Knowledge Distillation-Based Gesture Recognition Using CSI*

HUANG Zifei, ZHU Hai*, GONG Haocheng, YANG Mingze, WU Fei
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: In the field of gesture recognition, current methodologies are challenged by inefficiencies in fine-grained feature extraction, high deployment costs for models, and suboptimal robustness. To address these issues, a wireless gesture recognition method based on knowledge distillation is proposed. The approach begins with data preprocessing of channel state information (CSI), employing conjugate multiplication to eliminate phase offset noise. This is followed by the use of a Butterworth filter to eradicate static and high-frequency noise, and then principal component analysis is applied to extract the primary components of the signal. This method enables a student model to learn from the output of a teacher model through knowledge distillation, resulting in an understanding of CSI features and the teacher model's interpretation of body coordinate velocity spectrum. This technique not only maintains accuracy in recognition but also reduces the size of the model. Tests conducted on the Widar3.0 dataset demonstrate that the model's parameters are reduced by approximately 1.6 times, achieving a recognition accuracy of 94.75%. Compared to existing methods, the gesture recognition method proposed shows superior performance, offering a novel solution for wireless signal gesture recognition.

Key words: gesture recognition; channel state information; WiFi; knowledge distillation
EEACC: 7230 **doi**: 10.3969/j.issn.1004-1699.2025.11.009

基于知识蒸馏的 CSI 手势识别*

黄子非, 朱 海*, 龚浩成, 杨明泽, 吴 飞
(上海工程技术大学电子电气工程学院, 上海 201620)

摘 要: 在手势识别领域, 现有方法存在细粒度特征提取效率低、模型部署成本高、鲁棒性差等问题。为了解决这些问题, 提出一种基于知识蒸馏的无线手势识别方法。首先对信道状态信息(Channel State Information, CSI)进行数据预处理, 通过共轭乘法去除相位偏移噪声, 然后利用巴特沃斯滤波器消除静态和高频噪声, 接着使用主成分分析提取信号的主要成分。该方法通过知识蒸馏使学生模型学习教师模型的输出, 获得包含 CSI 特征和教师模型对身体坐标速度谱的理解表示。此方法在保持识别精度的同时压缩了模型规模。在 Widar3.0 数据集上的测试显示, 模型参数量减少约 1.6 倍, 识别准确率达 94.75%。与现有方法相比, 所提出的手势识别方法表现更优, 为无线信号手势识别提供新的解决方案。

关键词: 手势识别; 信道状态信息; WiFi; 知识蒸馏
中图分类号: TN929 **文献标识码**: A **文章编号**: 1004-1699(2025)11-1990-10

随着物联网和人工智能的深入发展, 手势在人机交互领域的重要性日益突出。其直观、易用和多样的特性使其成为连接人类与智能设备的纽带[1]。手势识别已广泛运用于物联网领域的各种创新应用, 如智能家居环境中的设备控制[2]、用户身份识别系统[3]和医疗健康[4]等。

目前, 手势识别技术的研究和应用主要有三种路径: 便携式传感器[[5]、视觉识别[6]和射频通信[7]。首先, 基于便携式传感器的手势识别技术主要通过捕获用户穿戴在身体某部位(如手腕或手指)上的设备所记录的手势来进行识别。设备内置加速度计、陀螺仪、磁力计等传感器, 能够准确捕捉手部的三维运动和角度变化。因其对环境光线和视野依赖较小, 能提供精确的手部动作数据, 故对精细手势控制的应用场景十分适用。但可穿戴式设备可能对用户使用造成不便, 且难以捕捉到全部复杂手势, 长时间穿戴还可能影响用户舒适度。其次, 基于视觉识别的手势识别技术通过摄像设备捕捉到图像, 再使用图像处理和深度学习算法识别和理解图像中的手势。此方法无需用户穿戴任何设备, 使用起来更为便捷。然而, 该技术对光照、背景复杂度以及视野有特定的要求, 复杂或低光照环境可能会降低识别效果, 并存在隐私问题。最后, 基于无线射频信号的手势识别通过获取设备上的手势动作信号进行识别。当手部在无线电波中移动时, 会改变周围无线电波的传播特性, 分析这些变化, 可识别特定手势。其优势在于能够在无需直接接触用户的情况下捕捉到手势, 但传统的射频设备成本高、操作复杂,

项目来源: 国家自然科学基金青年科学基金项目(61902237)
收稿日期: 2023-08-08 **修改日期**: 2024-01-05

且对环境干扰敏感,例如,其他电子设备、人体的移动等都可能影响其识别精度。

相对于上述技术,基于 WiFi 的手势识别技术因设备需求低、不受视觉限制、应用范围广、可穿透墙壁和障碍等优点,已逐渐成为研究重点。该技术通过捕获 WiFi 信号的微小变化,形成特殊的信号模式,进而识别特定手势。其中,WiFi 信号特征主要包括接收信号强度指示(Received Signal Strength Indicator,RSSI)和信道状态信息(Channel State Information,CSI)。虽然 RSSI 易于获取,但由于其不稳定性,主要适用于粗粒度的室内定位和行为感知。反之,CSI 能提供稳定的细粒度信息,因此被选为手势识别的主要信号特征。

在 IEEE 802.11n 协议下,Anmol 等[8]公开了能从商用无线网络设备中提取 WiFi 信道状态信息的 CSI Tool 工具,此后,基于 CSI 的动作识别研究逐步走向前沿。Wang 等[9]的工作独树一帜,对近年来基于 CSI 的行为识别应用进行了全方位的研究和深度的解析,精炼并归纳出三种主要的识别方法:基于模式、基于模型和基于深度学习的方法。

