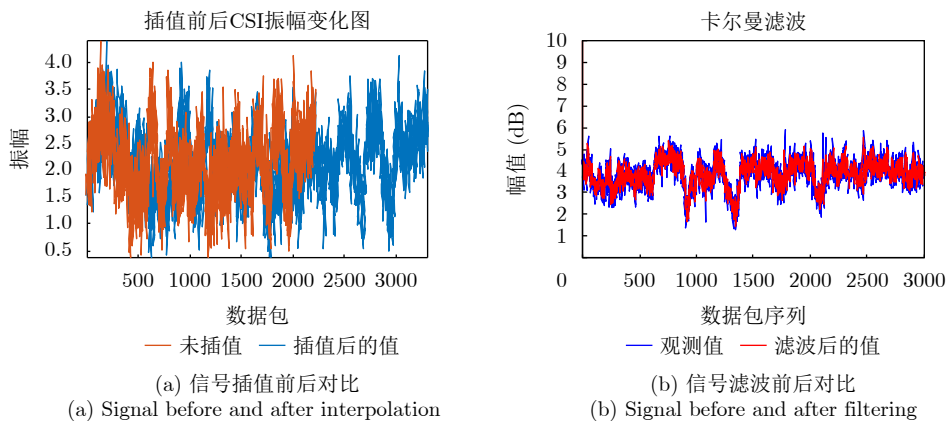


图 11 信号处理流程

Fig. 11 Signal processing flow



(a) 信号插值前后对比  
(a) Signal before and after interpolation

(b) 信号滤波前后对比  
(b) Signal before and after filtering

图 12 信号预处理效果对比

Fig. 12 Comparison of signal preprocessing effects

其中,  $\hat{\mathbf{y}}_{p,i}$  和  $\mathbf{y}_{p,i}$  分别为第  $p$  个样本中第  $i$  个关键点的预测值与真实值,  $\|\cdot\|_2$  为欧氏距离运算符,  $P$  为样本总数,  $i$  表示关键点索引。为进一步评估模型在整体关键点定位上的表现, 本文计算所有关键点 PJPE 的平均值, 作为全局绝对误差指标:

$$\text{MPJPE} = \frac{1}{N} \sum_{i=1}^N \text{PJPE}_i \quad (12)$$

其中,  $N$  为关键点总数。

人体姿态估计除关注每个关键点的位置误差, 关键点之间的相对关系是反映人体姿态整体空间重构能力的重要信息。为了综合关键点标注精度与空间位置误差, 并评估模型在不同容错阈值下的鲁棒性, 本文采用正确关键点百分比 PCK 作为评价指标。相比于绝对位置误差评估准则, PCK 指标通过引入人体边界框尺寸对误差进行归一化处理, 有效消除人体尺度差异和拍摄距离对误差评估的影响, 从而更直观地反映模型在可接受误差范围内的预测准确率, 适用于跨场景、跨尺度的姿态估计性能比较。PCK 定义如下:

$$\text{PCK}_i @ \alpha = \frac{1}{P} \sum_{p=1}^P \mathbb{I} \left( \frac{\|\hat{\mathbf{y}}_{p,i} - \mathbf{y}_{p,i}\|_2}{\sqrt{(w^p)^2 + (h^p)^2}} \leq \alpha \right) \quad (13)$$

其中,  $\alpha$  为判定阈值,  $w^p$  和  $h^p$  为第  $p$  个样本对应的人体边界框的宽度与高度, 用于对欧氏距离进行归一化。若预测关键点与真实关键点之间的欧氏距离小于阈值  $\alpha$ , 则视为预测正确。 $\mathbb{I}(\cdot)$  为指示函数: 当预测关键点与真实关键点之间的归一化距离小于阈值  $\alpha$  时, 视为预测正确。为综合评价模型在全身关键点上的整体预测性能, 本文在 PCK 基础上进一步定义平均正确关键点百分比 (Mean Percentage of Correct Keypoints, MPCK) 指标:

$$\text{MPCK} @ \alpha = \frac{1}{N} \sum_{i=1}^N \text{PCK}_i @ \alpha \quad (14)$$

其中, 本文所采用的人体姿态关键点共包含 17 个关节点, 具体定义如图 13 所示。MPCK 可作为整体姿态估计精度的全局度量, 为不同模型之间的性能对比提供统一标准。

## 4.4 实验结果与分析

### 4.4.1 自测数据集验证

为了全面评估所提方法 PyDNet 在非平稳 Wi-Fi 信号下的姿态估计性能, 本节选取了 PerUnet<sup>[18]</sup>, SDy-CNN<sup>[24]</sup> 及 WPFormer<sup>[20]</sup> 等主流方法作为对比基准。性能对比包含关键点定位准确率 (PCK)、平均关节位置误差 (PJPE) 及模型复杂度 3 个维度, 验

证 PyDNet 在精度、鲁棒性与计算效率方面的综合优势。

图 14 直观展示了所提方法与对比模型在多种典型人体活动下的姿态重构结果, 可以看出本方法在细粒度特征重构上具备明显优势。如图 14 行走场景所示, 对比模型 PerUnet, SDy-CNN 及 WPFormer 生成的姿态骨架在下肢摆动幅度上普遍存在偏差, 所提方法 PyDNet 准确重构了人体肢体角度, 同时保持了人体步态整体运动的一致性。图 14 抬手场景中, PyDNet 精准估计出了腕关节空间坐标, 而 PerUnet, SDy-CNN 存在一定程度的端点漂移现象, WPFormer 的手部关节明显缩短。此外, 在处理交叉遮挡及大幅度协同变化等复杂姿态时, 所提方法具有明显的优势。如图 14 弯腰场景, PyDNet 精准捕捉了弯腰过程中头部坐标变化, 除 SDy-CNN 模型生成的头部坐标有微幅弯曲外, PerUnet 与 WPFormer 生成的姿态头部坐标完全没有体现头部变化特征。图 14 下蹲场景中, PyDNet 生成的骨架具备躯干的旋转特征与深度变化, 未出现结构畸变现象, PerUnet, SDy-CNN 及 WPFormer 均没能有效表征下蹲行为。特别是在图 14 快速变化跌倒场景中, PyDNet 在严重的自遮挡位体情况下仍能实现姿态重构, 对比模型 PerUnet、SDy-CNN 及 WPFormer 模型发生失效并错误地将其混淆为下蹲行为。该实验直观展示了所提金字塔空洞结构在协同捕捉局部细粒度特征与全局空间依赖方面的优越性。

