

基于 CSI 深度融合的人类活动识别

李思思, 22321023 吴桐, 22321030 高植林, 22321039

1 引言

人类活动识别 (human activity recognition, HAR) 技术目前在多个领域内展现出广泛的应用潜力, 包括但不限于医疗健康监测、智能家居自动化以及紧急响应系统^[1]。传统 HAR 技术主要依赖于可穿戴传感器、视频摄像头以及雷达等设备来采集数据。这些方法普遍存在一定的局限性: 基于可穿戴传感器的系统要求用户佩戴额外的设备, 基于摄像头的系统则受到视线限制并可能引发隐私问题, 而基于雷达的技术因信号波长较短, 其探测范围受限。

针对这些问题, 研究者们提出了基于 Wi-Fi 信号的 HAR 系统^[2]。这类系统基于的核心原理是: 人体的运动能够引起 Wi-Fi 信号的变化, 通过分析环境中 Wi-Fi 信号的变化, 可以推断出人体的活动模式。与传统的 HAR 技术相比, 基于 Wi-Fi 的 HAR 系统具有显著的优势。首先, 它无需用户携带任何传感器或设备, 从而适用于入侵检测等特殊场景。其次, 由于 Wi-Fi 设备的广泛普及, 基于 Wi-Fi 的 HAR 系统具有低成本的优势。此外, Wi-Fi 信号的覆盖范围广泛, 不受光线条件限制, 且不涉及对用户隐私的侵犯。

近年来, 基于信道状态信息 (channel state information, CSI) 的 HAR 技术受到了越来越多的关注。CSI 记录了正交频分复用 (orthogonal frequency-division multiplexing, OFDM) 子载波的振幅和相位数据。然而, 目前的方案通常只关注 CSI 的振幅或相位信息, 或者简单地将振幅和相位信息串联起来进行活动识别, 而忽略了振幅和相位之间的相互作用^[2]。实际上, 测量的相位和振幅信息可以相互补偿。具体来说, 振幅信息在障碍物遮挡 (即非视线环境 (non line of sight, NLOS)) 和远距离传播时会大幅衰减。相比之下, 相位测量在远距离传播上显示出周期性变化, 在 NLOS 条件和远距离传播下信息相对丰富。因此, 如何整合 CSI 振幅和相位的重要共同信息, 同时保留每种信息的特定模式, 仍是一个挑战。

此外, 在无线感知研究领域, 跨域感知一直是研究热点^[3]。无线感知技术的核心在于从信号中提取具有区分性的特征, 然而接收到的无线信号并非独立于背景环境, 系统部署环境的变化会导致提取的无线信号特征出现较大的类内变化, 从而严重影响系统的感知性能。不同的使用环境、不同的用户, 甚至同一用户的不同位置和朝向等因素, 都会导致数据样本分布的差异, 从而降低人体活动识别的准确率。在这种情况下, 若重新采集数据进行训练, 可能导致模型泛化能力下降和学习成本上升。

为了应对上述挑战, 本文提出了一种基于 CSI 的特征深度融合解决方案, 能够动态交换振幅和相位的子网络通道, 构建深度融合特征图, 从而提高跨域人类活动识别的性能。具体来说, 本文采用批量归一化 (batch normalization, BN) 的缩放因子作为对应特征通道的评价指标, 当某个特征通道的缩放因子接近零时, 表明该通道的特征信息可能不够显著, 则使用相同位置的另一个特征信息来替换, 以此实现振幅与相位特征的深度融合。这种方法能够更加强调整类内的统一特征, 从而提升模型的泛化性和鲁棒性, 使模型能够更好地适应未知场景和未知用户。

本文的主要贡献可以总结如下:

(1) 为了提高识别准确性, 本研究采用预处理的 CSI 振幅信息和校准的相位信息作为人类活动识别“指纹”信息。

(2) 本文设计了一个深度融合框架, 能够动态地在不同特征的子网络之间交换通道。通过 BN 缩放因子来衡量不同特征通道的重要性, 实现了 CSI 振幅和相位信息的融合。