基于模式的方法侧重于寻找并识别人类行为的不同模式,需要手动提取输入数据特征。WiG 使用支持向量机(Support Vector Machine,SVM)对不同手势进行分类,该分类基于从异常数据序列(由不同手势引发)中提取的特征[10]。WiFinger 将 CSI 流划分为单独的手势轮廓,并从每个轮廓中提取一个特征向量,然后利用 k 最近邻(k-Nearest Neighbor,KNN)算法进行手势识别[11]。特征向量间的距离通过动态时间规整(Dynamic Time Warping,DTW)计算,有效解决了 WiFi 设备进行精细手势识别的问题。WiCatch 通过构建虚拟天线阵列并在时间域中捕捉手部的信号,重建手势运动轨迹,实现干扰消除,并使用 SVM 完成分类[12]。

基于模型的方法构建一个特定的物理模型以确定人类的活动。这种方法的优点是无需大量数据,即可使用模型以较低的计算成本实现。Ren 等[13]采用滑动窗口对 CSI 进行去噪处理,基于菲涅尔模型提取信道相位信息的特征值,来描述四种基本手势,并使用决策树分类器实现手势识别。张峻豪等[14]利用菲涅尔理论推断最佳动作捕获位置,使用改进后的定位算法对手势发生的时间段进行精准定位并标记,然后使用 DTW 进行模板匹配和手势识别。然而,实际环境中的信号变化复杂且具有不确定性,通过建模描述变化往往非常困难。即使能建立模型,其适用性也会受特定环境和数据的限制,且模型对微小变化的高度敏感性可能影响识别结果的准确性和稳定性。

随着深度学习算法的广泛应用,基于深度学习的 CSI 手势识别系统日益出现。这类系统能够通过神经网络自动学习并提取特征。例如,DeNum 使用包含一个输入层和三个隐藏层的深度学习模型进行提取振幅和相位,然后使用 SVM 进行分类,实现手势识别[15]。SignFi 则将预处理的 CSI 数据输入 9 层 CNN 进行手势分类[16]。Tang 等[17]使用并行长短期记忆-全卷积网络提取手势中各个维度上的信息,具有良好的鲁棒性。Widar3.0 通过在多普勒频移特征中加入位置和方位信息,获取领域无关的底层特征——身体坐标速度谱(Body-Coordinate Velocity Profile,BVP),然后将 BVP 作为输入,通过 CNN 和 GRU 网络提取时空特征,实现手势识别[18]。

然而,深度学习模型往往需要大量的计算资源和存储空间。为解决此问题,Guo 等[19]探讨了知识蒸馏的起源和概念,即可以将复杂模型(教师模型)的知识转移至简洁轻量的模型(学生模型),使模型在更小的计算资源下也能取得良好的性能。由 Yang 等[20]开发的 MobileDA 框架,通过统计度量减少域偏移,并通过创新的跨域蒸馏方法进一步提升深度模型性能。实验结果表明,仅使用一个简化的学生模型,就能在资源有限设备上实现高效的手势识别,并保持出色的性能。

为解决当前研究模型部署难度大、鲁棒性不足等问题,本文提出了一种基于知识蒸馏的手势识别方法。通过知识蒸馏,将教师模型中 BVP 的高级特征信息用以引导学生模型中 CSI 的低级特征信息的学习,从而提升手势识别性能。在保持优秀性能的同时,模型参数量和计算复杂性均得以降低,使得该模型更适合在资源受限的设备上运行。实验结果进一步证明了该方法的有效性。

1　基本理论概述

1.1　信道状态响应

信道状态信息是无线通信中用于描述信道特性的一种信息,其涵盖了信道在传输过程中的衰减、噪声及多径干扰等因素。CSI 数据准确描绘了无线信道在空间、时间和频率三个维度的变化特性,这在改进信号质量和提升数据传输效率等方面具有重要意义。现阶段,IEEE 802.11 系列的无线通信标准已支持使用正交频分复用-多输入多输出[21]技术实现并发数据传输。该技术在商用无线设备中得到广泛应用,以提升数据传输速度和网络性能。

在无线感知领域,信道响应分析常用方法有信道冲击响应(Channel Impulse Response,CIR)和信道频率响应(Channel Frequency Response,CFR)。CIR主要描绘了无线信道在时域中的多径效应,用以区分不同的传播路径。在线性时不变的假设下,可以使用式(1)表达CIR:

$$h(\tau)=\sum_{i=1}^{N}\alpha_i e^{-j\theta_i}\delta(\tau-\tau_i) \tag{1}$$

式中:i 表示在第 i 个信号路径,α、θ、τ 分别表示幅度衰减、相位差和传输延迟,N 表示总传播路径数量,$\delta(\tau)$表示单位脉冲函数。CIR直观地描绘了信号在各传播路径中的时延扩展情况。然而,由于CIR无法直接展示在频域中的信号幅度衰减,因此需利用CFR来提供这部分信息。CFR从频域的视角呈现了信号的多径传播现象,涵盖了信号的幅度特性和相位特性。假设系统是线性时不变的,通过对CIR执行快速傅里叶变换(Fast Fourier Transform,FFT),可以得到CFR,如式(2)所示:

$$H=\mathrm{FFT}(h(\tau)) \tag{2}$$

式中:H 表示CFR,$h(\tau)$ 表示CIR。在无线信号通过空气传播的过程中,可能会遇到各类物体,例如建筑物、地形或其他障碍物。信号在遇到这些物体时,可能会发生反射、折射或散射,从而形成新的不同的信号路径。这些路径往往比直射路径长,因此信号可能在不同的时间到达接收器,产生多径传播,如图1所示。多径效应会因多径传播的出现而产生,这是由于不同路径上的信号传播时延不同,当这些信号在接收端叠加时,可能会造成干扰和对原始信号的失真。对于室内环境中的无线信号,可以将其视作小尺度平坦衰落模型,其在频域上的表示如式(3)所示:

$$Y=\boldsymbol{H}X+\mathrm{Noise} \tag{3}$$

式中:Y 表示接受端信号频域大小;$\boldsymbol{H}$ 表示信道增益矩阵;X 表示发射端信号频域大小;Noise表示高斯白噪声。根据式(3),CSI信息可以使用矩阵表示为:

