基于端到端和Mamba注意力融合网络的毫米波雷达跨人手势识别

方 超[①] 王 勇*[①] 周 牧[②] 杨小龙[①] 庞 宇[③]

[①](重庆邮电大学通信与信息工程学院 重庆 400065)

[②](重庆邮电大学电子科学与工程学院 重庆 400065)

[③](重庆邮电大学生命健康信息科学与工程学院 重庆 400065)

摘要：毫米波雷达作为一种非侵入式、非接触的传感设备，在人机交互、智能家居、虚拟现实等领域具有广阔应用前景而备受关注。现有深度学习模型由于其强大的特征提取能力，对训练用户的手势能实现很好的性能，当面临不同手势习惯、手部大小存在差异的新用户时，识别性能会出现显著退化。为提升模型在跨人场景下的泛化能力，该文提出一种融合端到端学习与状态空间模型的毫米波雷达手势识别网络。该方法直接以原始雷达数据立方体作为输入，通过嵌入Mamba模块在时空维度建模长程依赖关系，从而实现对不同用户手势特征的自适应提取与鲁棒表示。实验结果表明，所构建的端到端架构能够有效捕捉与用户无关的判别性手势模式。在跨人测试集上，该文方法在11折实验中取得94.28%的平均识别准确率和2.55%的标准差，最佳单折准确率为97.50%，显著优于传统深度学习方法，表明其在受控采集条件下具有较好的跨人识别鲁棒性。

关键词：注意力机制；端到端神经网络；手势识别；毫米波雷达；多域融合

中图分类号：TN957 **文献标识码：**A **文章编号：**2095-283X(2026)x-0001-14

DOI: 10.12000/JR25260 **CSTR**: 32380.14.JR25260

引用格式：方超, 王勇, 周牧, 等. 基于端到端和Mamba注意力融合网络的毫米波雷达跨人手势识别[J]. 雷达学报(中英文), 待出版. doi: 10.12000/JR25260.

Reference format: FANG Chao, WANG Yong, ZHOU Mu, *et al*. End-to-end cross-person gesture recognition via mamba fusion network and millimeter-wave radar[J]. *Journal of Radars*, in press. doi: 10.12000/JR25260.

End-to-end Cross-person Gesture Recognition Via Mamba Fusion Network and Millimeter-wave Radar

FANG Chao[①] WANG Yong*[①] ZHOU Mu[②] YANG Xiaolong[①] PANG Yu[③]

[①](*School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

[②](*School of Electronic Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

[③](*School of Life Health Information Science and Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

收稿日期：2025-12-03；改回日期：2026-06-23；网络出版：2026-xx-xx

*通信作者： 王勇 yongwang@cqupt.edu.cn *Corresponding Author: WANG Yong, yongwang@cqupt.edu.cn

基金项目：国家自然科学基金(52302059,62571074,62501100)，重庆市技术创新与应用发展重大专项(CSTB2025TIAD-STX0022)，重庆市教育委员会科学技术研究计划(KJQN202400616)，新重庆青年创新人才项目(CSTB2025YITP-QCRCX0100)

Foundation Items: The National Natural Science Foundation of China (52302059, 62571074, 62501100), The Chongqing Major Project of Technological Innovation and Application Development (CSTB2025TIAD-STX0022), The Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202400616), The New Chongqing Youth Innovation Talent Project (CSTB2025YITP-QCRCX0100)

责任主编：方震 Corresponding Editor: FANG Zhen

Abstract: As a noninvasive and contactless sensing technology, millimeter-wave radar has attracted considerable attention because of its broad application potential in human-computer interaction, smart homes, and virtual reality. Existing deep learning models achieve strong performance in recognizing gestures from trained users owing to their powerful feature extraction capabilities; however, their recognition accuracy degrades significantly when applied to new users with different gesture habits and hand sizes. To improve model generalization in cross-user scenarios, this paper proposes a millimeter-wave radar gesture recognition network that integrates end-to-end learning with a state space model. The proposed method directly processes raw radar data cubes and incorporates a Mamba module to capture long-range spatiotemporal dependencies. This enables the adaptive extraction and robust representation of user-independent gesture features. Experimental results show that the proposed end-to-end architecture effectively captures discriminative gesture patterns that are invariant across users. On the cross-user test set, the proposed method achieved an average recognition accuracy of 94.28% with a standard deviation of 2.55% across 11 folds, while the highest single-fold accuracy reached 97.50%. These results substantially outperform those of conventional deep learning methods and validate the generalization capability of the proposed method in cross-user application scenarios.

Key words: Attention mechanism; End-to-end neural network; Gesture recognition; Millimeter-wave radar; Multi-domain Fusion

1 引言

近年来，基于毫米波雷达的人体手势识别在人机交互[1]、智能家居[2]等领域受到广泛关注。手势动作通常伴随距离、速度和角度等物理量的连续变化，而毫米波雷达能够对这些信息进行高精度测量[3]，因此已被广泛应用于人机交互、跌倒检测和行人活动分类等任务[4–6]。现有手势识别方法主要包括可穿戴式和非接触式两类。可穿戴传感器[7,8]能够提供较高精度的运动数据，但需要用户持续佩戴，容易造成使用不便并降低用户体验。非接触式方法无需附着于人体，更适合自然交互场景。其中，基于摄像头的方法[9]容易受到光照、遮挡和视场范围的影响，同时存在隐私泄露风险。相比之下，毫米波雷达[10]无需用户佩戴设备，对光照条件不敏感，且不直接采集人体外观图像，具有较好的隐私保护能力。此外，毫米波雷达还具有一定穿透能力，能够在弱光或部分遮挡环境下感知手部运动。因此，基于毫米波雷达的人体手势识别为实现自然、可靠和隐私友好的人机交互提供了有效技术途径。

许多基于毫米波的人体手势识别方法被提出[11–13]。这些方法将雷达回波转换成点云和微多普勒，然后将其输入到深度神经网络中以获得满意的结果。然而，微多普勒、点云以及距离-多普勒等传统雷达特征表示通常依赖预设的信号处理流程和人工设定的参数。尽管这些方法具有明确的物理意义和良好的可解释性，能够有效表征手势的运动特征，但其固定的处理流程可能难以针对跨人手势识别任务自适应保留和强化最具判别性的任务相关信息。众所周知，输入越丰富，识别精度就越高。因此，一些研究人员研究了多域雷达信息用于人体手势识别。例如，文献[14,15]从人体手势反射的雷达回波通常被处理成距离-多普勒(Range-Doppler, RD)图、时间-距离(Time-Range, TR)图、时间-多普勒(Time-Doppler, TD)图和时间-频率(Time-Frequency, TF)图，并构建多域融合识别网络，以实现对人体手势的有效识别。

然而，上述方法通常将雷达信号预处理与识别网络分开设计。该处理范式虽然具有清晰的物理解释，但预处理阶段的参数和特征映射方式通常独立于下游识别目标进行设定，可能限制模型对任务相关手势信息的自适应建模与充分利用。为了克服这一缺陷，近年来部分研究开始探索端到端的人体手势识别方法。该方法所面临的主要挑战在于原始雷达数据是复数，因此幅度和相位都包含重要的手势信息。而在以往的端到端雷达识别工作中，部分方法直接使用2D sinc滤波器对原始雷达数据进行特征学习[16]，另一些方法将复数雷达数据的实部和虚部分别作为独立通道进行处理[17]，这些方法不仅造成信息缺失，还限制了模型在跨人部署场景下的鲁棒性。

现有多域融合模型的关键组成部分是卷积运算，它擅长捕捉局部特征，但无法表示手势中的长距离依赖关系，从而导致重要的全局上下文信息提取不足。对于多类型数据的融合[18]，必须设计更有效的融合策略，以利用每个数据源的独特特性，实现精确而全面的手势分类。此外，当前的卷积或基于注意力机制的主干网络难以在动态手势和用户变化场景中捕捉通用的局部-全局特征信息。当训练

好的模型应用于未见过的用户时，识别性能通常会下降。

为解决上述局限性与挑战，本文提出了一种基于毫米波雷达的新型端到端跨人手势识别系统，该系统利用多维特征表示和Mamba注意力融合网络(MambaFuse) 进行手势识别。本文选用距离-多普勒-方位角(Range-Doppler-Angle, RDA)与RD两种雷达数据立方序列进行手势识别，通过可学习的预处理模块将雷达回波信号处理为雷达数据立方。该系统采用多尺度特征模块，从RDA和RD立方体序列中捕获多尺度的局部距离、多普勒、方位角特征。在动态手势识别系统内部，本文创新性地提出Mamba注意力模块，该模块采用的Mamba注意力的Transformer架构，其核心是用mamba替代传统自注意力机制。该改进在保证低计算复杂度的同时显著提升了识别精度。本文在包含11名志愿者执行8类典型手势的原始雷达数据集进行了实验，验证了所提MambaFuse方法的性能与可行性。本文的主要贡献如下：