(3) 本研究通过实验来评估所提出方法的有效性和鲁棒性。实验结果表明, 本研究提出的系统具有良好的泛化性和稳定性, 并且在性能上实现了显著的提升。

2 系统架构

本文提出的 HAR 系统架构如图 1 所示。该系统主要分为三个模块: 预处理、子网络训练、网络的深度融合。

其中预处理模块负责收集 CSI 信号, 并提取出振幅和相位数据, 但此时的数据受噪声等影响, 存在较大偏差, 因此对振幅数据使用 Hampel 滤波器进行预处理, 消除测量噪声和异常值; 对相位数据进行线性转换, 降低随机相位偏移的影响。子网络训练模块使用 Resnet50 网络分别构建振幅子网络和相位子网络, 由于振幅和相位在活动识别任务中表现出不同的稳定性和敏感性, 根据其识别效果, 深度融合模块将两个子网络进行融合, 得到充分利用二者优势的融合网络, 利用该融合网络可以实现对人类活动的精准识别。

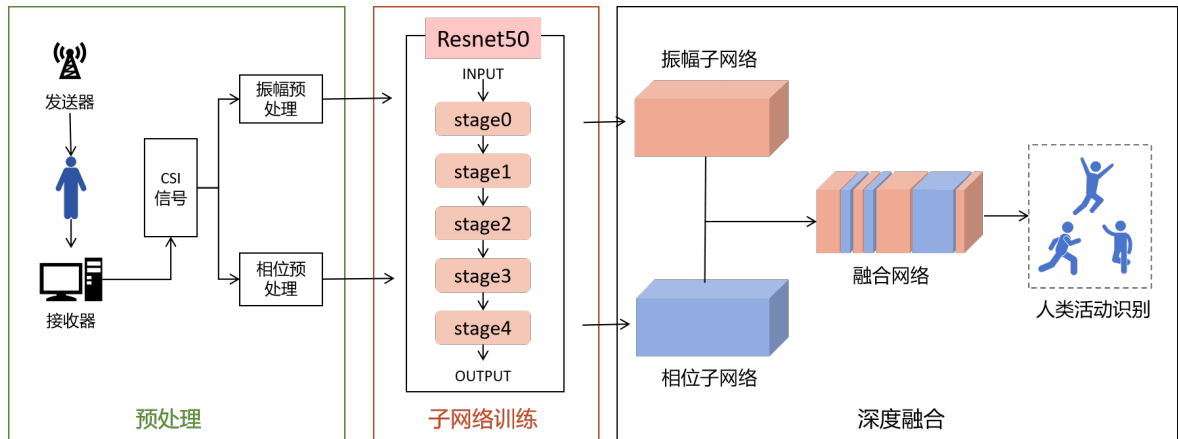


图 1 HAR 系统架构图

3 系统实现

3.1 信号预处理

CSI 是一种用于表征无线通信链路特征的信息, 其主要用于估计链路中的信道参数, 如信道增益、衰落等。CSI 的数据源自 OFDM 系统中各个子载波的振幅和相位信息, 如式 1 所示:

$$H_t = [H_t^1, \dots, H_t^i, \dots, H_t^N] \quad (1)$$

其中, H_t 为时刻 t 测得的 CSI 向量, N 为子载波总数。 H_t^i 表示子载波 i 在 t 时刻的 CSI 测量值, 可以表示为:

$$H_t^i = A_t^i e^{j\angle\theta_t^i} \quad (2)$$

其中 A_t^i 和 θ_t^i 分别表示其振幅和相位。

既往研究表明,在人类活动识别领域,CSI 的振幅表现出相对稳定性,而相位对于辨别某些细微运动具有更高的敏感性^[1-2]。因此,相较于单独利用振幅或相位信息,将两者融合在一起能更加有效地识别人类活动。基于此,本文选择采用振幅和相位相结合的方法进行人类活动识别。