### (1) 关键点定位准确率分析

本节主要评估所提模型在不同阈值下的关键点定位准确性 (PCK)。如表 2 所示, 随着阈值  $\alpha$  收紧,

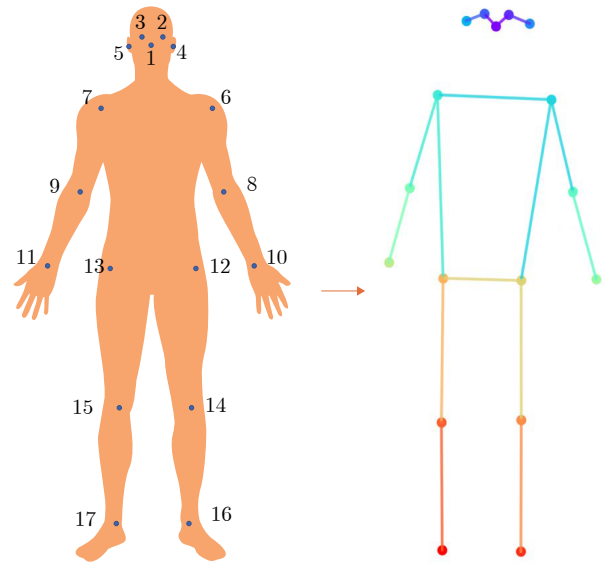


图 13 人体姿态关键点示意图

Fig. 13 Illustration of human body keypoints

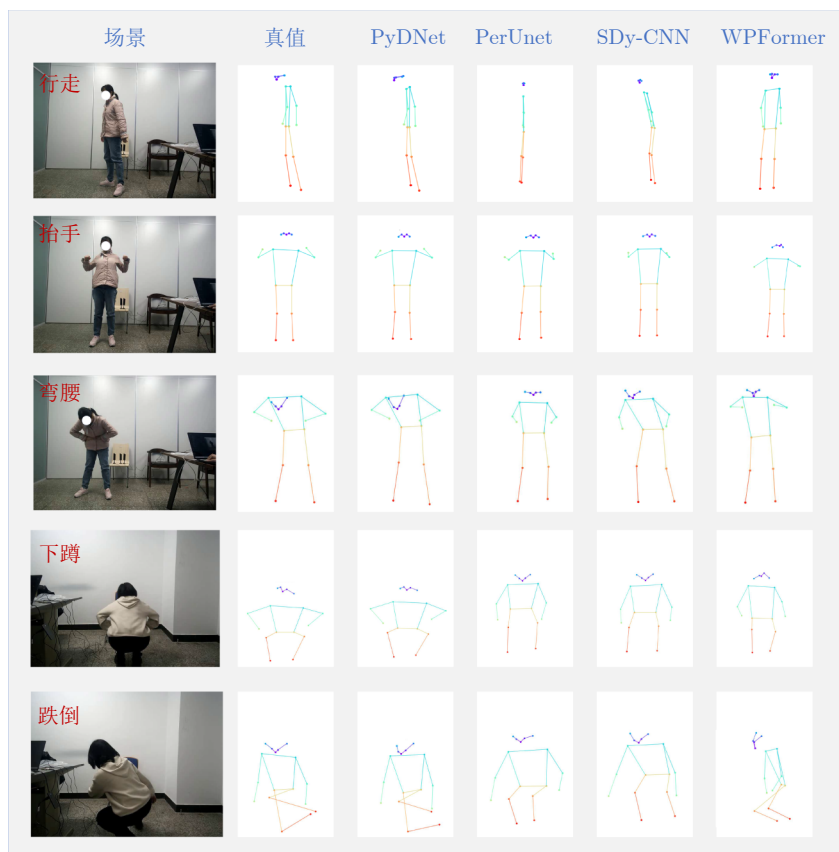


图 14 人体姿态估计效果对比图

Fig. 14 Comparison of human pose estimation results across different models

表 2 各模型PCK结果(%)  
Tab. 2 PCK results of different models (%)

人体关键点	PCK@0.10				PCK@0.05				PCK@0.01			
	PyDNet	PerUnet	SDy-CNN	WPFormer	PyDNet	PerUnet	SDy-CNN	WPFormer	PyDNet	PerUnet	SDy-CNN	WPFormer
鼻子	<b>95.00</b>	89.25	94.28	78.79	<b>85.76</b>	69.90	79.08	57.96	<b>42.35</b>	18.26	19.85	9.33
耳朵	<b>94.94</b>	88.97	94.09	78.74	<b>85.60</b>	70.65	79.56	58.69	<b>43.01</b>	18.55	20.03	11.33
眼睛	<b>95.48</b>	90.56	94.86	81.19	<b>87.00</b>	73.88	81.24	62.44	<b>44.35</b>	21.08	21.38	15.16
肩关节	<b>96.33</b>	92.44	96.01	83.74	<b>88.40</b>	74.64	81.74	63.75	<b>42.38</b>	18.92	18.81	13.66
肘关节	<b>93.67</b>	86.03	91.83	75.31	<b>80.83</b>	57.52	67.47	47.07	<b>23.76</b>	5.39	7.63	3.50
手腕	<b>87.17</b>	65.76	80.08	53.74	<b>66.74</b>	30.38	41.62	22.53	<b>13.29</b>	2.06	2.22	1.65
髋关节	<b>98.17</b>	95.30	97.18	90.81	<b>92.80</b>	81.90	86.52	71.89	<b>48.90</b>	24.91	25.63	11.53
膝关节	<b>97.53</b>	93.06	95.14	89.55	<b>91.63</b>	78.53	82.47	71.22	<b>48.31</b>	24.78	25.06	12.03
脚踝	<b>96.34</b>	91.08	93.90	86.67	<b>90.13</b>	74.22	79.26	68.03	<b>47.07</b>	20.32	15.33	11.84
均值	<b>94.96</b>	87.98	92.97	79.90	<b>85.41</b>	67.84	75.23	58.19	<b>39.09</b>	17.07	17.18	10.04