(1) 本文提出了一种用于毫米波雷达跨人手势识别的端到端多域特征融合网络，其中可学习的预处理模块旨在从原始雷达回波中自适应地生成多维RDA和RD数据立方体，消除了对手工信号处理的依赖，并使网络能够更关注于特定任务的雷达信号。

(2) 本文设计了一个双分支多尺度特征提取模块，从RDA和RD立方体中提取判别性浅层特征，以实现稳健的手势表示。另外，设计了基于Mamba注意力的轻量级Transformer模块，它用Mamba取代了传统的自注意力机制。这种设计显著降低了计算复杂度，同时保留了捕获长程依赖关系的能力。另外，本文还设计了一种新颖的跨域注意力融合策略进一步以跨域融合交互方式融合互补特征。

(3) 在自收集的原始雷达数据上进行实验，结果表明本文方法在11折实验中取得94.28%的平均识别准确率和2.55%的标准差，最佳单折准确率为97.50%，展示了本文方法的鲁棒性和有效性。

2　RD和RDA立方体序列构建

毫米波雷达具有高频、易于集成等优势，能够感知人体手势状态。不同的人体手势在雷达回波中表现出不同的特征。图1展示了雷达回波信号的生成过程，这些信号将作为所提出的手势系统的输入数据。

雷达通过天线发射连续的调频波信号，经人体目标反射后由接收天线接收。发射信号和接收信号可以表示为

$$s_{\mathrm{t}}(t)=A_{\mathrm{t}}\exp\left[\mathrm{j}\left(2\pi f_{\mathrm{c}}t+\pi kt^{2}+\varphi_{0}\right)\right] \tag{1}$$

$$s_{\mathrm{r}}(t)=A_{\mathrm{r}}\exp\left[\mathrm{j}\left(2\pi f_{\mathrm{c}}(t-\tau)+\pi k(t-\tau)^{2}+\varphi_{0}\right)\right] \tag{2}$$

其中，A_{t}和A_{r}分别表示发射信号和接收信号的幅度；$s_{\mathrm{t}}(t)$和$s_{\mathrm{r}}(t)$分别表示发射信号和接收信号；φ_0表示发射信号的初始相位；f_{c}表示初始频率，k为线性调频信号的斜率，τ表示信号从发射到接收所经历的时间。电磁波照射到目标物体后反射回来的信号称为回波信号。将发射信号与接收信号进行混频处理，并将混频后的信号通过低通滤波器，即可生成中频(Intermediate Frequency, IF)信号。因此，中频信号可表示为

$$\begin{aligned}s_{\mathrm{IF}}(t)&=\mathrm{LPF}\left\{s_{\mathrm{t}}(t)s_{\mathrm{r}}^{*}(t)\right\}\\&=A_{\mathrm{t}}A_{\mathrm{r}}\exp\left\{\mathrm{j}2\pi\left(f_{\mathrm{c}}\tau+k\tau t-\frac{1}{2}k\tau^{2}\right)\right\}\end{aligned} \tag{3}$$

其中，$s_{\mathrm{IF}}(t)$为中频信号。中频信号经模数转换采样后，数字信号可进一步用于提取距离和多普勒信息。对于FMCW雷达，目标距离可由中频信号的频率估计得到。具体而言，线性调频斜率为

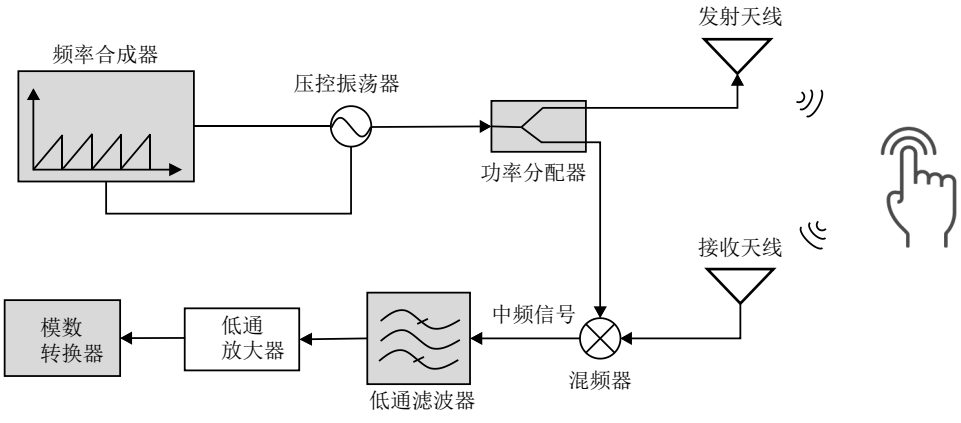


图 1 毫米波雷达系统架构

Fig. 1 Architecture of the millimeter-wave radar system

$k = B/T_c$，往返传播时延为$\tau = 2R/\mathrm{c}$，因此目标距离可表示为

$$R = \frac{\mathrm{c}f_{\mathrm{IF}}}{2k} = \frac{\mathrm{c}T_{\mathrm{c}}f_{\mathrm{IF}}}{2B} \tag{4}$$

其中，c为光速，T_{c}为线性调频脉冲的周期，B为带宽，f_{IF}为中频信号的频率，R表示目标物体与雷达之间的距离。f_{IF}可由中频信号的瞬时相位求导获得

$$f_{\mathrm{IF}} = \frac{\mathrm{d}\left(f_{\mathrm{c}}\tau + k\tau t - \frac{1}{2}k\tau^2\right)}{\mathrm{d}t} = k\tau \tag{5}$$

通过研究相邻chirp之间的相位变化，可以估计手势相对于雷达的径向速度，其表达式为

$$v = \frac{\lambda\Delta\phi}{4\pi T_{\mathrm{c}}} \tag{6}$$

其中，λ是发射信号的波长，$\Delta\phi$是脉冲之间的相位差。由于手势动作是由手指、手掌和手腕等多个局部散射点共同形成的非刚体动态运动，不同散射点相对于雷达具有随时间变化的径向速度，因此会在多普勒维度上产生时变的微多普勒频移。微多普勒频移可表示为

$$f_{\mathrm{md}}(t) = \frac{2v_{\mathrm{r}}(t)}{\lambda} \tag{7}$$

其中，$f_{\mathrm{md}}(t)$表示微多普勒频移，$v_{\mathrm{r}}(t)$表示手部局部散射点相对于雷达的瞬时径向速度。

由于FMCW 毫米波雷达接收天线之间存在距离，同一目标的回波信号在不同接收天线之间会产生路径差异，从而产生相位差。根据该相位差，可以推导出目标的角度。本文假设接收天线沿方位角方向构成均匀线性阵列，相邻接收天线间距为l。在远场单个距离－多普勒分辨单元内的主导散射中心下，目标回波可近似为平面波入射到接收阵列。对于入射方位角θ的目标，相邻接收天线之间的路径差为$l\sin\theta$，由此产生的相位差可表示为$\Delta\phi_{\mathrm{RX}} = 2\pi l\sin\theta/\lambda$。因此，可以通过提取接收天线间中频信号的相位差来获取方位角，该角度可表示为

$$\theta = \sin^{-1}\left(\frac{\lambda\Delta\phi_{\mathrm{RX}}}{2\pi l}\right) \tag{8}$$

其中，ϕ_{RX}为接收天线间的相位差，l为接收天线间距。

对于接收到的中频信号，可以使用雷达成像算法将人体手势数据转换为距离、多普勒和角度特征。然而，使用雷达成像算法本身存在计算成本，并且需要进行一些归一化处理。一个好的雷达人体手势识别系统应该能够利用时域中非归一化的输入，并学习适合后续下游任务的变换。端到端人体手势识别网络可以满足解决这一问题，例如通过一个复数加权可学习线性层[19]来取代传统的预处理过程。因此，该线性层模块不仅模拟了标准离散傅里叶变换 (Discrete Fourier Transform, DFT) 的动作，还能学习最优权重。传统FFT 流程与本文可学习线性层在处理维度上是一一对应的，但区别在于传统处理核固定，而本文线性层由 DFT 初始化后可通过识别损失端到端更新。各层的初始化权重如式(9)所示，其值对应于标准 DFT。

$$w(a,b) = \exp\left(-\mathrm{j}\frac{2\pi}{N}ab\right), \begin{cases} 0 \le a \le M-1 \\ 0 \le b \le N-1 \end{cases} \tag{9}$$