在振幅的预处理方法上,本文使用 Hampel 滤波器来消除数据的测量噪声和异常值,进而提高活动识别的准确性和可靠性。预处理后,得到振幅测量矩阵 $A = [A_1, \dots, A_t, \dots, A_T]$, 其中 $A_t = [A_t^1, \dots, A_t^i, \dots, A_t^N]^T$ 。

相位数据在人类活动识别领域中的应用较为有限,这主要是由于测量相位过程中硬件设备的不完善导致的误差。事实上,误差主要源于两个方面。首先是下变频器对接收机信号产生的载波频偏(carrier frequency offset, CFO), 这是因为接收机和发射机之间的中心频率难以实现完全同步。另一个原因是模数转换器(analog-to-digital converter, ADC)产生的采样频率偏移(sampling frequency offset, SFO),主要源于时钟信号的不同步。此外,在不同子载波上的测量相位误差也是不同的。针对这一问题,Wang 等人^[4]提出了一种简单而有效的方法,通过线性变换来降低随机相位偏移的影响。

Algorithm 1: Phase Calibration

```

1 Input: measured phase values  $M_P$  of 30 subcarriers;
2 Output: calibrated phase values  $C_P$  of 30 subcarriers;
3 Set  $T_P$  as a vector as the same size of  $M_P$ ;
4 Set  $m$  as a vector from -28 to 28;
5 Set  $\text{diff} = 0$ ;
6 Set  $\eta = \pi$ ;
7 Set  $T_P(1) = M_P(1)$ ;
8 for  $i = 2 : 30$  do
9   if  $M_P(i) - M_P(i-1) > \eta$  then
10    |  $\text{diff} = \text{diff} + 1$ ;
11   end
12    $T_P(i) = M_P(i) - \text{diff} * 2 * \pi$ ;
13 end
14 Compute  $k = \frac{T_P(30) - T_P(1)}{m(30) - m(1)}$ ;
15 Compute  $b = \text{sum}\{T_P\}/30$ ;
16 for  $i = 1 : 30$  do
17   |  $C_P(i) = T_P(i) - k * m(i) - b$ ;
18 end

```

图 2 相位校准算法^[4]

在第 i 个子载波上测量到的 CSI 相位信息 θ_i 可表示为

$$\theta_i = \tilde{\theta}_i + \frac{2\pi K_i \Delta t}{M} + \beta + Z_f \quad (3)$$

其中 $\tilde{\theta}_i$ 为真实相位值, Δt 为 SFO 造成的延迟, β 为器件引起的随机相位偏移, Z_f 为测量过程中的随机噪声, K_i 为子载波指数, M 为 FFT 点数。事实上,由于 Δt 和 β 不可知,实际研究中不可能获得真正的相位信息。然而,考虑到整个频带的相位,可以对原始相位进行线性变换以去除 Δt 和 β 项。设 a 和 b 分别表示相位斜率和整个频带的偏移量,可以用下面的表达式估计相位斜率和偏移量,最终得到校准相位 $\tilde{\theta}_i$ 。

$$a = \frac{\theta_N - \theta_1}{K_N - K_1} \quad (4)$$

$$b = \frac{1}{N} \sum_{i=1}^N \theta_i \quad (5)$$

$$\tilde{\theta}_i = \theta_i - aK_i - b \quad (6)$$

由于相位的递归特性,测量到的相位被折叠(范围在 $-\pi$ 到 π 之间)。因此,本文需要将测量相位转换为真实值。为了获得真实的测量相位,可以通过减去多个 2π 来恢复折叠相位。具体算法流程如图 2 所示。在算法的第 8-13 行中,通过判断相邻子载波之间的测量相位变化是否大于给定阈值,对测量相位进行多个 2π 的补偿。在第 14-18 行中,通过上述相位校准算法获得校准相位。图 3 展示了测量相位值、恢复折叠后的真实相位值和校准相位值。可以看到,随着载波索引的增加,真实相位逐渐减小,而校准相位的范围比测量相位的范围小得多。