$$\boldsymbol{H}=\frac{Y}{X} \tag{4}$$

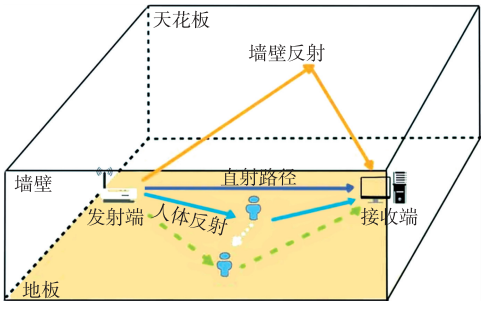


图1 传播模型

在某一子载波上,CSI可以用CFR表示为:

$$H(k)=\|H(k)\|e^{j\angle H(k)} \tag{5}$$

式中:$\|H(k)\|$ 表示第 k 条子载波的幅度,$\angle H(k)$表示第 k 条子载波的相位。在特定的采样频率和固定的时间窗口下,每条天线可以得到一个CSI序列矩阵:

$$\boldsymbol{H}=[H_1,H_2,H_3,\cdots,H_s] \tag{6}$$

式中:H_s 表示第 s 条子载波上的CSI,s 表示子载波总数量。当人员进行特定动作时,CSI所反映的振幅和相位会相应地发生变化。通过研究这些变化规律,可以实现有效的人体动作识别。

1.2 身体坐标速度谱

在无线通信领域,多普勒效应描述的是由于发送者或接收者的运动导致接收信号频率的改变,而多普勒效应会造成多普勒频移(Doppler Frequency Shift,DFS)。当物体或人体在WiFi信号的覆盖范围内移动,该移动会影响WiFi信号的传播,并因此改变接收信号的频率。通过对CFR执行短时傅里叶变换来获得DFS,DFS可由式(7)表示:

$$f_D(t)=-\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}t}d(t)=-f\frac{\mathrm{d}}{\mathrm{d}t}\tau(t) \tag{7}$$

式中:f 是载波频率,λ 表示信号的波长,$\tau(t)$是时延,$d(t)$表示反射路径的传播距离。Widar3.0从原始CSI测量的多普勒频移谱中,以较低的信号水平推导并提取出人类手势不受环境影响的特征,这被称为身体坐标速度谱(Body-Coordinate Velocity Profile,BVP)。BVP的核心思想是将人体各部分(如双手、双臂及躯干)在执行手势时的不同速度反映在一个速度分布图上,这些分布图能唯一标示出人体的活动类型。

在多路径传播环境下,可以通过将多路径信号的相位与相应的DFS相关联,将CSI转变为多普勒频移形式。当人执行特定的手势时,各部分以不同速度移动,导致这些部分反射的信号经历不同的DFS。这些DFS在接收端累积,生成了一个维度为 $F\times M$ 的DFS频谱图,其中 F 表示频率采样点的数量,M 表示收发链路的数量。虽然DFS频谱图包含了手势信息,但这些信息与人体的位置和方向等环境因素密切相关。而人体坐标系统下的物理运动速度的功率谱分布更为稳定和可靠,只会受到手势特性的影响。而BVP矩阵被量化为 $N\times N$ 的离散矩阵,其中 N 表示速度分量对应的潜在数值的数量,每个速度分量都沿着人体坐标系的各个轴进行分解。如图2所示,建立了一个局部的人体坐标系统,其原点定位在人的位置,正轴与人的方向对齐。在

此坐标系统中,手势运动的任一速度成分 v 可被映射到 DFS 频谱图的某个频率成分上。基于此,可以在 DFS 谱图与 BVP 之间建立连续映射关系。因此,为消除环境影响,从多个链路的 DFS 谱图中,推导出了与环境无关的 BVP,从而准确识别和解析手势,有效地消除了环境因素的影响。

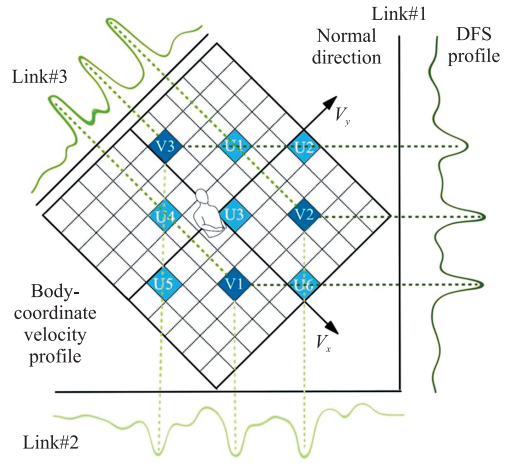


图 2　BVP 和 DFS 频谱图之间的关系[18]

1.3　知识蒸馏

知识蒸馏的目标是通过将复杂的知识转化为更易于理解的形式,以提升知识传递和理解的效率。学生模型通过优化知识蒸馏损失函数以尽可能地模拟教师模型的行为和性能。

首先,需要对教师模型执行预训练。在训练阶段中的每个样本通过教师模型生成相应的预测结果,这些结果通常以概率分布的形式存在,被称为“软标签”。软标签与原始的“硬标签”(实际的类别标签)相比,提供了更多信息,例如,模型对不同类别的不确定性。为了提高负标签对损失函数的影响,引入了一个“温度 T”参数,并将其用于蒸馏中的 Softmax 函数,这使得概率输出更为平滑。具体的数学计算过程如式(8)所示:

$$\hat{y}_i=\frac{\mathrm{e}\left(\frac{z_i}{T}\right)}{\sum_j \mathrm{e}\left(\frac{z_i}{T}\right)} \tag{8}$$