其中，M表示线性层的维度,即输入序列的长度，N表示离散傅里叶变换的点数。为了获取距离、多普勒和角度信息，本文对中频信号进行距离线性层、多普勒线性层、方位角线性层操作，生成RDA和RD数据立方体序列[20]。由于人体手势是一个连续的过程，每帧数据都相互关联。因此，需要将每帧数据连接在一起，以捕捉时间特征信息。本文记录了每个动作的开始时间，选择由32帧组成的序列来表示每个完整的手势动作，在40 ms帧周期下，总时长为1.28 s。

3 基于端到端的MambaFuse跨人手势识别网络

本节详细描述所提出的具有多域融合的MambaFuse模型。如图2所示，MambaFuse的整个模型框架可分为4部分：多尺度特征提取模块、通道注意力模块、Mamba注意力模块和交叉注意力融合模块。

3.1 框架概述

为进一步清晰展示所提出 MambaFuse 网络的整体结构，图2 给出了完整的网络架构示意图。整体流程包括可学习预处理、双分支多尺度特征提取、通道注意力增强、Mamba 注意力建模、交叉注意力融合以及最终分类6个阶段。首先，雷达回波可以表示为由快时间采样点、慢时间chirp序列和天线通道构成的多维数据。其中，快时间采样点对应每个chirp内的ADC采样维度，可用于提取距离信息。慢时间chirp序列表示同一帧内连续发射的多个chirp，利用相邻chirp之间的相位变化，可以提取速度信息。虚拟天线是由多个发射天线和多个接收天线组合后等效形成的虚拟阵元，利用与目标到达方向相关的相位差，可以提取目标的方位角信息。本文将雷达回波信号构成的多维数据输入可学习预处理模块，生成RDA和RD数据立方体序列，RDA数据能够描述手势在距离—速度—方位

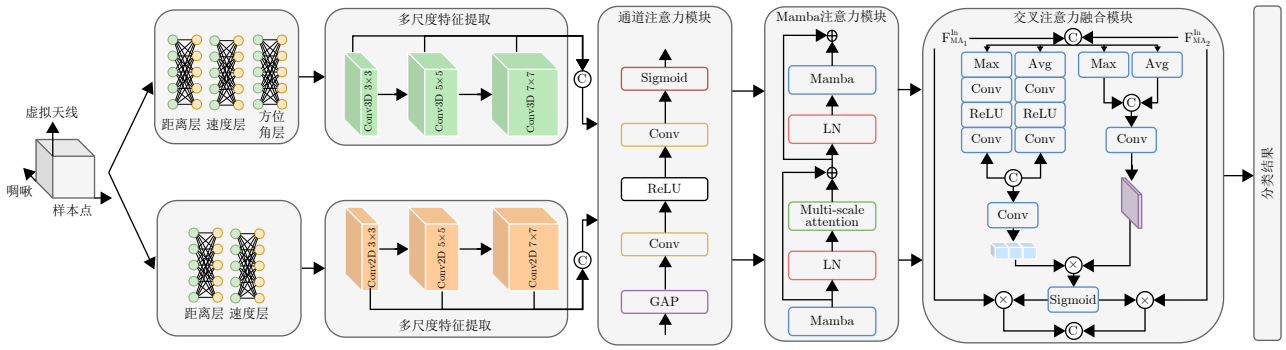


图 2 所提MambaFuse手势识别模型总体框架

Fig. 2 Overall framework of the proposed MambaFuse gesture recognition model

角空间中的全局运动分布，RD数据更突出手势动作中的局部运动分布。随后，RDA与RD数据分别输入两个并行多尺度特征提取分支，以捕获不同尺度雷达信号中的互补局部手势特征。之后，通道注意力模块用于突出关键特征通道，Mamba注意力模块进一步建模长程全局时空依赖关系。最后，交叉注意力融合模块对两个分支的特征进行全局-局部交互式融合，并将融合特征输入分类器以获得最终识别结果。

3.2 多尺度特征提取模块

给定两个输入数据$\mathbf{I}_1 \in \mathrm{R}^{T\times R\times D\times A}$和$\mathbf{I}_2 \in \mathrm{R}^{T\times R\times D}$，其中$R$, D, A和T分别代表距离、多普勒、方位角和时间信息。利用卷积神经网络在空间上下文特征提取方面的卓越能力，将其与雷达立方体序列相结合可充分发挥其在特征处理中的优势。该技术充分利用距离、速度、方位角信息，有效解决RDA与RD立方体序列中的空间关联性问题。本文所提网络首先采用多尺度特征提取捕获浅层局部抽象特征。如图2所示，为了充分利用来自不同雷达表示的多尺度互补手势信息，以分层多尺度方式设计了两个并行的特征提取分支，即RDA分支和RD分支。RDA分支利用时间分布式3D卷积提取距离-多普勒-角度体空间特征，RD分支利用时间分布式2D卷积提取距离-多普勒平面特征。时间维度在卷积特征提取过程中被保留，并随后由Mamba注意力模块进行建模。该设计将时间分布式卷积提取的局部空间特征与基于Mamba的全局时序特征相结合，从而捕获动态手势中的局部-全局特征依赖关系。在每个分支中，使用3个连续的特征提取块逐步提取手势特征，这些特征提取块的核大小和特征图维度不断增加，通过逐步增加卷积层通道数(从16到64再到256)，模型能够学习更复杂的高层特征。首先，将输入的RDA和RD数据立方体分别输入到第1个特征提取块中，以捕获浅层时空特征，其表示为

$$\mathbf{F}_{\mathbf{fem}}^{1} = \mathbf{Re}\left(\mathbf{BN}^{1}\left(\mathbf{Co}_{3}^{1,16}\left(\mathbf{I}_{z}\right)\right)\right) \tag{10}$$

其中，$\mathbf{I}_z$表示输入雷达立方体，即$\mathbf{I}_1$和$\mathbf{I}_2$。$\mathbf{F}_{\mathrm{fem}}^{i}$ $(i \in [1,3])$表示从第i个特征提取块提取的特征表示。其次，$\mathbf{F}_{\mathbf{fem}}^{2}$被输入到第2个特征提取模块，该模块具有更大的核大小，以捕获中层时空特征，其公式为

$$\mathbf{F}_{\mathbf{fem}}^{2} = \mathbf{Re}\left(\mathbf{BN}\left(\mathbf{Co}_{5}^{16,64}\left(\mathbf{F}_{\mathbf{fem}}^{1}\right)\right)\right) \tag{11}$$

最终，被传递至第3特征提取块，该模块采用最大尺度的特征图以捕获大尺度特征。该过程可表述为

$$\mathbf{F}_{\mathbf{fem}}^{3} = \mathbf{Re}\left(\mathbf{BN}\left(\mathbf{Co}_{7}^{64,256}\left(\mathbf{F}_{\mathbf{fem}}^{2}\right)\right)\right) \tag{12}$$

随后，采用通道注意力模块以聚焦感知关键的多尺度特征，该过程可表述为

$$\mathbf{F}_{\mathbf{fem}}^{i^*} = \mathrm{Sig}\left(\mathrm{Co}_{1}^{n_i/r_i,n_i}\left(\mathrm{Re}\left(\mathrm{Co}_{1}^{n_i,n_i/r_i}\left(\mathrm{G}\left(\mathbf{F}_{\mathbf{fem}}^{i}\right)\right)\right)\right)\right) \times \mathbf{F}_{\mathbf{fem}}^{i} \tag{13}$$

其中，$\mathrm{G}(\cdot)$表示全局平均池化操作，$\mathbf{F}_{\mathbf{fem}}^{i^*}$代表通道注意力模块的输出特征图。多尺度模块的作用并不是建模不同目标距离，而是在固定距离条件下增强模型对不同手势动作尺度差异的特征提取能力，从而更好的提取不同感受野下的多粒度RD和RDA手势运动特征。

3.3 Mamba注意力模块

状态空间模型(State Space Models, SSMs)[21]的基本概念源于连续线性时不变系统。该模型以一维信号$x(t) \in \mathrm{R}$作为输入，旨在通过一个维数为N的隐含状态$\mathbf{h}(t) \in \mathbf{R}^N$将其映射为输出序列$y(t) \in \mathrm{R}$。此映射过程可通过以下线性常微分方程进行表述：

$$\begin{aligned} \mathbf{h}'(t) &= \mathbf{A}\mathbf{h}(t) + \mathbf{B}x(t) \\ y(t) &= \mathbf{C}\mathbf{h}(t) \end{aligned} \tag{14}$$

其中，$\mathbf{h}'(t) \in \mathrm{R}^N$表示隐含状态$\mathbf{h}(t)$的时间导数，$\mathbf{A} \in$

$\mathrm{R}^{N\times N}$为状态转移矩阵，而$\mathbf{B}\in\mathbf{R}^{N\times 1}$与$\mathbf{C}\in\mathbf{R}^{N\times 1}$则分别为输入与输出的投影矩阵。