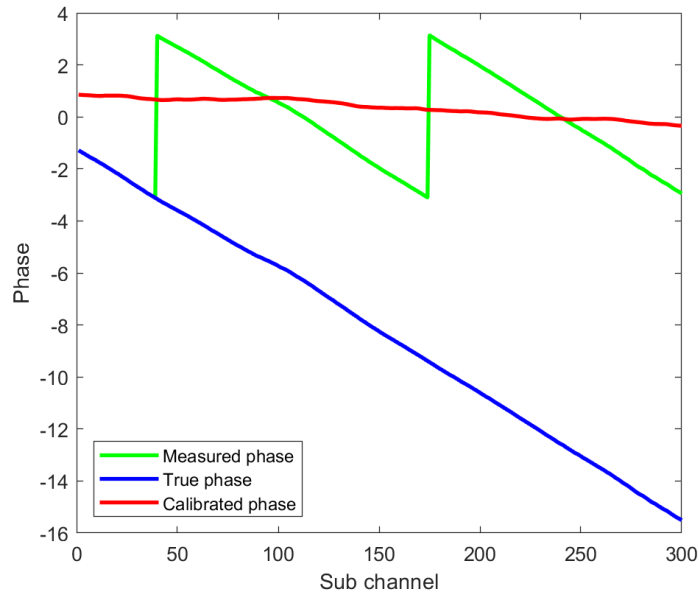


图 3 测量相位值、真实相位值和校准相位值

3.2 残差网络

本文选择残差网络 ResNet^[5]作为从振幅和相位信息提取特征的主干网络。深度卷积网络在 NLP、CV 等领域已经取得了较大成功, VGGnet 和 GoogLeNet 在增加网络层方面都实现了较好的性能。深度网络性能优于浅层网络的原因是它们能提取出差异更大的特征,然而,随着层数的增加,深度网络会出现梯度消失或爆炸问题,为了解决这些问题,He 等人提出了 ResNet 网络,通过引入残差学习的概念改变 CNN 架构的发展。ResNet 的主要思想是利用快捷连接建立的残差块,如图 4 所示。

残差建立块的数学定义可以由下面的表达式表示:

$$y = F(x, W_i) + x \quad (7)$$

其中, x 和 y 分别表示网络层的输入和输出, $F(x, W_i)$ 是残差函数,表示学习得到的残差映射, $F(x) + x$ 表示通过快捷方式的一个直接连接,添加元素对元素的标识映射不会给模型带来额外的参数和计算。本文使用 ResNet50 建立振幅子网络和相位子网络,网络细节如图 5 所示。