式中:z_i 表示模型的原始输出,T 表示温度,$\hat{y}_i$ 表示经过温度 T 调整后的预测。将温度 T 设定为 1 后,方程收敛为 Softmax 函数的标准形式。在训练学生模型时,采用了软标签和硬标签的联合训练策略。通过使用硬标签,可以在一定程度上减少教师网络出错的可能性。软标签的交叉熵损失的具体计算过程如式(9)所示:

$$L_{\mathrm{soft}}=-\sum_i^n y_i'\log(\hat{y}_i') \tag{9}$$

式中:y_i'表示软标签,$\hat{y}_i'$表示学生模型对软标签的预测值,n 表示类别的数量,i 表示类别索引。硬标签的交叉熵损失计算过程如下:

$$L_{\mathrm{hard}}=-\sum_i^n y_i\log(\hat{y}_i) \tag{10}$$

式中:y_i 表示硬标签,$\hat{y}_i$ 表示学生模型对硬标签的预测值,n 表示类别的数量,i 表示类别索引。学生模型的目标是尽可能地模拟教师模型的行为,即尽可能地拟合教师模型的软标签。这个过程可以通过最小化以下损失函数来实现:

$$L=\alpha L_{\mathrm{soft}}+(1-\alpha)L_{\mathrm{hard}} \tag{11}$$

式中:α 表示一个权重参数,用于控制硬标签和软标签的相对重要性。学生模型从教师模型那里获得了更加丰富的知识,从而提升了自身的分类精度。

2　本文方法

2.1　方法设计

如图 3 所示,本文提出的手势识别框架主要由数据预处理和知识蒸馏两个部分构成。首先,借助巴特沃斯滤波器对 CSI 原始数据进行去噪,并通过主成分分析(Principal Component Analysis,PCA)对子载波实现降维,从而提取出动态信号的主分量。在知识蒸馏阶段,预训练一个由 BVP 驱动的教师模型,并通过引

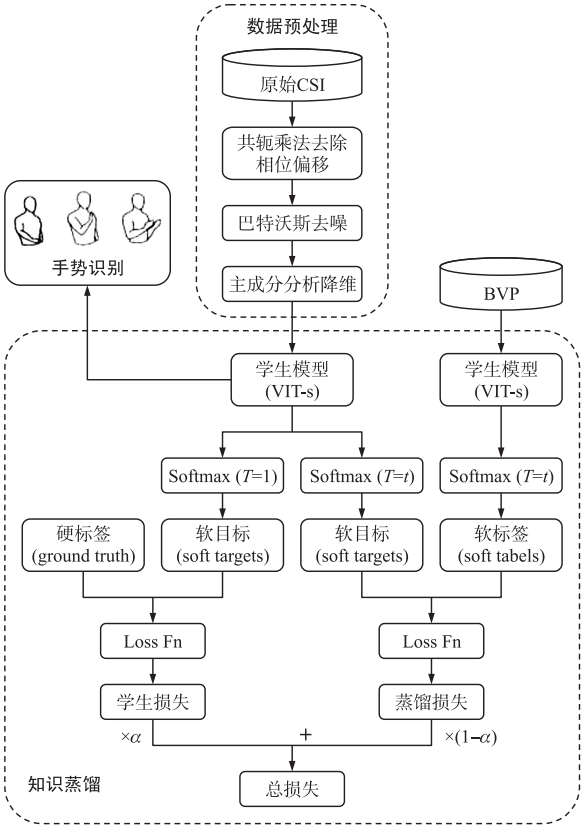


图 3　框架概要

入由温度调节的 Softmax 函数,将教师模型和学生模型的输出概率分布转化为软目标,其中温度参数 T 被用于调整各个软目标的权重。以 KL 散度作为蒸馏损失,鼓励学生模型向教师模型学习,同时尽量减小教师模型与学生模型间的回归差异。此外,学生模型不仅要模仿教师模型的行为,同时还要增强其对分类数据的学习能力,为此,为了提高学生模型对真实标签的准确性,采用了交叉熵函数。

2.2 数据预处理

由于 WiFi 信号易受环境干扰,可能导致 CSI 数据噪声过大或信号混乱,从而影响后续分类任务。因此,尽管 Widar3.0 提供原始的 CSI 数据,但仍需要进行预处理以优化信号质量、减少噪声干扰,并提取更准确、鲁棒的特征。

2.2.1 数据集

本文使用清华大学杨峥团队公开的 Widar3.0 手势数据集,该数据集是在 CSI 手势识别领域中公认的高质量数据集之一,数据集中包含丰富的 CSI、DFS 和 BVP。每个 BVP 文件都是一个 $20\times20\times T$ 的矩阵,其中,第一维代表 x 轴速度范围在[−2,+2] m/s,第二维则表示 y 轴速度范围在[−2,+2] m/s,而第三维则对应时间戳,其采样率为 10 Hz。

图 4 展示了六种手势的示意图。这些手势的数据是由 17 位用户在教室、礼堂和办公室这三种环境下生成的,期间,用户面向五个不同方向,而接收器则有八个不同位置。对于每种手势的数据集,选择其中的 80%作为训练样本,剩余的 20%则用作测试样本。

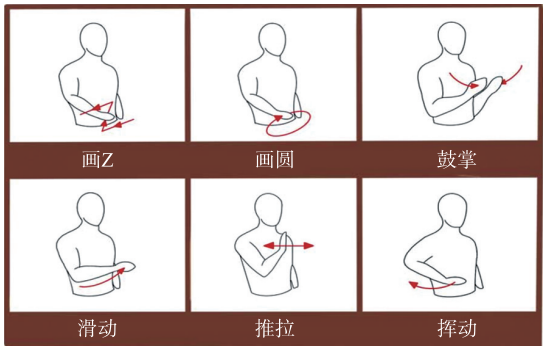


图 4 6 种手势的示意图[18]