由式(14)所描述的连续时间系统，在集成至基于离散序列的深度学习模型时，通常面临兼容性挑战。为此，研究随后采用了时间尺度参数为Δ的零阶保持器技术，以促成一种直接的离散化转换。该步骤将连续参数$\mathbf{A}$和$\mathbf{B}$转化为其离散形式的对应参数$\bar{\mathbf{A}}$和$\bar{\mathbf{B}}$。

$$\begin{aligned}\bar{\mathbf{A}}&=\exp(\Delta\mathbf{A})\\ \bar{\mathbf{B}}&=(\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A})-\mathbf{I})\cdot\Delta\mathbf{B}\end{aligned}\tag{15}$$

详细的离散化过程可参阅文献[22]。经离散化后，该状态空间系统可表述为

$$\begin{aligned}\mathbf{h}_{\mathrm{t}}&=\bar{\mathbf{A}}\mathbf{h}_{t-1}+\bar{\mathbf{B}}x_{\mathrm{t}}\\ y_{\mathrm{t}}&=\mathbf{C}\mathbf{h}_{\mathrm{t}}\end{aligned}\tag{16}$$

为提升模型的计算效率与可扩展性，本研究采用卷积运算 $*$ 以加速上述线性递推过程。因此，系统的最终输出可被整合表示为

$$\begin{aligned}\bar{\mathbf{K}}&=\left(\mathbf{C}\bar{\mathbf{B}},\mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}},\cdots,\mathbf{C}\bar{\mathbf{A}}^{L-1}\bar{\mathbf{B}}\right)\\ \mathbf{y}&=\mathbf{x}*\bar{\mathbf{K}}\end{aligned}\tag{17}$$

其中，L表示输入序列的长度，而$\bar{\mathbf{K}}\in\mathrm{R}^{L}$则作为结构化的卷积核。

传统状态空间模型主要基于线性时不变的简化假设。该假设虽享有线性时间复杂度的优势，但难以有效捕捉输入序列中的上下文信息。为突破此局限，Mamba引入了一种选择机制，并提出了选择性状态空间模型，以实现序列状态间的动态交互。Mamba模块的详细架构如图3所示。与采用静态参数化的传统SSMs不同，S6模型使投影矩阵能够依据输入进行动态调整，从而实现对每个序列单元的选择性注意力聚焦。具体而言，参数$\mathbf{B}$, $\mathbf{C}$及时间间隔Δ均由输入序列$\mathbf{x}\in\mathbf{R}^{B\times L\times D}$动态投影生成，其过程可表述为

$$\begin{aligned}\mathbf{B}&=\mathrm{Projection}_B(\mathbf{x})\\ \mathbf{C}&=\mathrm{Projection}_C(\mathbf{x})\\ \Delta&=\tau_\Delta(\mathrm{Parameter}+\mathrm{Projection}_\Delta(\mathbf{x}))\end{aligned}\tag{18}$$

其中，$\mathbf{B}\in\mathbf{R}^{B\times L\times D}$, $\mathbf{C}\in\mathbf{R}^{B\times L\times D}$和$\Delta\in\mathrm{R}^{B\times L\times D}$。函数$\mathrm{Projection}_B(\cdot)$和$\mathrm{Projection}_C(\cdot)$执行向$N$维空间的线性投影。$\tau_\Delta$表示Softplus激活函数。该选择性机制使Mamba能够有效滤除时序任务中的无关噪声，同时选择性地保留或丢弃与当前输入相关的信息。

Transformer虽能有效捕获长程上下文信息并广泛应用于识别任务，但其核心的自注意力机制存在二次计算复杂度问题，且难以捕捉局部上下文信息。为解决上述问题，本文设计了一种新颖的Mamba注意力网络，该网络使用Mamba模块[22]替代Transformer中的自注意力机制，不仅能够捕获全局上下文信息，同时兼顾了模型的有效性和计算效率。具体而言，图2中的Mamba注意力模块以图3所示的Mamba模块作为核心序列建模单元，并结合层归一化、残差连接和MLP前馈映射构成完整的特征增强结构，其计算过程可表述为

$$\mathbf{F}_{\mathbf{AD}}^{i}=\mathbf{MAMBA}\left(\mathbf{LN}\left(\mathbf{F}_{\mathbf{fem}}^{i^*}\right)\right)+\mathbf{F}_{\mathbf{fem}}^{i^*}\tag{19}$$

$$\mathbf{F}_{\mathbf{Mamba}}^{i}=\mathbf{MLP}\left(\mathbf{LN}\left(\mathbf{F}_{\mathbf{AD}}\right)\right)+\mathbf{F}_{\mathbf{AD}}\tag{20}$$

其中，$\mathbf{F}_{\mathbf{AD}}^{i}$表示$\mathbf{F}_{\mathbf{fem}}^{i^*}$与Mamba运算结果之和，$\mathrm{LN}(\cdot)$代表层归一化操作，$\mathbf{F}_{\mathbf{Mamba}}^{i}$为经Mamba注意力模块处理后的最终输出。对于雷达跨人手势识别任务，Mamba模块的输入为经过多尺度特征提取和通道注意力优化后的RDA和RD时序特征。这些特征保留了手势执行过程中连续帧之间的时间关系。动态手势通常由距离、多普勒速度和方位角的联合变化来体现，因此仅依赖卷积操作可能难以充分捕获跨帧长程依赖。Mamba通过其选择性状态空间机制对输入序列进行自适应建模，保留对手势类别

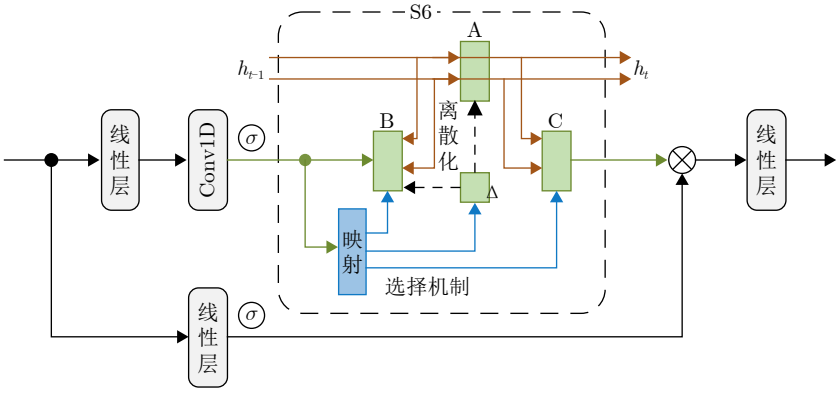


图 3 Mamba模块的架构

Fig. 3 The architecture of the mamba module

具有判别作用的运动模式，同时减少局部冗余响应。由此，Mamba模块能够对局部RDA和RD特征进行全局时序优化，并使网络学习更加鲁棒的用户无关手势特征。

3.4 跨域注意力融合模块

本研究提出的跨域注意力融合模块(Cross-domain Attention Fusion Module, CAFM)受卷积块注意力模块(Convolutional Block Attention Module, CBAM)[23]启发，旨在融合来自不同域的本质特征。与CBAM不同，本模块以两个卷积层及一个拼接操作取代了原有的共享多层感知机与加法层。值得注意的是，本文还为识别网络重新设计了一种并行注意力结构。

在通道独立分支中，首先对融合特征并行施加最大池化与平均池化操作，以生成初始通道注意力向量。随后，这些向量经由两个卷积层与PReLU激活函数处理，经拼接操作后通过最终卷积层生成通道注意力向量$\mathbf{F}_{\mathrm{F}}^{\mathbf{ca}} \in \mathbf{R}^{T\times E}$。

$$\begin{aligned}\mathbf{F}_{\mathbf{f}}^{\mathbf{ca}} = \mathrm{Co}_3\left(\left[\mathrm{Co}_3\left(\mathrm{Re}\left(\mathrm{Co}_3\left(\mathrm{MP}\left(\mathbf{F}_{\mathrm{fu}}\right)\right)\right)\right),\right.\right.\\ \left.\left.\mathrm{Co}_3\left(\mathrm{Re}\left(\mathrm{Co}_3\left(\mathrm{AP}\left(\mathbf{F}_{\mathrm{fu}}\right)\right)\right)\right)\right]\right)\end{aligned} \tag{21}$$

其中，$\mathrm{MP}(\cdot)$和$\mathrm{AP}(\cdot)$分别表示最大池化与平均池化操作，$[\cdot]$表示特征拼接操作，$\mathbf{F}_{\mathrm{fu}}$则代表$\mathbf{F}_{\mathbf{Mamba}}^{1}$和$\mathbf{F}_{\mathbf{Mamba}}^{2}$的初始融合结果。