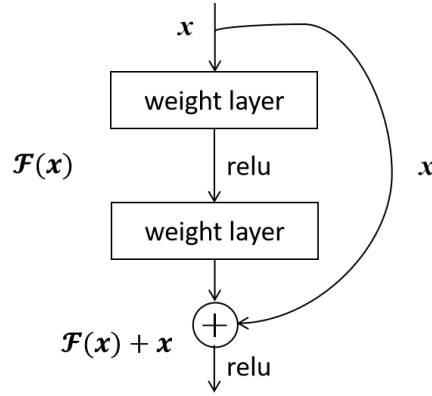


图 4 残差块

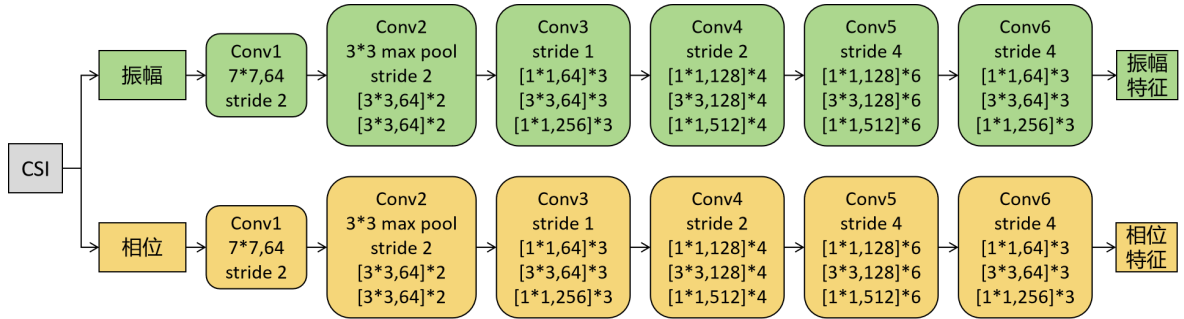


图 5 由 ResNet50 构建的子网络

3.3 融合网络

神经网络的深度融合技术是一种先进的机器学习方法，它涉及将多个神经网络层或模型结合在一起，以实现更复杂、更精细的数据处理和决策制定。这种技术通常用于处理需要高度抽象和综合能力的任务，如图像识别、自然语言处理和复杂模式识别。

在深度融合技术中，每个神经网络层负责从数据中提取不同层次的特征。较低层可能专注于基本特征，如边缘或声音的基本频率，而更高层则可以识别更复杂的模式，如面部特征或语言结构。通过这种层次结构，神经网络能够学习如何将简单的特征组合成更复杂的概念，从而实现对复杂数据的高效处理。

为了高效地融合振幅和相位信息，本文采用了一种通道交换方案来深度融合振幅和相位的子网络。在深度融合模块中，本文使用 BN 层缩放因子来确定在两个子网络之间交换和融合通道的位置。BN 层已被广泛应用于卷积神经网络中，其目的是为了消除协变量偏移，并提升模型的泛化能力。如果 BN 层中的缩放因子 γ 足够接近 0，那么在反向传播中，其对于梯度的影响也会接近 0，这表明相关通道是不重要的。基于此，本文采用 BN 层中的缩放因子 γ 作为在两个 ResNet 网络中交换通道以进行深度融合的评估指标。

针对每个子网络中与较小缩放因子相关的通道，本文认为这些通道在模型性能的贡献上相对较小，可以被视为不重要的通道。因此，本文用另一个子网络中同一位置的通道来替换那些其缩放因子小于阈值 $\epsilon = 0+$ 的通道。在实现过程中，本文将整个通道分为两个相等的子部分，并将稀疏约束应用于不相交的通道，以阻止特征信息在部分通道之间的传递。通过这种方式，可以促使模型在训练过程中学习到更加重要的通道，从而提高模型的性能。深度融合的通道交换过程如图 6 所示。