注：表内加粗数值表示各指标下的最优结果。

所有模型的性能均有所下降，但所提PyDNet的衰减幅度最小。在 $\alpha=0.10$ 的高精度要求下，PyDNet实现了94.96%的平均准确率(MPCK)，相较于SDy-CNN (92.97%)提升了1.99%，较PerUnet (87.98%)和WPFormer (79.90%)分别提升了6.98%和15.06%。当阈值进一步收紧至 $\alpha=0.05$ 时，PyDNet仍保持85.41%的高可用性，对比模型WP-

Former已降至58.19%，充分证明了所提方法具备更出色的细粒度姿态重构能力。横向对比表2中各关节维度表现，PyDNet在上肢关键点(肩、肘、腕)的检测精度提升最为显著。以PCK@0.10标准为例，肩关节与肘关节的检测准确率分别达到了96.33%和93.67%，显著优于WPFormer的83.74%和75.31%。这一性能增益主要归因于人体上肢运动幅度大且速

度快，单一尺度的卷积核难以兼顾局部细节与运动轨迹。所提出金字塔空洞卷积结构PyDBlock通过多分支空洞卷积并行提取特征，在不丢失分辨率的前提下有效扩展了感受野，使得模型既能捕捉肢体末端的瞬态变化，又能维持全局的结构约束，从而显著提升了复杂动作下的重构能力。

### (2) 关节位置误差分析

本节主要用于分析所提模型绝对关节重构误差性能，即各模型的平均关节位置误差(MPJPE)。如表3所示，PyDNet在所有单一关键点及整体均值上均实现了最低误差，其MPJPE仅为11.90，相对于PerUnet (MPJPE = 20.49), SDy-CNN(MPJPE = 16.58)和WPFormer(MPJPE = 27.14)，误差分别降低了约41.9%、28.2% 和56.2%。这一显著性能差异，进一步验证了残差结构在特征传递中的信息保留作用。通过恒等映射机制，PyDNet有效缓解了深层网络对底层物理特征的遗忘，使得回归头能够利用更丰富的原始信息来校准预测坐标，从而保持精准的定位能力。

### (3) 蒙特卡罗实验

为排除单一数据集划分带来的随机性偏差，本文引入蒙特卡罗交叉验证以评估模型性能的统计稳定性。对原始数据集进行5次独立的随机划分，每次均按训练集:测试集:验证集 = 6:2:2的比例重新划

分，并分别对所提方法进行独立训练与评估，量化测试结果如表4所示。

如表4所示，在5次独立随机划分下，模型的各项评估指标均保持稳定。在绝对定位误差方面，平均单关节位置误差(MPJPE)的统计均值约为12.50，最大极差为0.60(区间12.08~12.68)，标准差约为0.23，呈现出较小的波动范围。在准确率评估方面，各阈值下的正确关键点百分比(MPCK)高度一致。表4结果表明本文方法在面对多重数据分布随机扰动时具有显著的鲁棒性。其稳定的性能表现并非源于特定数据划分下的过拟合，而是得益于PyDBlock模块在多尺度特征解耦与时空隐式聚合方面的机制设计。实验结果进一步证实了该模型特征提取能力的统计显著性与泛化可靠性。

### (4) 算法鲁棒性分析

为全面评估模型在真实复杂电磁环境情况下(突发干扰导致的数据包丢失或子载波缺失)的稳健性，实验对测试集CSI数据进行了随机数据包丢弃与子载波维度置零模拟。丢弃率设置区间为0~1.0，间隔0.1。各对比算法的平均单关节位置误差(MPJPE)性能衰减曲线与量化结果如图15所示。

#### (a) 基于数据包丢失维度的时序鲁棒性分析

如衰减曲线图15(a)所示，随着数据包丢失率的增加，所有对比模型的平均估计误差均呈单调递增趋势。然而，本文方法PyDNet展现出最为显著的抗丢包干扰能力。在0~0.8的宽泛丢包率区间内，PyDNet的MPJPE始终维持在全场最低水平。具体分析，对于数据丢包率为0.3时，PyDNet的绝对误差仅为27.56，显著低于常规卷积模型PerUnet (38.99)与基于Transformer架构的WPFormer (47.09)，具备时序维度上的高鲁棒性。这一效果得益于PyDNet架构中多层空洞卷积的叠加，有效扩张了时间维度的感受野，使得模型能够利用时间序

表 3 各模型PJPE结果

Tab. 3 PJPE results of different models

人体关键点	PyDNet	PerUnet	SDy-CNN	WPFormer
鼻子	<b>11.71</b>	19.67	14.97	28.49
左耳	<b>11.63</b>	19.49	14.80	28.30
右耳	<b>11.78</b>	19.67	14.87	28.50
左眼	<b>10.90</b>	17.72	14.20	25.23
右眼	<b>10.96</b>	17.79	13.86	25.68
左肩	<b>10.21</b>	16.75	13.89	22.72
右肩	<b>9.95</b>	16.50	13.44	23.34
左肘	<b>14.38</b>	24.50	19.47	30.84
右肘	<b>14.88</b>	25.29	20.48	33.62
左腕	<b>21.15</b>	37.94	29.47	45.64
右腕	<b>23.05</b>	41.63	32.21	51.74
左髌	<b>7.88</b>	13.19	11.42	18.51
右髌	<b>7.61</b>	12.94	11.14	17.47
左膝	<b>8.27</b>	14.31	12.73	17.90
右膝	<b>8.51</b>	15.23	13.31	19.49
左脚踝	<b>9.42</b>	17.20	15.52	20.99
右脚踝	<b>10.08</b>	18.44	16.12	22.86
均值	<b>11.90</b>	20.49	16.58	27.14

注：表内加粗数值表示各指标下的最优结果。

表 4 基于蒙特卡罗仿真的不同数据划分下的实验结果

Tab. 4 Performance across different data partitions based on Monte Carlo simulations

实验次数	MPCK@ 0.01	MPCK@ 0.05	MPCK@ 0.10	MPCK@ 0.20	MPJPE
1	38.36	85.25	94.99	98.83	12.08
2	37.22	84.49	94.54	98.76	12.45
3	36.58	83.93	94.32	98.78	12.68
4	36.28	84.04	94.43	98.73	12.65
5	37.24	84.10	94.26	98.67	12.64
平均值	37.14	84.36	94.51	98.75	12.50
标准差	0.72	0.48	0.26	0.05	0.23