类似地，在空间独立分支中，同样采用最大池化与平均池化以生成初始空间注意力图谱，这些图谱经拼接后通过卷积层处理，最终生成空间注意力图谱$\mathbf{F}_{\mathbf{F}}^{\mathbf{spa}}$，其计算过程表述为

$$\mathbf{F}_{\mathbf{f}}^{\mathbf{spa}} = \mathbf{Co}_3\left(\left[\mathbf{MP}\left(\mathbf{F}_{\mathbf{fu}}\right), \mathbf{AP}\left(\mathbf{F}_{\mathbf{fu}}\right)\right]\right) \tag{22}$$

随后，采用Sigmoid激活函数生成注意力权重图谱，其计算公式为

$$\mathbf{W} = \mathbf{Sig}\left(\mathbf{F}_{\mathbf{f}}^{\mathbf{ca}} \times \mathbf{F}_{\mathbf{f}}^{\mathbf{spa}}\right) \tag{23}$$

接下来，将注意力权重$\mathbf{W}$与$(1-\mathbf{W})$分别分配给RDA分支与RD分支，并通过如下计算得到最终融合特征$\mathbf{F}_{\mathrm{F}} \in \mathrm{R}^{T\times E}$。

$$\mathbf{F}_{\mathrm{F}} = \left[\mathbf{W} \times \mathbf{F}_{\mathbf{Mamba}}^{1}, (1-\mathbf{W}) \times \mathbf{F}_{\mathbf{Mamba}}^{2}\right] \tag{24}$$

最终，采用Softmax函数获取最终识别结果$\mathbf{Y}_{\mathrm{F}}$。

4 实验结果与分析

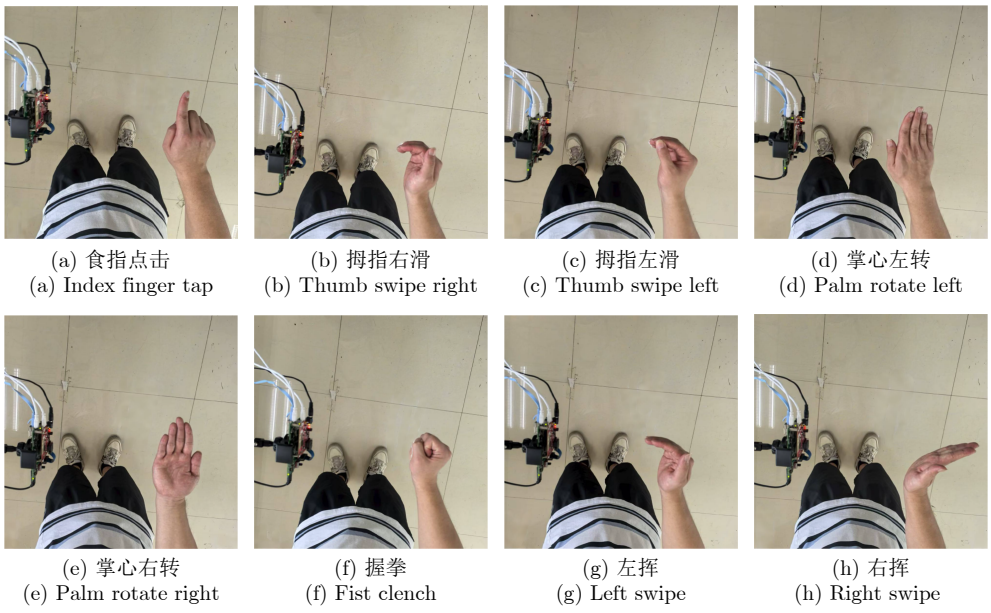
4.1 实验数据

本文数据集通过定制的毫米波雷达传感平台采集，该平台集成德州仪器AWR1642雷达与DCA1000EVM数据采集模块。详细雷达配置与系统参数汇总于表1，采集手势示例如图4所示。数据

表 1　毫米波雷达参数配置

Tab. 1　Millimeter-wave radar parameter configuration

参数	数值
开始频率	77 GHz
调频斜率	98 MHz/us
ADC采样点	128
调频带宽	3.92 GHz
帧周期	40 ms
每帧chirp数	128
调频脉冲周期	40 us



(a) 食指点击 (a) Index finger tap　(b) 拇指右滑 (b) Thumb swipe right　(c) 拇指左滑 (c) Thumb swipe left　(d) 掌心左转 (d) Palm rotate left

(e) 掌心右转 (e) Palm rotate right　(f) 握拳 (f) Fist clench　(g) 左挥 (g) Left swipe　(h) 右挥 (h) Right swipe

图 4 本文采集的雷达手势数据集

Fig. 4 Radar gesture dataset collected in this study

集包含8类手势：食指点击、拇指右滑、拇指左滑、掌心左转、掌心右转、握拳、左挥、右挥。除手势1外，所有手势均要求掌心朝向雷达传感器，且每次动作前手掌距传感器25 cm。这些手势因其在人机交互与智能家居控制领域的应用价值而被选取。总共11名志愿者(4名女性、7名男性，年龄22～40岁)参与数据采集，每位受试者对每个手势重复20次，最终构建包含1760个样本的数据集用于模型训练与测试，该数据集主要是面向青年跨人手势识别，尚不能充分覆盖儿童、青少年或老年用户的手势差异，未来工作将通过纳入更大年龄跨度的参与者来扩展数据集。在训练阶段，本文随机选择10名志愿者的数据用于训练和验证，其中用80%手势数据用于训练，20%用于验证；在测试阶段，用未被选择的1名志愿者所有样本进行测试。

本文在NVIDIA RTX A6000 GPU上基于PyTorch框架训练所有模型。训练周期设置为60轮，采用Adam优化器对模型进行优化。学习率固定为0.0001。

4.2 评估指标

关于性能评估，本研究采用分类准确率作为评估指标，该指标在动态手势识别任务中已被广泛使用。其计算公式为

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100\% \tag{25}$$

其中，N_{correct}表示正确预测的样本数量，N_{total}代表样本总数。此外，为评估模型稳定性，本文还采用标准差作为辅助指标，其计算方式为

$$\text{STD} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{26}$$

其中，n表示所有手势样本的总数，x_i代表单个手势的识别准确率，$\bar{x}$则为所有手势准确率的平均值。

4.3 网络性能分析

为充分验证本文提出的MambaFuse模型识别性能，本文将其与3种先进模型进行对比，即DSTFF [24], PLCN [25]和DCS-CTN [26]。为保证对比的公平性，为充分验证本文提出的MambaFuse模型的识别性能，本文选取 DSTFF, PLCN和DCS-CTN 3种先进方法作为对比模型。为保证比较的公平性，所有对比方法均在本文自采集的毫米波雷达跨人手势数据集上进行重新训练与测试，并采用相同的训练/验证/测试划分、相同的跨人测试协议和相同的评价指标。对于各对比方法，本文尽可能保持其原文献中的网络结构、优化策略和关键超参数设置，仅根据本文数据集的输入维度和手势类别数量对输入层与输出层进行必要调整。本文复杂场景主要是不同用户在手部尺寸、运动速度、手势幅度和执行习惯等方面带来的雷达特征分布差异。

图5展示了不同方法在毫米波雷达跨人手势识别任务中的混淆矩阵结果。从图中可以看出，不同模型在各手势类别上的识别准确率存在明显差异。整体而言，在图5所示测试折次中，本文提出的MambaFuse方法取得了97.50%的识别准确率，高于 DSTFF, PLCN和DCS-CTN等对比方法，表现出更高的稳定性与泛化能力。本文方法在食指点击、掌心左转、握拳和右挥等类别具有较高的识别率，说明MambaFuse能够很好的提取在距离、多普勒和角度维度上的判别性特征。相较之下，拇指左滑、掌心右转等类别的识别率相对较低，主要原因在于这些手势与相邻方向类别之间存在较强的运动模式相似性，从而会导致出现判别错误。

具体而言，DSTFF在部分动作上面能够实现较高识别准确率，但由于缺乏对全局特征捕捉的能力，导致动作5和动作7上的识别准确率较低。动作5为掌心右转，该动作的判别信息不仅来自局部手势特征分布，还来自掌心旋转过程中方位角、径向速度随时间的全局变化。若仅关注局部瞬时特征，掌心右转容易与其他掌心旋转类动作产生混淆。动作7为左挥，其局部手势特征可能与右挥或其他侧向移动手势相似，而其关键判别线索主要体现在整个挥动过程中的起止位置、运动方向、方位角变化趋势和时间顺序。本文方法通过多尺度RD和RDA分支提取短时局部运动特征，并利用Mamba模块建模完整手势序列中的长程时空依赖，最后通过跨域注意力融合模块实现局部细粒度特征与全局运动轨迹信息的互补融合。因此，MambaFuse能够更有效地区分局部形态相似但整体运动方向或时序结构不同的手势，从而使动作5和动作7的识别准确率均高于90%，并显著优于DSTFF方法。这是由于本文的方法能同时提取手势的局部与全局特征。此外，MambaFuse相较于其他采用时序信息和多维融合特征的分类算法也展现出竞争优势，表明其通过多维信息融合策略的有效性。具体来说MambaFuse在平均准确率比DSTFF提升了5.00%，比基于PLCN的方法的平均准确率高出27.50%，比基于DCS-CTN方法的平均准确率高出5.83%，这一优势主要得益于MambaFuse在模型结构中引入的多尺度特征融合与动态通道交互机制，能够同时捕获