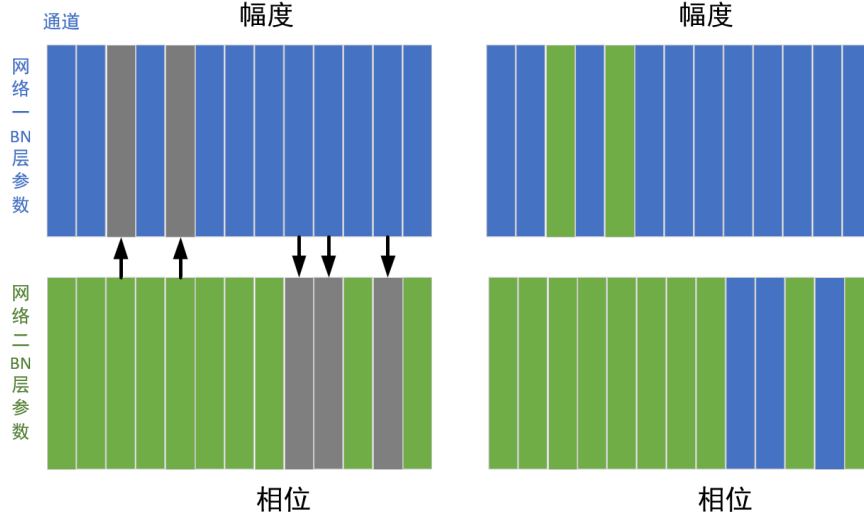
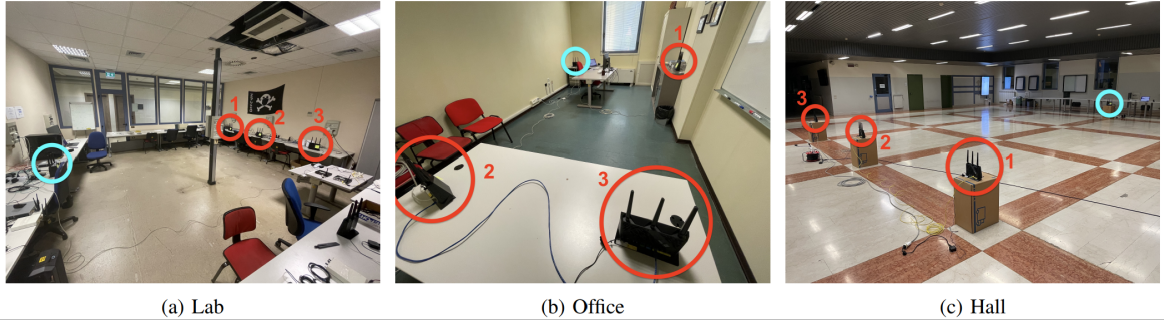


图 6 深度融合的通道交换

4 实验设置

4.1 数据集

考虑到实际采集数据的软硬件要求和复杂性，以及目前基于 CSI 的人类运动识别公开数据集较为丰富，本文最终选定了一个较新的公开数据集^[6]。该数据集对 3 名受试者进行了 12 种运动（表 1）的 CSI 数据收集，涉及 7 个不同的场景（表 2）。图 7 为实际数据采集环境，包括三个不同的场地：实验室、办公室和大厅。每个场地内部署了一个信号发射装置（标识为蓝色）和三个信号接收装置（标识为红色）。

图 7 数据采集环境^[6]

数据集集中的每个 CSI 数据点表示为一个 $W \times N$ 的复数矩阵，其中 W 代表 OFDM 子载波的数量， N 表示接收天线的数量，将 T 个连续的 CSI 数据点沿着时间维度进行聚合，构建出大小为 $T \times W \times N$ 的 3D 张量。这种数据结构便于查询特定时间点的 CSI 信息，以及研究时间维度上 CSI 的变化或分析子载波之间的数据特性。

4.2 实验设置

本文设置了三个基准实验，用于评估系统的整体性能：

- (1) 同一个人，同一环境（下简称为 PE，数据集为 S1+S4+S5）
- (2) 不同的人，同一环境（下简称为 XP，数据集为 S1+S2+S3）

表 1 数据采集环境设置

场景	参与人员	采集环境	采集时间
S1	A	Lab	1
S2	B	Lab	1
S3	C	Lab	1
S4	A	Lab	1
S5	A	Lab	2
S6	A	Office	2
S7	A	Hall	2

表 2 数据采集人类活动列表

A. Walk	D. Sitting	G. Wave hands	J. Wiping
B. Run	E. Empty room	H. Clapping	K. Squat
C. Jump	F. Standing	I. Lay down	L. Stretching