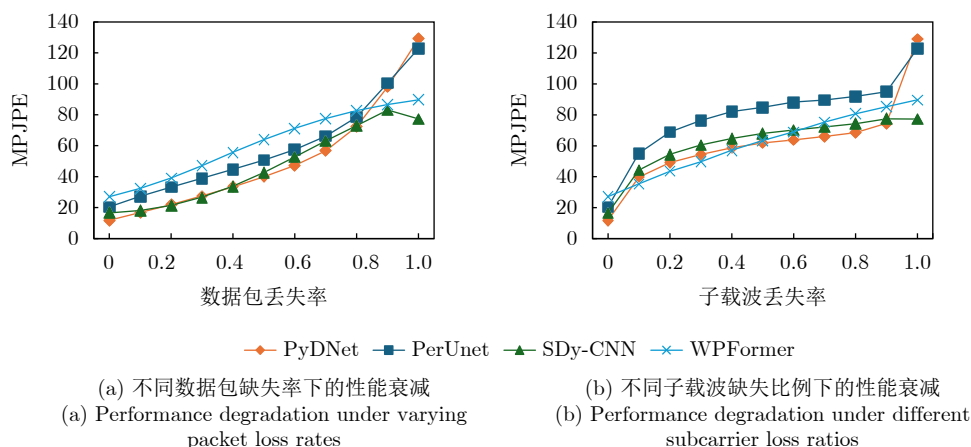


图 15 数据缺失对所提算法的影响

Fig. 15 Impact of incomplete data on the proposed algorithm

列的物理冗余度，自适应地填补瞬时数据缺失带来的特征断层。

#### (b) 基于子载波缺失维度的频域鲁棒性分析

针对频域维度的特征恶化情况，衰减曲线图15(b)进一步证实了PyDNet的结构优势。在子载波丢失率处于0.4~0.9的严重恶化区间内，PyDNet性能衰减过程较为平缓。在高达80%的子载波严重受损情况下，PyDNet的MPJPE稳定在68.58，优于SDy-CNN (74.28)与WPFFormer (80.70)。尽管序列建模方法WPFFormer在极低缺失率(<0.2)下具备一定的抗扰动能力，但在高缺失率下误差加剧。相反，PyDNet凭借多分支金字塔空洞卷积的并行架构，能够有效地从残存的子载波维度中提取互补的高频局部细节与低频全局特征，抵御了单一频带受损引发的特征空间塌陷。

#### (c) 极限缺失状态下的模型“先验偏置”现象分析

对比分析图15(a)、图15(b)中各算法表现，当数据或子载波丢失率达到1.0 (即100%信号盲区，输入特征完全失效)时，对比模型SDy-CN和WPFFormer在绝对误差上呈现出异常的收敛态势(MPJPE分别被限制在77.29与89.66)，而本文方法的误差则符合物理直觉地大幅发散(129.06)。

从深度学习回归机理分析，对比算法在零有效输入情况下表现出的低误差实质上是模型坍塌至数据集均值的过拟合表现，模型参数过分依赖训练集的空间分布先验。在特征缺失时，网络偏置项主导输出了一个静态的“平均安全姿态”，在数学上掩盖了预测失效的本质，反映了SDy-CNN和PFormer算法模型对实际输入信号的低敏感度。相反，PyDNet模型在极限状态下的误差迅速增加，从反

面有力地印证了其预测过程是强信号驱动的。这表明PyDNet所提取的多尺度特征高度反映了CSI信号的实时物理变化，而非单纯对数据集空间分布的统计记忆。在实际的无线感知部署中，这种高输入敏感度能有效避免模型在无目标或设备断联状态下输出虚假的高置信度姿态，从而具备更高的系统安全界限。

#### 4.4.2 公开数据集验证

为验证模型在跨受试者与复杂场景下的泛化能力，本文基于公开的大规模Wi-Fi人体姿态估计数据集Wi-Pose<sup>[26]</sup>开展交叉验证实验。Wi-Pose数据集基于5 GHz Wi-Fi信号进行CSI数据采集。该项目招募了12名体型各异的志愿者，在室内环境下执行12种日常动作(包含弯腰、下蹲、跳跃、行走等)。数据集共包含166600组样本。实验按8:2比例随机划分训练集与测试集。测试时设置置信度0.2为可信阈值，置信度低于该阈值的关键点将被系统判定为无效数据，并在模型测试过程中予以剔除。实验采用平均单关节位置误差(MPJPE)及不同阈值(0.01, 0.05, 0.10, 0.20)下的正确关键点百分比(MPCK)作为评估指标，测试集评估结果如表5所示。

由表5可知，PyDNet在各评估尺度上均取得最优性能。在严格评估阈值( $\alpha=0.01$ 与0.05)下，PyDNet准确率分别达15.50%与62.14%，显著优于其余基线模型。这表明本文方法在捕捉细微姿态变化及高频局部特征方面具有优势。在绝对关节误差方面，PyDNet的MPJPE降至27.74，较次优基线模型PerUnet (39.83)的误差降低约30.3%。验证结果表明，所提方法能有效克服CSI信号多径干扰，在面对未知受试者与复杂测试环境时具备良好的泛化能力与实际部署潜力。

表 5 Wi-pose数据集下不同模型的性能对比

Tab. 5 Performance comparison of different models on the Wi-pose dataset

模型	MPCK@0.01 (%)	MPCK@0.05 (%)	MPCK@0.10 (%)	MPCK@0.20 (%)	MPJPE
PyDNet	<b>15.50</b>	<b>62.14</b>	<b>78.72</b>	<b>91.06</b>	<b>27.74</b>
PerUnet	3.15	39.88	65.84	86.68	39.83
SDy-CNN	1.90	30.14	59.71	87.15	42.68
WPFormer	4.79	39.35	61.84	83.76	43.14

注：表内加粗数值表示各指标下的最优结果。

#### 4.4.3 跨域自适应性能评估

现有的基于深度学习的Wi-Fi感知模型通常在独立同分布假设下表现良好，但在实际场景部署中，CSI信号高度依赖于物理环境的多径效应。当测试环境、收发设备位置或受试者发生改变时，会产生显著的域偏移现象，导致模型性能严重退化。为了评估本文算法在未知环境下的泛化能力与鲁棒性，本文借鉴Ph-Wri<sup>[27]</sup>中消除内容干扰项的特征解耦思想，设计了跨域自适应评估实验，旨在通过引入辅助的域对抗微调策略来验证所提算法跨域应用的潜力。