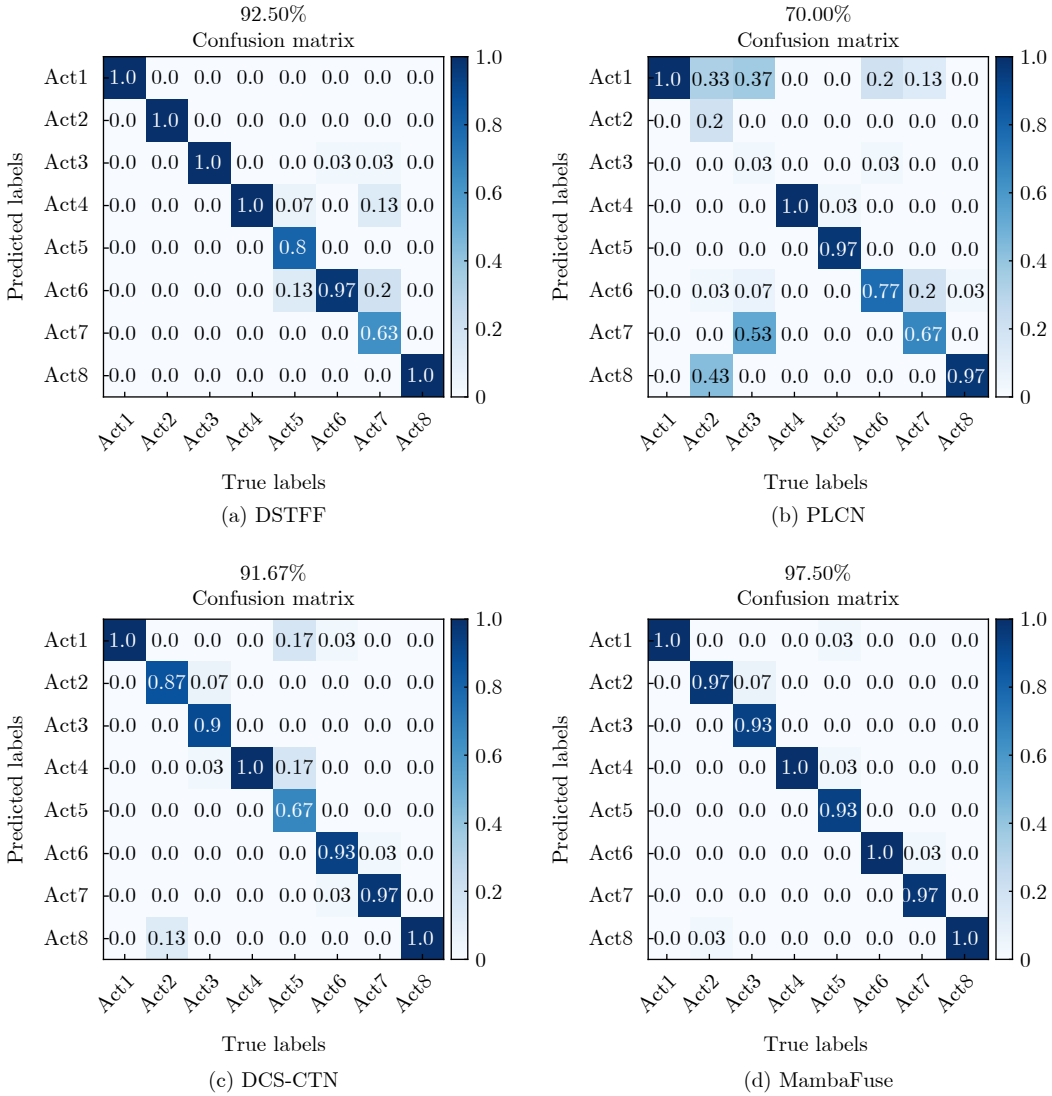


图 5 与其他方法对比的混淆矩阵结果

Fig. 5 Confusion matrices comparing the proposed method with other methods

局部时序特征与全局语义依赖，从而有效缓解个体差异带来的分布偏移问题。这些结果充分表明，本文提出的MambaFuse模型在手势识别准确率方面具有显著优势，并在跨人场景中表现出更好的鲁棒性与泛化能力。

4.4 消融实验

为验证MambaFuse中各组件的有效性，本研究从网络输入与网络架构两个关键维度进行消融实验：

(1) 不同数据立方体输入的消融分析：由于MambaFuse是端到端多域融合网络，其输入特征对最终手势识别结果具有决定性影响。为此，本文设计了RDA数据与RD数据两种输入形式，使网络能够充分捕获互补的手势特征。为验证两种输入的有效性，本文开展两组实验：第1分支仅使用RDA数据立方体，第2分支仅使用RD数据立方体。表2结果显示，缺失RDA或RD输入会导致平均识别准确率下降，表明单一数据输入无法取得理想识别效果。而融合网络通过更高的平均识别准确率在所有验证样本上取得最优结果，这证明本文的融合模型能充分挖掘不同域数据中有益的互补信息，对提升识别精度具有关键作用。另外，我们还进行了RDA和RA作为网络输入的消融实验，来进一步验证RD分支中多普勒运动信息对动态手势识别的必要性，结果表明其识别准确率比RDA和RD作为输入降低了3.92%，这主要是由于RA分支缺少Doppler维度，它更容易保留静态背景杂波、多径反射以及非手部散射体带来的干扰，对微小的手势特征造成了干扰，从而降低了模型的识别性能。

(2) 不同网络组件的消融分析：为深入探究MambaFuse各结构对识别结果的影响，本文在手

表 2 不同组件的消融实验结果

Tab. 2 Ablation results of different components

模型变体	网络结构						评价指标
	RDA	RD	RA	多尺度模块	Mamba注意力模块	融合模块	准确率(%)
01	×	√	×	√	√	√	95.50
02	√	×	×	√	√	√	94.33
03	√	√	×	×	√	√	96.31
04	√	√	×	√	self-attention	√	95.17
05	√	√	×	√	√	×	96.76
06	√	×	√	√	√	√	93.58
07	√	√	×	√	√	√	97.50

势数据集上评估3个核心组件：多尺度特征提取模块、Mamba注意力模块和跨域注意力融合模块。如表2所示，本文设计了4种模型变体，实验表明，多尺度模块、Mamba注意力模块、融合模块分别使识别准确率提升1.19%, 2.33%和0.74%，该结果进一步验证了所有主要结构的有效性和必要性。其中模型变体04表示用self-attention模块来替换 Mamba注意力模块，与我们所提模型相比其准确率降低了2.33%，这进一步表明我们所提Mamba注意力模块的具有更好的识别性能和更好的跨人泛化能力。

(3) 留一法(Leave-One-Subject-Out, LOSO)跨人交叉验证：为进一步验证模型的跨人泛化能力，本文采用留一被试交叉验证LOSO策略。由于数据集共包含11名志愿者，因此LOSO实验共进行11轮。在每一轮实验中，将其中1名志愿者的全部样本作为测试集，其余10名志愿者的数据用于训练与验证。最终，对11轮测试得到的准确率进行平均。本文对比方法DSTFF, PLCN和DCS-CTN在11折上的平均识别准确率分别为85.20%±6.38%, 61.71%±7.59%和86.88%±4.74%。本文方法实验结果如图6所示，在所有被试轮换为测试用户的情况下，各被试的识别准确率介于88.37%(P11)至97.50%(P9)之间，其余9名志愿者的准确率均处于94%～96%之间。在11折上的平均准确率为94.28%±2.55%。进一步分析LOSO结果可以发现，所提出方法在不同测试用户之间整体保持了较稳定的识别性能。除P10和P11外，其余被试的准确率均集中在94%～96%区间，说明模型的跨人识别能力并非依赖于某一特定测试用户，而是在多数未见用户上均能保持较好的泛化性能。P9取得最高准确率97.50%，表明当测试用户的手势运动模式与训练用户在距离、多普勒和角度动态特征方面具有较高一致性时，模型能够有效提取用户无关的判别性表示。相比之下，P11的准确率相对较低，为 88.37%。