(3) 同一个人，不同环境（下简称为 XE，数据集为 S5+S6+S7）

本文选择 80% 的数据进行训练，其余 20% 的数据进行测试，模型统一训练 50 轮。

之后，本文研究了模型在跨域问题上的性能，即使用 PE 实验中训练得到的模型，测试其面对未知人员和未知环境时的识别准确率：

(1) 不同的人，同一环境（XP）：使用 S2+S3 作为测试集

(2) 同一个人，不同环境（XE）：使用 S6+S7 作为测试集

5 结果和讨论

基准实验结果如表 3 和图 8 所示，跨域实验结果如表 4 和图 9 所示。

在 PE 和 XP 场景下，使用振幅信息进行人体活动识别时，测试准确率能够达到 94% 以上；然而，在 XE 场景下，识别准确率显著下降。导致这种现象的原因可能是环境的变化对信号传输产生了较大的干扰。因此，为了提高 XE 场景下的识别准确率，可以考虑研究和采用更为先进的信号预处理技术，以消除或减轻环境噪声对信号的影响。

在 XE 场景下，使用相位信息反而实现了较好的识别效果。从理论上讲，振幅信息能够更好地表征人类行为模式，而相位信息则侧重于细微动作及姿态变化的捕捉。然而在实际应用场景中，相位信息呈现出了更好的稳定性。特别是在 XE 场景下，环境因素导致振幅信息的变化更为复杂，相比之下，相位信息对于环境变化的敏感度较低，从而在该场景中达到了更高的识别精度。这一发现提示研究人员在信号处理及模式识别的实际应用中，应根据具体场景选择合适的信号特征，以优化识别系统的性能。

深度融合 CSI 振幅和相位信息的方法在三种场景下均能提高识别准确率。具体而言，在 PE 和 XP 场景下，识别准确率可达到 98% 左右；而在 XE 场景下，识别准确率显著提升至 87.5%。在 PE 和 XP 场景下，环境噪声对信号的影响较小，振幅和相位信息均能较好地反映人体活动。在这两个场景下，深度融合方法能够充分发挥两种信息的特点，提高识别准确率。然而，在 XE 场景下，环境噪声对振幅信号的影响更

表 3 Baseline 实验测试结果

场景	数据	识别准确率 (%)
PE	Amplitude	96.3
	Phase	83.1
	Amplitude+Phase	97.8
	Deep fusion	98.5
XP	Amplitude	94.1
	Phase	67.6
	Amplitude+Phase	95.8
	Deep fusion	97.8
XE	Amplitude	43.4
	Phase	65.4
	Amplitude+Phase	86.0
	Deep fusion	87.5

为显著，而相位信息对环境变化的敏感性较低，深度融合方法能够有效结合两种信息，降低环境噪声对识别性能的影响，从而提高识别准确率。

跨域测试结果显示，仅使用振幅信息或相位信息的识别准确率在 XP 场景下均在 30% 以下，而在 XE 场景下也仅在 50% 左右。此外，基于振幅信息的识别准确率在每种情况下都要略高于基于相位信息的识别准确率。这些结果表明，在跨域测试中，未知人员对于识别准确率的影响更大。

而深度融合方法在 XP 和 XE 两种场景下的识别准确率均在 85% 以上，这说明该方法具有较好的泛化能力，能够适应未知的场景和未知的人员。深度融合方法能够有效地结合振幅信息和相位信息，从而提高识别准确率。在跨域测试中，尤其是 XP 和 XE 场景下，人员或环境的变化会导致振幅信息和相位信息的特征表现出现较大差异。然而，深度融合方法能够有效地结合这两种信息，降低这些变化对识别性能的影响。

表 4 跨域实验测试结果

场景	数据	识别准确率 (%)
XP	Amplitude	28.5
	Phase	21.7
	Amplitude+Phase	81.5
	Deep fusion	85.9
XE	Amplitude	54.6
	Phase	52.9
	Amplitude+Phase	82.6
	Deep fusion	88.0