2.2.2 共轭乘法去除相位偏移

在多输入多输出无线通信系统中,利用多个天线,每个天线接收不同的信号,在空间传播过程中经历不同程度的衰减和相位变化。由于接收器和发射器之间缺乏精确的时间对齐,与时间相关的不可预测的随机相位移动会出现,从而导致时间域的失真。因此,需要特定的信号预处理技术来抵消相位移动,目标是最大限度地保持 DFS 的完整性。共轭乘法算法的应用有助于识别在特定时刻信号质量优越的天线,并选择这些天线来提高整个系统的性能。图 5 展示了使用此技术处理的天线对的结果。

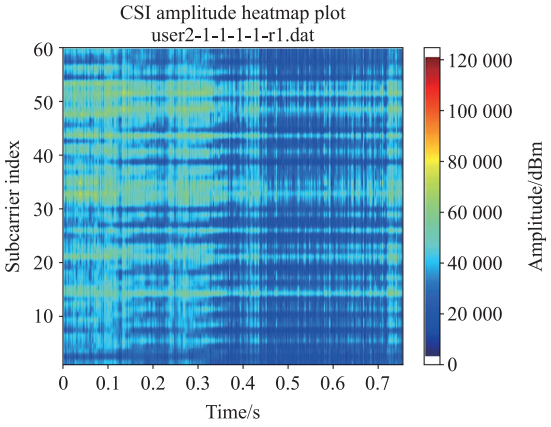


图 5 共轭乘法的结果

2.2.3 巴特沃斯去噪

在手势识别领域,主要的研究焦点集中在手指及手臂的动作。CSI 信号中,常含有大量低频干扰和突发噪声。这些干扰和噪声的振动频率多在低频区域,噪声信息则主要分布在高频段。针对这个问题,使用巴特沃斯滤波器来滤除高频部分。由于人类活动导致的 CSI 数据变化频率较低,因此低通滤波器能够有效消除由多径效应和环境变化引起的高频噪声。该滤波器的特性在于,其通带的频率响应平稳,阻带能快速削减噪声或干扰。这种滤波器适合处理 CSI 各子载波。

如图 6 所示,在实际操作中,上截止频率一般设定为 60 Hz,这是由手势动作的经验速度(2 m/s)决定的。对于工作在 5.8 GHz 的 WiFi 设备,其多普勒频率的上限约为 60 Hz。下截止频率设定为 2 Hz,这是基于人为考虑,需在完全消除高频干扰和减少多普勒效应低频分量之间找到平衡点[22]。

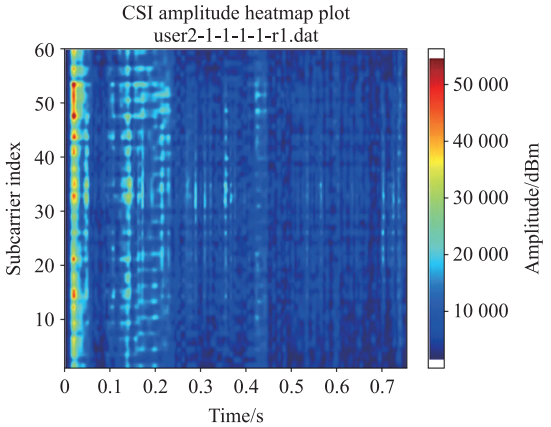


图 6 巴特沃斯滤波器去噪

2.2.4 主成分分析降维

在每个 CSI 数据流中,各子载波间的信道效应

是相互关联的。若所有 CSI 子载波被用于特征提取,可能产生冗余信息。解决此问题的办法是利用 PCA 进行数据降维。数据降维是一种在高维数据集中减少特征维度的技术。它尽可能保持数据特征的同时,减少冗余信息和噪声的影响,以提高数据处理和分析的效率。

CARM[23]方法首先采用 PCA 算法对 CSI 在同一网卡上的三根天线的所有子载波进行降维处理,进而提升频谱分析和特征工程的效率,同时减少识别框架的复杂度。不同子载波的 CSI 流是同一组时变信号的线性组合,因此它们高度相关。此外,这些子载波各自反映了不同的中心频率和与信号波长有关的 DFS。因此,本文选用 PCA 算法对子载波进行降维。PCA 的详细步骤首先包括数据中心化,即将所有数据点减去其均值,使得新数据集的均值为 0。中心化的公式如下:

$$X_{\text{centered}} = X - \mu \tag{12}$$

式中:X 表示原始数据,μ 表示数据的均值。然后计算协方差矩阵,反映出数据的各个维度之间的相关性。如式(13)所示:

$$C = \frac{1}{N} X_{\text{centered}}^{\mathrm{T}} X_{\text{centered}} \tag{13}$$

式中:N 是数据点的数量。之后计算协方差矩阵的特征值和特征向量,特征值和特征向量反映了数据的主要变化方式。计算特征方程如式(14)所示:

$$\boldsymbol{C}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{14}$$

式中:$\boldsymbol{C}$ 是协方差矩阵,$\boldsymbol{v}$ 是特征向量,λ 是特征值。构造投影矩阵 $\boldsymbol{P}$,其由前 k 个最大特征值所对应的特征向量组成。将原始数据投影到选定的主成分上,得到降维后的数据。

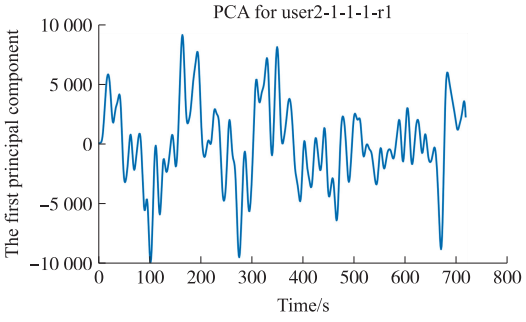


图 7　使用 PCA 获取信号主分量

$$\boldsymbol{X}_{\text{projected}} = \boldsymbol{X}_{\text{centered}} \boldsymbol{P} \tag{15}$$