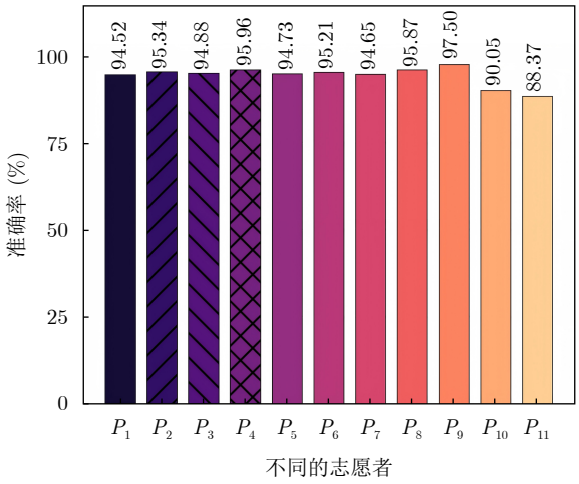


图 6 LOSO 实验中各志愿者的识别准确率

Fig. 6 Recognition accuracy of each participant in the LOSO experiment

这可能与该用户在手势执行速度、运动幅度、手掌朝向、动作持续时间及手部几何特征等方面与其他用户存在较大差异有关，从而导致部分类似手势之间的特征边界更加模糊。图7展示测试人员中手势动作差异性最大人员的原始雷达回波对比。从该图可以观察到，P9与P11在ADC采样点维度上的平均幅值分布存在明显差异。P9的回波平均幅值整体较高，且在多个采样点区间内呈现较连续的高能量分布；相比之下，P11的整体回波幅值较低，能量分布模式也与P9不同。这表明，即使在动作前手掌距雷达25 cm且多数手势掌心朝向雷达的约束条件下，不同测试人员由于手部几何特征、反射面积、动作幅度和速度习惯不同，仍会产生明显的原始回波差异。尽管如此，MambaFuse仍在所有LOSO折次中保持了较高的整体准确率，说明端到端可学习预处理、多尺度RDA/RD特征提取、Mamba长程时序建模以及跨域注意力融合能够有效增强模型对不同用户手势的鲁棒性。

(4) 模型复杂度和效率分析：表3将所提出方法的模型复杂度和推理效率与现有方法进行了比较。推理时间计算为所有测试样本中单个手势样本分类所需的平均时间。如表3所示，所提出方法的推理时间为30.12 ms，表明其能够实时识别手势样本。虽然其推理时间略高于DSTFF和PLCN，但远低于DCS-CTN的54.67 ms。此外，所提出方法的推理延迟远低于实时应用通常可接受的100 ms阈值。在模型复杂度方面，所提出方法的模型大小为65.47 MB，仅略大于DSTFF和PLCN，但远小于DCS-CTN的183.32 MB。同时，所提出方法的参数量为737 M，低于PLCN和DCS-CTN，但略高于DSTFF。这些结果表明，所提出方法在识别性能、模型复杂度和推理效率之间取得了良好的平衡。与DCS-CTN相比，所提出方法在保持较强识别能力的同时，显著降低了模型规模和推理时间。因此，所提出方法在实时手势识别场景中展现出良好的实际应用潜力。

(5) 可学习预处理方法与传统预处理方法对比：为了验证所提出的预处理方法的有效性，我们在本节中进行了消融实验。首先，所提出的预处理方法将雷达回波信号转换为RD数据立方体序列。基于图像的预处理方法用于将原始雷达回波信号转换为RD图。之后，分别使用RD数据立方体序列和RD图训练CNN-LSTM网络。图8展示了采用不同预处理方法的 CNN-LSTM 网络在训练和验证过程中的性能变化。结果表明，虽然两种方法最终均能收敛，但基于图像的预处理方法在训练和验证过程中不够稳定。相比之下，采用所提出预处理方法的模型收敛速度更快，准确率提升也更明显。此外，如图8(c)和图8(d)所示，当两个模型收敛到相近的训练损失和训练准确率，并充分拟合训练集时，其验证损失和验证准确率仍存在显著差异。这表明所提出的可学习预处理方法能够更好地学习和保留原始雷达数据特征，从而获得更高的识别准确率。为了更直接地验证本文可学习预处理模块的优势，我们进行了传统信号处理方法生成 RDA和RD表征数据作为网络输入的对比实验，结果如表4所示。从结果可以看出，在RD序列作为输入时，可学习权重预处理方法的识别准确率由90.24%提升至92.33%，提高了2.09%；在RDA序列作为输入时，识别准确率由92.61%提升至 93.78%，提高了1.17%。该结果表明，传统信号处理方法能够有效构建具有明确物理意义的RD和RDA表征，但其处理核和参数通常是固定的，难以根据下游跨人手势识别任务进行自适应调整。相比之下，本文采用的可学习预处理模块以传统DFT形式进行初始化，同时可在网络训练过程中根据分类损失进行端到端更新，因此能够更充分地保留和增强与手势类别判别相关的距离、多普勒和角度特征。

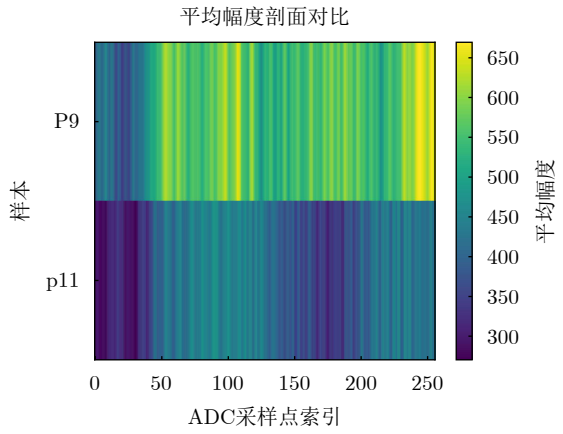


图 7 测试人员中手势动作差异性最大人员的原始雷达回波对比

Fig. 7 Comparison of raw radar echoes from the two participants exhibiting the largest gesture differences in the test set

5　结语

本文提出了一种融合端到端学习与状态空间模型的毫米波雷达跨人手势识别系统。本文设计了一个可学习的预处理模块，用于从雷达回波信号中提取判别性多维手势特征，并基于此构建了Mamba-Fuse进行手势识别。实验结果表明：相较于3种先进对比方法，本文方法在11折LOSO跨人交叉验证中取得94.28%的平均识别准确率和2.55%的标准差，最佳单折准确率为97.50%，通过跨人手势识别与人体活动识别任务的验证，本模型所展现的优异性能证明了其强大的泛化能力。总体而言，Mamba-Fuse能有效捕获更具判别力的手势特征，整合多域互补信息，实现基于RDA与RD的融合跨人手势分类。因此，本文方案有望推动了端到端和多域特征融合在非接触式手势识别技术的应用。在未来工作中，我们将进一步扩展数据集，引入更大年龄跨度的志愿者，并开展年龄分组和跨年龄泛化实验，以更全面地评估所提方法在真实应用场景中的鲁棒性。

表 3　推理时间和模型复杂度的定量比较结果

Tab. 3　Quantitative comparison of inference time and model complexity

方法	模型大小(MB)	参数量(M)	推理时间(ms)
DSTFF	58.77	14.69	22.33
PLCN	62.24	15.56	21.48
DCS-CTN	183.32	45.83	54.67
本文方法	65.47	16.37	30.12