实验将自采数据集定义为源域，开源数据集Wi-pose定义为目标域。在源域预训练模型(Baseline)的基础上，本文于特征提取网络末端引入域分类器与梯度反转层构建对抗架构(Proposed)。在微调阶段，将源域和目标域的数据混合输入。通过引入负向惩罚损失机制，系统在优化姿态估计精度的同时，强制特征提取器最大化域分类器的误差，构造极小极大博弈。这种博弈机制迫使模型遗忘与特定房间相关的多径特征，提取出纯粹的、跨域不变的人体骨架物理特征。

不同策略在目标域上的跨域姿态估计性能如表6所示。从绝对物理误差来看，当模型直接进行跨域零样本推理时，受限于目标域中截然不同的设备部署与多径效应干扰，模型产生了预期的域偏移现象，MPJPE高达210.40。然而在通过对抗微调剥离了环境特异性噪声后，模型的MPJPE大幅降低至42.99。在MPCK的评估中，该泛化潜力得到了进一步印证。在较宽松的阈值( $\alpha=0.3\sim0.5$ )下，对齐环境干扰后的模型准确率高达97.14%以上，证明算法在跨设备场景中依然能极其精准地重建人体的宏观躯干结构。在要求预测点与真实点极度一致的严苛阈值( $\alpha=0.1$ )下，Baseline的细粒度定位因环境噪声掩盖而近乎失效，但消除环境干扰后，算法的特征提取优势得以彻底释放，准确率大幅跃升至58.70%，绝对精度提升53.29%。综合MPJPE与MPCK的显著性能跨越表明，本文提出的骨干网络

表 6 目标域Wi-pose上的跨域姿态估计性能对比

Tab. 6 Performance comparison of cross-domain pose estimation on the target domain (Wi-pose dataset)

评估指标	Baseline	Proposed	性能变化
MPCK@0.1	5.41	58.70	53.29%
MPCK@0.2	17.09	88.18	71.09%
MPCK@0.3	30.51	97.14	66.63%
MPCK@0.4	43.72	99.41	55.69%
MPCK@0.5	56.03	99.90	43.87%
MPJPE	210.40	42.99	-167.42

架构实际上已经成功学习到了高质量的人体姿态本征特征。一旦环境干扰变量被对齐，算法即可展现出极强的空间定位恢复能力，具备在复杂真实环境中规模化部署的鲁棒性基础。

#### 4.4.4 模型参数量

为验证模型在物联网及边缘计算设备上部署的理论潜力，实验对各对比算法的计算开销与运行效率进行了综合量化评估。评估指标包含模型参数量(Parameters)、浮点计算数(FLOPs)、单帧推理延时(Inference time)、系统吞吐量(FPS)以及数据预处理耗时。硬件测试基准条件严格保持一致，具体对比数据如表7所示。

从空间复杂度来看，所提方法PyDNet展现出显著的轻量化优势。其模型参数量仅为6.35 M，相较于基于Transformer架构的WPFormer (26.73 M)和常规卷积架构PerUnet (17.49 M)，参数量分别大幅削减了约76.2%与63.7%。同时，PyDNet的浮点计算数(FLOPs)控制在12.22 G，这意味着该模型在理论上具备较低的内存占用和算力门槛，展现出适配边缘端侧设备的潜力。

在时间复杂度方面，PyDNet的单帧推理延时为4.70 ms，系统吞吐量达到212.93 帧/s。客观而言，由于引入了多分支金字塔空洞卷积结构以提取多尺度特征，其推理速度差于结构极简的SDy-CNN (1.07 ms)。从底层执行逻辑分析，该时间开销主要源于两方面因素：其一，多分支架构增加了内存访

表7 对比算法计算量对比

模型	参数量 (M)	浮点计算数 Flops(G)	单帧推理 延时(ms)	吞吐量 (帧/s)	数据预处理 耗时(ms)
PyDNet	6.35	12.22	4.70	212.93	14.45
PerUnet	17.49	30.85	4.24	236.06	14.39
SDy-CNN	6.56	7.10	1.07	930.37	15.19
WPFormer	26.73	48.45	4.21	237.63	3.95

问成本。相较于直筒型串行结构，并行分支需独立分配内存并频繁读写中间态特征图，引发了更高的底层算子调度与内核启动开销。其次，空洞卷积导致空间局部性退化。其跳跃式的特征采样方式打破了内存访问的连续性，降低了硬件缓存命中率，从而在物理计算层面拉长了实际的数据吞吐周期。然而，结合前文泛化性实验结果可知，本文方法牺牲了极少量的毫秒级延时，换取了高标准阈值(MP-CK@0.01)下准确率近8倍的提升(15.5% vs 1.90%)，以及绝对误差(MPJPE)的大幅下降。虽然当前212.93 帧/s的吞吐量是基于高性能GPU测试得出的理论上限，但结合其较低的FLOPs指标可预见，模型在算力受限的边缘设备上依然具备实现实时推理(通常为30 帧/s)的潜力。

在数据预处理阶段，各算法需完成CSI信号的时间戳对齐等操作。本文方法的预处理耗时为14.45 ms，与同类基于2D CNN 架构的基准模型(如PerUnet 的14.39 ms，SDy-CNN的15.19 ms)处于同一基准水平。综上分析，PyDNet在模型参数量这一核心轻量化指标上达到了最优，并在感知精度与推理效率之间实现了较好平衡。其计算开销特征契合无线感知技术向端侧轻量化发展的演进趋势。

#### 4.4.5 实验设置

本研究中的所有网络模型均基于PyTorch深度学习框架构建，使用单张NVIDIA RTX 4090 GPU (24 GB 显存)完成训练与推理。模型的初始学习率设定为 $1 \times 10^{-3}$ 。为保证模型在训练后期的稳定收敛，实验引入了基于验证集损失的动态学习率衰减策略：当模型在验证集上的损失连续50个Epochs未出现明显下降时，学习率将自动衰减为当前值的10%，具体参数如表8所示。