式中:$\boldsymbol{X}_{\text{projected}}$ 是投影后的数据。如图 7 所示,通过 PCA 提取出一条与手势引起的信号相关的信号主分量。

2.3　教师模型和学生模型

本文中的教师模型和学生模型都以 VIT(Vision Transformer)模型为基准,如图 8 所示。

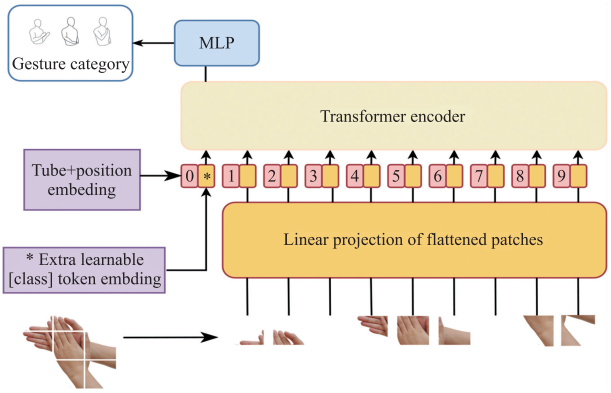


图 8　模型示意图

教师模型首先对 BVP 特征向量进行线性投射,之后,Token 序列经过位置嵌入后输入到 Transformer 编码器中。如图 9 所示,Transformer 编码器主要由两个核心组件构成,即多头自注意力(Multi-Head Self-Attention,MSA)模块和多层感知机(Multi Layer Perceptron,MLP)模块。MSA 模块的主要作用是在序列数据中捕捉长距离的依赖关系。为了避免深度网络性能的衰退,每个模块之前都部署了批归一化(Batch Normalization,BN),并在串行结构上采用了残差连接。MLP 模块与分类器中的 MLP 层类似。以下是编码器的详细运算过程:

$$y'_{\xi} = \mathrm{MSA}(\mathrm{BN}(y_{\xi-1})) + y_{\xi-1}, \xi = 1 \cdots L \tag{16}$$

$$y_{\xi} = \mathrm{MSA}(\mathrm{BN}(y'_{\xi})) + y'_{\xi}, \xi = 1 \cdots L \tag{17}$$

$$z = \mathrm{BN}(y_L^0) \tag{18}$$

在上述公式中,y_L^0 表示未经归一化的分类器输入。具体而言,编码器能够从手势运动所引起的空间内信号功率的特征变化来提取信息。MSA 能同时处理所有位置的输入,并使神经网络从多个不同的角度理解和编码深度信息。经过编码器的建模处理后,在

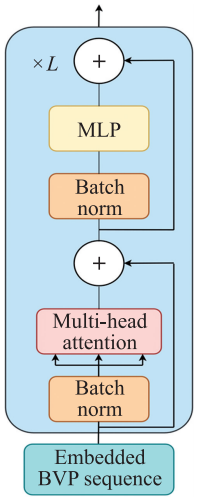


图 9　Transformer 编码器结构

预设的特征空间中,MLP 将分类器的输入映射以产生与分类标签对应的概率分布。在知识蒸馏框架下,经过 Softmax 函数处理,教师模型的输出概率分布转化为软目标,从而驱动学生模型的学习过程。

学生模型的输入数据和模型结构是其与教师模型的主要区别。学生模型输入的是经过预处理的 CSI 数据 $S_{csi} \in \boldsymbol{R}^{T\times L}$,其中 T 代表采样后的 CSI 信号样本,L 代表推导出 BVP 的链路数量。通过 6 个 WiFi 链路的 CSI 时间序列信号推导出 BVP 序列,这个序列集成了发射器位置、识别对象的位置和方向等信息。因此,学生模型将教师模型中的集成特征表示替换为包含环境信息的特征表示。这些环境信息向量在通过全连接层的线性投射后,被映射到 D 维特征向量空间。在模型架构上,学生模型的编码器仅有一层,而教师模型的编码器则有三层。这样的设计使学生模型在处理复杂的环境信息时能保持较低的计算复杂度。

教师模型与学生模型都选择了 MLP 作为分类器,MLP 可以快速且有效地处理并行生成的特征表示,并且捕获输入的重要信息,而忽略无关的细节。MLP 的网络结构由两层构成,如图 10 所示。为了提升模型对数据的拟合能力,在每一层都引入了 ReLU 非线性激活函数。训练过程中,交叉熵损失函数被选为优化目标,以最小化预测误差。为避免模型过拟合并降低分类器的计算复杂度,两层 MLP 都添加了 Dropout 层。在训练过程中,Dropout 层能够随机丢弃部分神经元,以此提升模型的泛化能力,确保了模型在维持高性能的同时,计算复杂度也保持在较低水平。

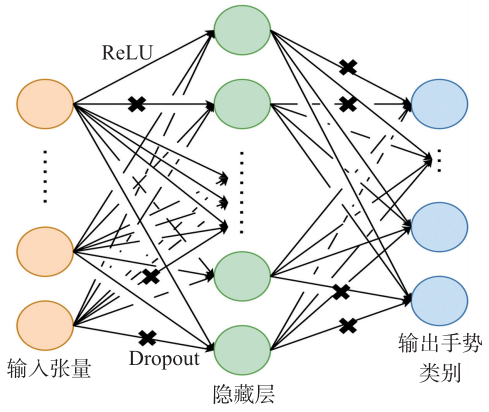


图 10 MLP 结构图

3 实验结果与分析

3.1 评估指标

本文采用三种评价指标来全面验证模型效果,包括精确率(Precision)、召回率(Recall)和 F1 分数(F-measure score)。

精确率是在预测为正例的样本中,实际为正例的比例。其计算公式如式(19):

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{19}$$

召回率是在实际为正例的样本中,被预测为正例的比例。其计算公式(20):

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{20}$$

F1 分数是精确率和召回率的调和平均数。其计算公式如式(21):

$$\text{F1}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{21}$$

式中:TP 是真正例(True Positive)的样本数量,TN 是真负例(True Negative)的样本数量,FP 是假正例(False Positive)的样本数量,FN 是假负例(False Negative)的样本数量。