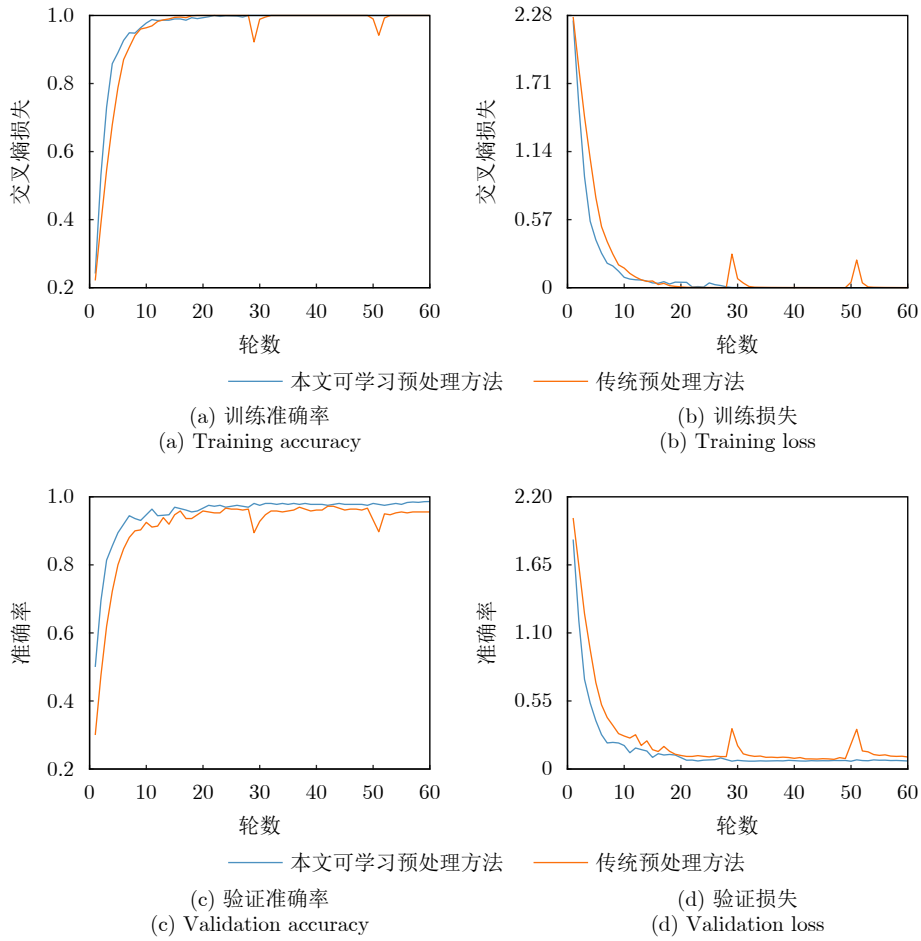


图 8 训练和验证过程中不同预处理方法的比较

Fig. 8 Comparison of different preprocessing methods during training and validation

表 4 不同预处理方法的网络输入对比结果(%)

Tab. 4 Comparison results of network inputs generated by different preprocessing methods(%)

预处理方法	RD序列作为输入	RDA序列作为输入
传统信号预处理方法	90.24	92.61
可学习权重预处理方法	92.33	93.78

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] 靳标, 孙康圣, 吴昊, 等. 基于毫米波雷达三维点云的人体动作识别数据集与方法[J]. 雷达学报(中英文), 2025, 14(1): 73–90. doi: 10.12000/JR24195.
JIN Biao, SUN Kangsheng, WU Hao, *et al*. 3D point cloud from millimeter-wave radar for human action recognition: Dataset and method[J]. *Journal of Radars*, 2025, 14(1): 73–90. doi: 10.12000/JR24195.

[2] WANG Yong, SHU Yuhong, JIA Xiuqian, *et al*. Multifeature fusion-based hand gesture sensing and recognition system[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 3507005. doi: 10.1109/LGRS.2021.3086136.

[3] LIU Zhaoyu, XIONG Yuyong, WU Gaoyang, *et al*. Super-resolution and accurate full-field displacement measurement with millimeter-wave radars[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 8507011. doi: 10.1109/TIM.2023.3327467.

[4] 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.
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.

[5] 赵雅琴, 宋雨晴, 吴晗, 等. 基于DenseNet和卷积注意力模块的

高精度手势识别[J]. 电子与信息学报, 2024, 46(3): 967–976. doi: 10.11999/JEIT230165.

ZHAO Yaqin, SONG Yuqing, WU Han, *et al*. High-precision gesture recognition based on DenseNet and convolutional block attention module[J]. *Journal of Electronics & Information Technology*, 2024, 46(3): 967–976. doi: 10.11999/JEIT230165.

[6] ZHANG Lin, YUAN Kang, CHU Hongqing, *et al*. Pedestrian collision risk assessment based on state estimation and motion prediction[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(1): 98–111. doi: 10.1109/TVT.2021.3127008.

[7] LU Jianchao, ZHENG Xi, SHENG M, *et al*. Efficient human activity recognition using a single wearable sensor[J]. *IEEE Internet of Things Journal*, 2020, 7(11): 11137–11146. doi: 10.1109/JIOT.2020.2995940.

[8] QIN Zhen, ZHANG Yibo, MENG Shuyu, *et al*. Imaging and fusing time series for wearable sensor-based human activity recognition[J]. *Information Fusion*, 2020, 53: 80–87. doi: 10.1016/j.inffus.2019.06.014.

[9] DING Chuanwei, ZHANG Li, CHEN Haoyu, *et al*. Human motion recognition with spatial-temporal-ConvLSTM network using dynamic range-Doppler frames based on portable FMCW radar[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2022, 70(11): 5029–5038. doi: 10.1109/TMTT.2022.3200097.

[10] MLIKI H, BOUHLEL F, and HAMMAMI M. Human activity recognition from UAV-captured video sequences[J]. *Pattern Recognition*, 2020, 100: 107140. doi: 10.1016/j.patcog.2019.107140.

[11] DING Chuanwei, ZHANG Li, CHEN Haoyu, *et al*. Sparsity-based human activity recognition with PointNet using a portable FMCW radar[J]. *IEEE Internet of Things Journal*, 2023, 10(11): 10024–10037. doi: 10.1109/JIOT.2023.3235808.

[12] LI Xinyu, HE Yuan, FIORANELLI F, *et al*. Semisupervised human activity recognition with radar micro-Doppler signatures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5103112. doi: 10.1109/TGRS.2021.3090106.

[13] ZHU Simin, GUENDEL R G, YAROVOY A, *et al*. Continuous human activity recognition with distributed radar sensor networks and CNN–RNN architectures[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5115215. doi: 10.1109/TGRS.2022.3189746.

[14] DING Wen, GUO Xuemei, and WANG Guoli. Radar-based human activity recognition using hybrid neural network model with multidomain fusion[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(5): 2889–2898. doi: 10.1109/TAES.2021.3068436.

[15] WANG Xiang, GUO Shisheng, CHEN Jiahui, *et al*. GCN-enhanced multidomain fusion network for through-wall human activity recognition[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 4024005. doi: 10.1109/LGRS.2022.3176117.

[16] STADELMAYER T, SANTRA A, WEIGEL R, *et al*. Data-driven radar processing using a parametric convolutional neural network for human activity classification[J]. *IEEE Sensors Journal*, 2021, 21(17): 19529–19540. doi: 10.1109/JSEN.2021.3092002.

[17] ZHAO Running, MA Xiaolin, LIU Xinhua, *et al*. An end-to-end network for continuous human motion recognition via radar radios[J]. *IEEE Sensors Journal*, 2021, 21(5): 6487–6496. doi: 10.1109/JSEN.2020.3040865.

[18] WANG Shuai, MEI Luoyu, LIU Ruofeng, *et al*. Multi-modal fusion sensing: A comprehensive review of millimeter-wave radar and its integration with other modalities[J]. *IEEE Communications Surveys & Tutorials*, 2025, 27(1): 322–352. doi: 10.1109/COMST.2024.3398004.

[19] ZHAO Peijun, LU C X, WANG Bing, *et al*. CubeLearn: End-to-end learning for human motion recognition from raw mmWave radar signals[J]. *IEEE Internet of Things Journal*, 2023, 10(12): 10236–10249. doi: 10.1109/JIOT.2023.3237494.

[20] EROL B and AMIN M G. Radar data cube processing for human activity recognition using multisubspace learning[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2019, 55(6): 3617–3628. doi: 10.1109/TAES.2019.2910980.

[21] HE Yan, TU Bing, LIU Bo, *et al*. 3DSS-Mamba: 3D-spectral-spatial Mamba for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5534216. doi: 10.1109/TGRS.2024.3472091.

[22] GU A and DAO T. Mamba: Linear-time sequence modeling with selective state spaces[C]. The First Conference on Language Modeling, Philadelphia, USA, 2024.

[23] WOO S, PARK J, LEE J Y, *et al*. CBAM: Convolutional block attention module[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: 10.1007/978-3-030-01234-2_1.

[24] LI Jianjun, XU Hongji, ZENG Jiaqi, *et al*. Radar-based human activity recognition using dual-stream spatial and temporal feature fusion network[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2024, 60(2): 1835–1847. doi: 10.1109/TAES.2023.3344685.

[25] QIAN Yujia, CHEN Chuan, TANG Longzhen, *et al*. Parallel LSTM-CNN network with radar multispectrogram for human activity recognition[J]. *IEEE Sensors Journal*, 2023, 23(2): 1308–1317. doi: 10.1109/JSEN.2022.3224083.

[26] WANG Congming, ZHAO Xiaohui, and LI Zan. DCS-CTN:

Subtle gesture recognition based on TD-CNN-Transformer via millimeter-wave radar[J]. *IEEE Internet of Things Journal*, 2023, 10(20): 17680–17693. doi: 10.1109/JIOT.2023.3280227.

作 者 简 介

方　超，博士生，主要研究方向为雷达信号处理、深度学习、人体手势识别。

王　勇，副教授，主要研究方向为新体制雷达系统、智能感知与处理。

周　牧，教授，主要研究方向为量子人工智能，量子雷达。

杨小龙，副教授，主要研究方向为无线感知与定位技术。

庞　宇，教授，主要研究方向为深度学习，目标识别。

(责任编辑：于青)