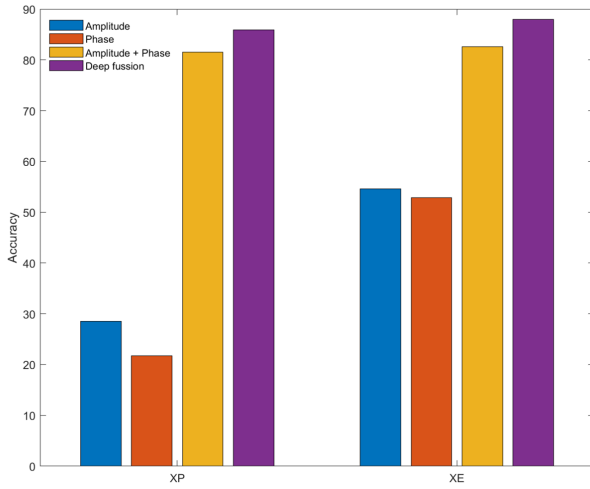


图 8 基准实验结果

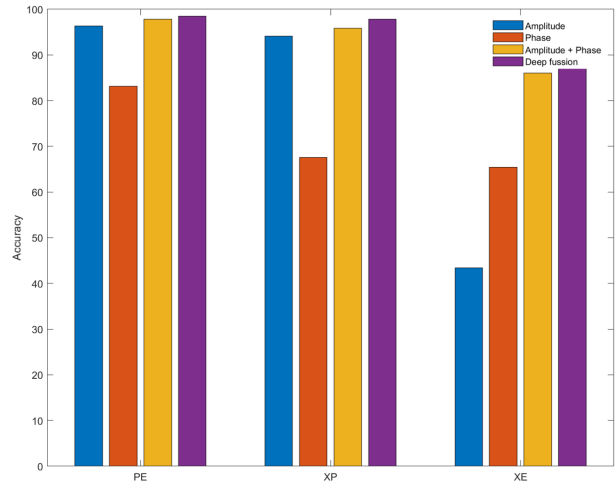


图 9 跨域实验结果

6 总结

本文设计并实现了一种基于 CSI 深度融合的人类活动识别系统。该系统利用预处理的 CSI 振幅数据和校准的 CSI 相位数据作为人类活动识别“指纹”信息，使用 ResNet 构建子网络分别提取振幅和相位特征，随后通过动态交换振幅和相位的子网络通道，构建深度融合特征图，以提高跨域人类活动识别的性能。此外，本文设置实验验证了系统的可用性和有效性。实验结果表明，本文提出的方法在不同场景下均展现了较高的识别准确率，同时模型具备良好的泛化能力和鲁棒性，能够很好地适应未知场景和未知用户。

参考文献

- [1] SHASTRI A, VALECHA N, BASHIROV E, et al. A review of millimeter wave device-based localization and device-free sensing technologies and applications[J/OL]. IEEE Communications Surveys Tutorials, 2022, 24(3): 1708-1749. DOI: 10.1109/COMST.2022.3177305.
- [2] LIU J, TENG G, HONG F. Human activity sensing with wireless signals: A survey[J/OL]. Sensors, 2020, 20(4). <https://www.mdpi.com/1424-8220/20/4/1210>. DOI: 10.3390/s20041210.
- [3] 钟怡, 毕添琪, 王菊, 丁雪, 黄延. 面向通信感知一体化的无线跨域感知研究综述[J/OL]. 信号处理, 2023, 39(1003-0530 (2023) 06-0951-12): 951. <https://www.signalpro.com.cn/article/doi/10.16798/j.issn.1003-0530.2023.06.001>.
- [4] WANG X, GAO L, MAO S. Csi phase fingerprinting for indoor localization with a deep learning approach[J/OL]. IEEE Internet of Things Journal, 2016, 3(6): 1113-1123. DOI: 10.1109/JIOT.2016.2558659.
- [5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778. DOI: 10.1109/CVPR.2016.90.
- [6] COMINELLI M, GRINGOLI F, RESTUCCIA F. Exposing the csi: A systematic investigation of csi-based wi-fi sensing capabilities and limitations[C/OL]//2023 IEEE International Conference on Pervasive Computing and Communications (PerCom). 2023: 81-90. DOI: 10.1109/PERCOM56429.2023.10099368.