3.2 实验结果与对比分析

图 11 所示为 Widar3.0 数据在教室中的六种手势动作识别结果的混淆矩阵。其中,“推拉”的准确率最小,为 88.41%,而手势较为复杂的“画 O”和“画 Z”的准确率分别为 99.16%和 99.54%,从实验结果可以看出,蒸馏后的学生模型手势识别的准确度都达到了较高水平,在 88%以上,表示模型在整体识别效果上的优异性能。

图 11 教室中 6 种手势的混淆矩阵

表 1 展示了经过知识蒸馏后的学生模型对 6 种不同手势的评估结果。所有类别的精确率均超过 90%,表明模型预测的精确性高,误报率低。所有类别的召回率都超过 88%,说明模型能够在大部分情况下正确地识别出实际的手势。所有类别的 $F1$ 分数均超过 90%,表明模型的整体性能较高。这样的结果对于手势识别应用具有重要的实际意义,能够提供可靠的用户体验和准确的交互操作。

表 1　教室中 6 种手势的 3 个评价指标　　%

类别	精确率	召回率	*F*1 分数
推拉	92.82	88.40	90.56
挥动	90.64	92.30	91.46
鼓掌	90.78	95.77	93.21
滑动	97.39	94.23	95.78
画 O	99.03	99.15	99.09
画 Z	99.00	99.54	99.27

不同模型的实验对比如图 12 所示,应用知识蒸馏技术后,学生模型的性能有了显著的提升。在没有进行蒸馏的情况下,学生模型的准确率仅为 62%。但在蒸馏过程中,学生模型不只是学习了如何分类输入数据,也吸收了教师模型的处理方式及在各种情况下的决策方法。这种额外的“软”知识显著提升了学生模型的性能,准确率大幅度提高至 94.75%,比 Widar3.0 高出 2.05%。这说明,通过知识蒸馏,学生模型可能获得了教师模型在处理 BVP 数据中学到的有价值的模式和特征,这是直接从 CSI 数据中学习难以获取的。因此,学生模型在处理 CSI 数据时,可能具备了比教师模型更丰富且有效的知识和理解,从而提升了其性能。此外,如表 2 所示,教师模型的参数量为学生模型的 2.6 倍左右,验证了学生模型在保持优异的识别性能的同时,也可以对模型进行有效的压缩。

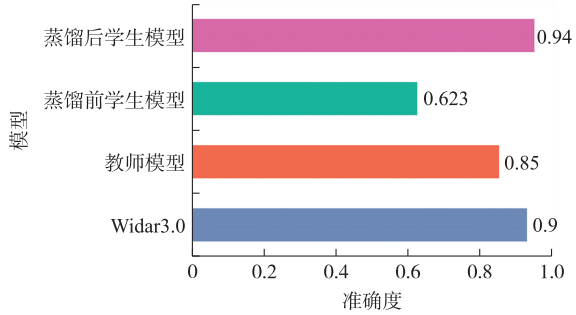


图 12　不同模型的实验对比

表 2　模型参数量

模型	参数量/万个
教师模型	260
学生模型	99.8

图 13 展示了在三种不同环境下的六种手势实验结果。教师、办公室和礼堂中所有手势的准确率分别达到了 88%、96%和 94%以上,这说明即使在不同环境中模型也能较容易地从数据中学习到这些手势特征。办公室环境下的识别准确率几乎全为最高,可能由于办公室环境更安静、稳定,噪声和干扰较少,因此模型能更精准地识别出手势动作。学生模型在这三种环境下的手势识别准确率均表现优异,进一步验证了模型的鲁棒性。

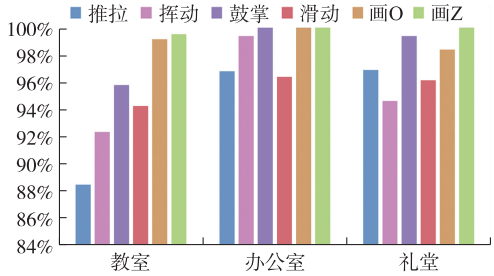


图 13　三种环境下 6 种手势的实验结果

温度主要影响了教师模型输出的“软化”程度,使学生模型不仅能学习到数据的硬标签,也能学习到教师模型对其他类别的理解。从图 14 可见,随温度上升,模型准确度先显著提升,然后达到 20 后趋于平稳。较高的温度能使软标签的概率分布更均匀,因而更多的信息能被传递给学生,从而提升性能。但当温度继续升高,教师模型的输出可能变得过于"软",即所有类别的概率都趋向等同。这可能使学生模型难以从教师模型的输出中获取足够信息,从而影响性能。这里的 T 是模型的一个数学超参数(蒸馏温度),不是物理温度,所以本身是无单位的。

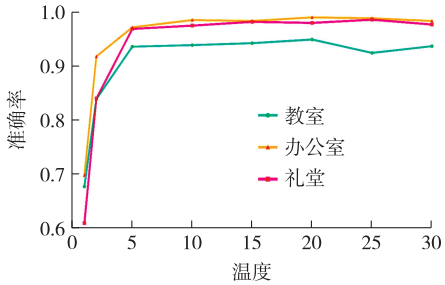


图 14　三种环境下不同蒸馏温度 *T* 的实验结果

4　结论

为了同时达成模型压缩、鲁棒性高和准确度高的需求,本文提出了一种基于知识蒸馏的手势识别方法。首先对原始的 CSI 数据进行预处理,然后通过学生模型从教师模型中获取有价值的知识,从而有效地识别出六种常用的交互手势。实验结果表明,尽管学生模型的数据与教师模型存在差异,学生模型依然能从教师模型中获取有效的特征表示。这一过程不仅提升了识别精度,而且显著减少了模型的参数量,从而提高了模型效率和可部署性。未来的研究将继续探索各类知识蒸馏策略,并寻求解决跨域能力问题。将知识蒸馏与其他先进的机器学习技术(如迁移学习)相结合,以克服数据稀疏和领域差异的问题,有望进一步提升 CSI 手势识别的性能、鲁棒性以及跨域能力。