### 4.5 消融实验

#### 4.5.1 不同卷积分支率对肢体误差影响

为了进一步揭示网络多尺度空洞设计与人体空间拓扑结构之间的深层映射关系，并验证模型对复杂人体骨架信息的解耦能力，本文引入了推理期分

表8 实验参数设置

Tab. 8 Experimental parameter settings	
设置	设定值
优化器	Adam
初始学习率	$1 \times 10^{-3}$
学习率调度策略	ReduceLROnPlateau (factor=0.1, patience=50)
批处理大小	16
训练集/测试集/验证集划分比例	60%/20%/20%

支掩蔽分析。传统的重新训练消融方法可能会破坏网络已收敛的联合特征分布，相比之下，本文在严格冻结预训练模型权重的条件下，于测试阶段系统性地阻断特定尺度空洞率分支的特征传递。通过独立量化躯干节点(大面积的主干核心结构)与肢体节点(局部的末端细微动作)的平均单关节定位误差(MPJPE)变化，能够直观且精确地评估各感受野分支在物理空间中的独立贡献，实验结果如表9所示。

当屏蔽大空洞率分支时，网络的整体空间感知能力发生严重退化，其中躯干节点的定位误差激增了230.0%，增幅显著高于肢体节点(+170.3%)。大感受野负责捕捉空间中大面积的躯体轮廓与整体姿态。一旦切断该分支，网络便丧失了对人体核心拓扑结构的全局视野，导致肩、髋等核心关节的定位彻底失效。由于肢体在运动学树中依附于躯干，躯干的绝对位移不可避免地连带提升了肢体误差。这一极端的数据变化强有力地证明了 $d=3$ 分支在网络中全局视野的关键作用。

当切断小空洞率分支时，由于网络仍具备大分支的全局保底能力，整体误差的上升幅度相对收敛。然而，肢体节点的误差增幅(+36.8%)明显高于躯干节点(+32.7%)。这一现象印证了 $d=1$ 分支结构的小感受野聚焦于捕捉手腕、脚踝等肢体末端的局部空间特征，具备四肢细微空间位移的精确刻画能力。在屏蔽中等空洞率分支时，躯干误差(+40.5%)与肢体误差(+44.7%)的性能退化比例介于 $d=1$ 与 $d=3$ 的实验结果之间。这一居中的定量指标不仅符合预期，也从侧面证实了金字塔形空洞卷积这一设计成功实现了从局部末端细节到全局核心拓扑的连续、平滑的空间特征映射。

#### 4.5.2 多尺度特征提取模块对比

为了深入探究PyDNet架构中各核心组件对模型性能的具体贡献，本节以标准ResNet<sup>[28]</sup>为基准骨干，构建了3组变体模型进行多维度的对比实验。

(1) 基准模型：标准残差网络ResNet，作为性能参照的基准。

表 9 不同卷积分支率下MPJPE  
Tab. 9 MPJPE under different convolution branch rates

掩蔽分支(Masked Branch)	躯干误差(MPJPE)	肢体误差(MPJPE)	躯干误差增量( $\Delta\%$ )	肢体误差增量( $\Delta\%$ )
小空洞率( $d=1$ )	11.82	18.76	+32.7%	+36.8%
中空洞率( $d=2$ )	12.52	19.85	+40.5%	+44.7%
大空洞率( $d=3$ )	29.42	37.07	+230.0%	+170.3%
基线模型(无掩蔽)	8.91	13.72	—	—

(2) 多尺度特征验证模型：在基准上引入金字塔卷积但未采用空洞策略的PyConvNet，旨在剥离空洞卷积的影响，单独验证金字塔结构在多尺度特征提取中的作用。

(3) 主流特征提取机制对比模型：分别引入主流注意力机制的SE-ResNet<sup>[29]</sup>和非局部模块的NL-ResNet<sup>[30]</sup>，将本文提出的结构与先进特征提取模块进行对比。

表10、表11和表12分别详细列出了各模型在关键点检测准确率、关节位置误差及参数规模上的量化对比结果。实验数据表明，本文提出的PyDNet在所有评价指标上均显著优于基准模型及其他变体，充分验证了金字塔空洞卷积结构的有效性。

如表10和表11所示，与基准模型ResNet相比，PyDNet表现出显著性能优势，在PCK@0.1阈值下准确率达到94.96%，较基准模型的84.38%提升了10.58%；同时，其平均关节位置误差(MPJPE)从23.58大幅降低至11.90，降幅接近50%。这一结果有力证明了针对CSI信号非平稳特性设计的金字塔空洞结构，相比标准残差网络具备更强的特征提取与时空建模能力。对比PyConvNet与PyDNet的表现，可以分离并验证空洞卷积策略的关键作用。虽然PyConvNet通过引入多尺度卷积核已将PCK@0.10提升至92.22%，证明了多尺度特征对姿态估计的增益，PyDNet在此基础上进一步实现了2.74%的精度提升。同时，在更为严格的PCK@0.05标准下，PyDNet的优势扩大至8.28% (85.41%对比77.13%)。这表明单纯的多尺度卷积虽然增加了特征粒度，但受限于较小的感受野，难以捕捉人体肢体的长距离依赖关系；而PyDNet通过空洞卷积在不增加参数的前提下有效扩展了感受野，结合多尺度结构实现了对局部微动与全局姿态的协同感知。此外，与引入通用注意力机制的变体SE-ResNet和NL-ResNet相比，PyDNet同样表现出色。SE-ResNet虽然提升了部分性能，但仍落后于PyDNet，而NL-ResNet的表现甚至低于基准模型(PCK@0.10仅为79.29%)，这可能是由于非局部模块在处理高噪Wi-Fi信号时引入了无关干扰，反证了本文所提结构对特定物理信号的适应性优势。

表 10 消融实验中各模型的平均关键点准确率(MPCK)对比(%)  
Tab. 10 Comparison of Mean Percentage of Correct Keypoints (MPCK) in the ablation study (%)

模型	MPCK@0.10	MPCK@0.05	MPCK@0.01
PyDNet	<b>94.96</b>	<b>85.41</b>	<b>39.09</b>
PyConvNet	92.22	77.13	25.43
ResNet	84.38	62.92	10.35
SE-ResNet	92.21	76.60	23.63
NL-ResNet	79.29	55.62	9.45

注：表内加粗数值表示最优结果。

表 11 消融实验中各模型的平均关节位置误差(MPJPE) 对比  
Tab. 11 Comparison of Mean Per Joint Position Error (MPJPE) in the ablation study

模型	MPJPE
PyDNet	<b>11.90</b>
PyConvNet	16.03
ResNet	23.58
SE-ResNet	16.28
NL-ResNet	27.70

注：表内加粗数值表示最优结果。

表 12 消融实验中各模型的计算量对比  
Tab. 12 Comparison of model parameters in the ablation study

模型	参数量(M)
PyDNet	6.35
PyConvNet	6.44
SE-ResNet	28.04
NL-ResNet	38.23
ResNet	29.13

PyDNet在模型轻量化方面也展现出极高的效率。如表12所示，得益于分组卷积与瓶颈层的高效设计，PyDNet的参数量仅为6.35 M，约为基准模型ResNet(29.13 M)的21.8%。相比之下，引入注意力机制的SE-ResNet与NL-ResNet并未能有效控制模型复杂度，参数量分别为28.04 M和38.23 M。PyDNet在大幅压缩参数规模(减少约78%)的同时实现了最高检测精度。这种高精度、低冗余的特性证实了该方法并非依赖参数堆叠，而是源于高效的结构优化，使其具备在资源受限边缘设备上实时部署的巨大潜力。

## 5 讨论

### 5.1 隐私保护与数据伦理

针对Wi-Fi感知技术应用中可能引发的隐私保护与数据安全问题,本文在数据的采集、处理及存储流程中制定并实施了严格的管控机制。在数据脱敏与存储层面,实验同步采集的视觉视频仅作为辅助生成人体骨架真值的中间参考。在完成骨架标签标注后,原始视频文件已离线加密存储,从物理链路层面切断了数据向云端或外部网络泄露的途径。在敏感信息推断风险层面,尽管部分研究表明CSI信号包含步态等微观生物特征,具备反向推断人员身份、性别或环境布局的潜在能力,但相较于高分辨率视觉传感器,CSI信号无法捕获面部纹理与精确的身体轮廓细节。本文所构建的PyDNet算法架构仅聚焦于人体空间关键点坐标的回归提取,不包含任何身份识别或敏感属性分类模块,使得身份推断等隐私风险处于严格受控状态。此外,在数据伦理授权方面,本研究的所有数据采集工作均在志愿者完全知情并签署书面授权协议的前提下开展。相关数据的使用授权请通过邮件联系通讯作者。

### 5.2 局限性与未来工作

尽管本文提出的PyDNet在室内人体姿态估计任务中实现了感知精度与计算开销的有效平衡,并在受损信道下展现出较好的鲁棒性,但受限于单一无线信号的物理特性与轻量化架构的设计折中,本研究仍存在一定的局限性。未来工作将主要从以下两个维度展开深化。

(1) 硬件访存瓶颈与架构优化:由于当前实验室硬件平台与边缘侧测试环境的客观条件限制,本研究暂无法在真实的IoT边缘硬件上完成严谨的物理功耗测定与绝对推理延迟量化,现阶段主要通过参数量与浮点运算量进行理论潜力验证。尽管PyDNet的模型参数量较低,但其多分支金字塔并行结构增加了内存访问成本,在物理执行层面导致单帧推理延时有增加。未来工作将考虑引入结构化剪枝等轻量化计算技术,进一步压缩多分支网络的计算冗余与访存开销。

(2) 跨域泛化边界与模型外推性验证:现有实验数据主要源于固定的室内场景与特定开源数据集,受试样本多样性有限。模型在面对全新物理场景及复杂多径分布时的跨域鲁棒性仍需系统性验证。未来工作将构建跨场景、跨设备及跨人员划分的实验设置,以提升结论外推性。

## 6 结语

本文提出了一种用于Wi-Fi人体姿态估计的轻量级网络PyDNet,旨在解决复杂多径环境下人体行为多尺度特征提取困难以及现有模型参数量大的核心问题。通过设计PyDBlock核心模块,本研究创新性地结合了多尺度空洞卷积与残差学习机制,在不增加模型复杂度的前提下,成功实现了对人体细粒度动作与全局姿态的时空变化特征的协同建模。实验结果表明,PyDNet在精度与效率之间取得了显著优于现有方法的平衡。在阈值 $\alpha = 0.10$ 条件下,模型实现了94.96%的平均检测准确率,并将平均关节位置误差降至11.90,显著优于PerUnet和WPFormer等主流方法。模型参数量仅为6.35 M,仅为对比模型WPFormer的约1/4,极大地降低了计算开销。综上所述,PyDNet不仅验证了金字塔空洞结构在无线感知领域的有效性,更为未来在资源受限边缘设备上实现高精度、实时人体姿态估计提供了具备理论潜力的高效解决方案。

**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] CAO Zhe, HIDALGO G, SIMON T, *et al.* OpenPose: Realtime multi-person 2D pose estimation using part affinity fields[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(1): 172–186. doi: [10.1109/TPAMI.2019.2929257](https://doi.org/10.1109/TPAMI.2019.2929257).
- [2] TOSHEV A and SZEGEDY C. DeepPose: Human pose estimation via deep neural networks[C]. *The IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 2014: 1653–1660. doi: [10.1109/CVPR.2014.214](https://doi.org/10.1109/CVPR.2014.214).
- [3] MEHRABAN S, ADELI V, and TAATI B. MotionAGFormer: Enhancing 3D human pose estimation with a Transformer-GCNformer network[C]. *The IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, 2024: 6905–6915. doi: [10.1109/WACV57701.2024.00677](https://doi.org/10.1109/WACV57701.2024.00677).
- [4] AN Xiaoqi, ZHAO Lin, GONG Chen, *et al.* ShaRPose: Sparse high-resolution representation for human pose estimation[C]. *The AAAI Conference on Artificial Intelligence*, Vancouver, Canada, 2024: 691–699.
- [5] ZHAO Mingmin, LI Tianhong, ABU ALSHEIKH M, *et al.* Through-wall human pose estimation using radio signals[C]. *The 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).