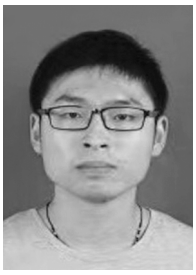
参考文献:

[1]　Guo L, Lu Z, Yao L. Human-Machine Interaction Sensing

Technology Based on Hand Gesture Recognition: A Review[J]. IEEE Transactions on Human-Machine Systems,2021,51(4):300-309.
[2] Nigam S, Shamoon M, Dhasmana S, et al. A Complete Study of Methodology of Hand Gesture Recognition System for Smart Homes[C]//2019 International Conference on Contemporary Computing and Informatics (IC3I), Mysuru, Karnataka, India, IEEE, 2019:289-294.
[3] Bianco S, Napoletano P, Raimondi A, et al. U-Wear: User Recognition on Wearable Devices Through Arm Gesture [J]. IEEE Transactions on Human-Machine Systems, 2022,52(4):713-724.
[4] Chen J, Zhao S Y, Meng H N, et al. An Interactive Game for Rehabilitation Based on Real-Time Hand Gesture Recognition[J]. Frontiers in Physiology, 2022, 13:1028907.
[5] 刘玉焘. 基于可穿戴式传感器的人体动作捕获与识别研究[D]. 哈尔滨:哈尔滨工业大学,2020.
[6] Hasan H S, Kareem S A. Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey[C]// 2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Kuala Lumpur, Malaysia, IEEE, 2012:55-60.
[7] Wang S, Zhou G. A Review on Radio Based Activity Recognition [J]. Digital Communications and Networks, 2015,1(1):20-29.
[8] Halperin D, Hu W, Sheth A, et al. Tool Release: Gathering 802.11 n Traces with Channel State Information [J]. ACM SIGCOMM Computer Communication Review, 2011,41(1):53-53.
[9] Wang Z, Jiang K, Hou Y, et al. A Survey on Human Behavior Recognition Using Channel State Information[J]. IEEE Access, 2019, 7:155986-156024.
[10] He W, Wu K, Zou Y, et al. WiG: WiFi-Based Gesture Recognition System[C]//2015 24th International Conference on Computer Communication and Networks (ICCCN), Las Vegas, NV, USA, IEEE, 2015:1-7.
[11] Li H, Yang W, Wang J, et al. WiFinger: Talk to Your Smart Devices with Finger-Grained Gesture[C]//2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, NY, USA, ACM, 2016:250-261.
[12] Tian Z, Wang J, Yang X, et al. WiCatch: A Wi-Fi Based Hand Gesture Recognition System [J]. IEEE Access, 2018,6:16911-16923.
[13] Ren S, Wang H, Gong L, et al. Intelligent Contactless Gesture Recognition Using WLAN Physical Layer Information[J]. IEEE Access, 2019, 7:92758-92767.
[14] 张峻豪, 吴飞, 朱海. 基于 GD-Kmeans 和菲涅尔理论的 WiFi 手势识别方法[J]. 计算机工程与应用, 2020, 56 (19):126-131.
[15] Zhou Q, Xing J, Li J, et al. A Device-Free Number Gesture Recognition Approach Based on Deep Learning[C]//2016 12th International Conference on Computational Intelligence and Security (CIS), Wuxi, China, IEEE, 2016:57-63.
[16] Ma Y S, Zhou G, Wang S Q, et al. SignFi: Sign Language Recognition Using WiFi[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018,2(1):23.
[17] Tang Z, Liu Q, Wu M, et al. WiFi CSI Gesture Recognition Based on Parallel LSTM-FCN Deep Space-Time Neural Network[J]. China Communications, 2021, 18(3):205-215.
[18] Zhang Y, Zheng Y, Qian K, et al. Widar3. 0: Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11):8671-8688.
[19] Gou J P, Yu B S, Maybank S J, et al. Knowledge Distillation: A Survey[J]. International Journal of Computer Vision, 2021, 129(6):1789-1819.
[20] Yang J, Zou H, Cao S, et al. MobileDA: Toward Edge-Domain Adaptation[J]. IEEE Internet of Things Journal, 2020, 7(8):6909-6918.
[21] Patil P, Patil M R, Itraj S, et al. A Review on MIMO OFDM Technology Basics and More[C]//2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, India, IEEE, 2017:119-124.
[22] Qian K, Wu C, Zhou Z, et al. Inferring Motion Direction Using Commodity Wi-Fi for Interactive Exergames[C]// 2017 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, ACM, 2017:1961-1972.
[23] Wang W, Liu A X, Shahzad M, et al. Understanding and Modeling of WiFi Signal Based Human Activity Recognition[C]//21st Annual International Conference on Mobile Computing and Networking, New York, NY, USA, ACM, 2015:65-76.

黄子非(1996—),男,上海人。上海工程技术大学硕士研究生,主要研究方向为无线感知、信号处理等;

朱 海(1992—),男,江苏泰州人。南京邮电大学工学博士,上海工程技术大学讲师。主要研究方向为无线感知、信号处理等;



龚浩成(1998—),男,江苏苏州人。上海工程技术大学硕士研究生,主要研究方向为无线感知、信号处理、机器学习等,njuptzhuh@ 163.com;

吴　飞(1968—),男,国防科大自动化控制博士后、南开大学(广发证券)金融数学博士后、上海工程技术大学教授。主要研究方向为计算机并行处理与节能控制、定位技术、基于无线信号的室内情景感知技术等。

杨明泽(1996—),男,甘肃人。上海工程技术大学硕士研究生,主要研究方向为无线感知、深度学习等;