- [6] 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*, 2024, 35(12): 18285–18299. doi: [10.1109/TNNLS.2023.3314031](https://doi.org/10.1109/TNNLS.2023.3314031).
- [7] ZHANG Rui, GENG Ruixu, LI Yadong, *et al.* RFMamba: Frequency-aware state space model for RF-based human-centric perception[C]. The Thirteenth International Conference on Learning Representations, Singapore, Singapore, 2025.
- [8] SENGUPTA A and CAO Siyang. *mmPose-NLP*: A natural language processing approach to precise skeletal pose estimation using mmWave radars[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(11): 8418–8429. doi: [10.1109/TNNLS.2022.3151101](https://doi.org/10.1109/TNNLS.2022.3151101).
- [9] SENGUPTA A, JIN Feng, ZHANG Renyuan, *et al.* mmPose: 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).
- [10] 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](https://doi.org/10.12000/JR24189).  
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](https://doi.org/10.12000/JR24189).
- [11] MA Yongsen, ZHOU Gang, and WANG Shuangquan. WiFi sensing with channel state information: A survey[J]. *ACM Computing Surveys (CSUR)*, 2020, 52(3): 46. doi: [10.1145/3310194](https://doi.org/10.1145/3310194).
- [12] WEI Bo, SONG Hang, KATTO J, *et al.* RSSI-CSI measurement and variation mitigation with commodity Wi-Fi device[J]. *IEEE Internet of Things Journal*, 2023, 10(7): 6249–6258. doi: [10.1109/JIOT.2022.3223525](https://doi.org/10.1109/JIOT.2022.3223525).
- [13] HALPERIN D, HU Wenjun, SHETH A, *et al.* Tool release: Gathering 802.11n traces with channel state information[J]. *ACM SIGCOMM Computer Communication Review*, 2011, 41(1): 53. doi: [10.1145/1925861.1925870](https://doi.org/10.1145/1925861.1925870).
- [14] WANG Fei, ZHOU Sanping, PANEV S, *et al.* Person-in-WiFi: Fine-grained person perception using WiFi[C]. The IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 5451–5460. doi: [10.1109/ICCV.2019.00555](https://doi.org/10.1109/ICCV.2019.00555).
- [15] HE Kaiming, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN[C]. The IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2980–2988. doi: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).
- [16] RONNEBERGER O, FISCHER P, and BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. The 18th International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany, 2015: 234–241. doi: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [17] YANG Jianfei, ZHOU Yunjiao, HUANG He, *et al.* MetaFi: Device-free pose estimation via commodity WiFi for metaverse avatar simulation[C]. The IEEE 8th World Forum on Internet of Things, Yokohama, Japan, 2022: 1–6. doi: [10.1109/WF-IoT54382.2022.10152057](https://doi.org/10.1109/WF-IoT54382.2022.10152057).
- [18] ZHOU Yue, ZHU Aichun, XU Caojie, *et al.* PerUnet: Deep signal channel attention in UNet for WiFi-based human pose estimation[J]. *IEEE Sensors Journal*, 2022, 22(20): 19750–19760. doi: [10.1109/JSEN.2022.3204607](https://doi.org/10.1109/JSEN.2022.3204607).
- [19] DENG Jie, CHEN Kaiqi, JING Pengsen, *et al.* CSI-channel spatial decomposition for WiFi-based human pose estimation[J]. *Electronics*, 2025, 14(4): 756. doi: [10.3390/electronics14040756](https://doi.org/10.3390/electronics14040756).
- [20] ZHOU Yunjiao, HUANG He, YUAN Shenghai, *et al.* MetaFi++: WiFi-enabled transformer-based human pose estimation for metaverse avatar simulation[J]. *IEEE Internet of Things Journal*, 2023, 10(16): 14128–14136. doi: [10.1109/JIOT.2023.3262940](https://doi.org/10.1109/JIOT.2023.3262940).
- [21] 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, UK, 2020: 23. doi: [10.1145/3372224.3380900](https://doi.org/10.1145/3372224.3380900).
- [22] GIAN T D, TRAN D T, PHAM Q V, *et al.* Multi-modal human pose estimation: A Wi-Fi-driven approach with adaptive kernel selection[J]. *IEEE Transactions on Artificial Intelligence*, 2025. doi: [10.1109/TAI.2025.3631005](https://doi.org/10.1109/TAI.2025.3631005).
- [23] GIAN T D, NGUYEN T H, NGUYEN N T, *et al.* WiLHPE: WiFi-enabled lightweight channel frequency dynamic convolution for HPE tasks[C]. The Tenth International Conference on Communications and Electronics, Danang, Vietnam, 2024: 516–521. doi: [10.1109/ICCE62051.2024.10634628](https://doi.org/10.1109/ICCE62051.2024.10634628).
- [24] NGUYEN X H, NGUYEN V D, LUU Q T, *et al.* Robust WiFi sensing-based human pose estimation using denoising autoencoder and CNN with dynamic subcarrier attention[J]. *IEEE Internet of Things Journal*, 2025, 12(11): 17066–17079. doi: [10.1109/JIOT.2025.3535156](https://doi.org/10.1109/JIOT.2025.3535156).
- [25] FANG Haoshu, LI Jiefeng, TANG Hongyang, *et al.* AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 7157–7173.
- [26] ZHOU Yue, XU Caojie, ZHAO Lu, *et al.* CSI-Former: Pay more attention to pose estimation with WiFi[J]. *Entropy*, 2023, 25(1): 20. doi: [10.3390/e25010020](https://doi.org/10.3390/e25010020).
- [27] HUANG Jinyang, FENG Yuanhao, CUI Fengqi, *et al.* Identifying who you are no matter what you write through

- abstracting handwriting style[J]. *IEEE Transactions on Dependable and Secure Computing*, 2026. doi: [10.1109/TDSC.2026.3668275](https://doi.org/10.1109/TDSC.2026.3668275).
- [28] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [29] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141. doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
- [30] WANG Xiaolong, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7794–7803. doi: [10.1109/CVPR.2018.00813](https://doi.org/10.1109/CVPR.2018.00813).

### 作者简介

刘 淼, 博士生, 主要研究方向为智能无线感知与物联网技术。

曾小路, 副教授, 主要研究方向为智能无线感知与物联网技术、穿墙雷达静止目标成像。

杨小鹏, 教授, 主要研究方向为生命雷达技术、穿墙雷达技术、探地雷达技术、相控阵雷达及自适应阵列信号处理。

邢程荐, 硕士生, 主要研究方向为智能无线感知与物联网技术。

刘 宇, 硕士生, 主要研究方向为Wi-Fi智能感知技术。

(责任编辑: 于青)