DOI: 10.13382/j.jemi.B2407427

# 基于 DSC-SGRU 模型的 Wi-Fi 手势识别系统研究*

何育浪[1,2] 赵志彪[1,2] 李 振[1,2] 李珊珊[1,2]
(1. 天津职业技术师范大学自动化与电气工程学院 天津 300222;
2. 天津市信息传感与智能控制重点实验室 天津 300222)

**摘 要**:Wi-Fi 无线感知技术已成为感知领域的研究热点,能够实现对人体活动和周围环境的智能感知。现有的无线感知模型参数量较大,在移动设备等算力有限的场景中难以实时感知。为此,提出了一种基于深度可分离卷积的轻量级特征提取模块和堆叠的门控循环单元混合的分类识别模型。首先基于深度可分离卷积构建了轻量的特征提取模块,用以捕获人体手势的空间特征,并保持特征的时序性不发生变化;然后使用双层堆叠的 GRU 网络学习人体手势的时空特征;最后使用开源数据集 Widar 对模型的性能进行验证,提取 CSI 信息中的 BVP 特征以提高跨域场景的识别准确率,并利用加权的损失函数来解决样本不均衡问题。结果表明,提出的模型在跨域场景下准确率达到 77.6%,参数量仅有 236.891 K。与现有的其他 Wi-Fi 手势识别模型相比,提出的模型在性能基本保持不变的情况下,极大地降低了模型的参数和计算复杂度,为 Wi-Fi 无线感知技术在实际应用中的推广奠定了基础。

**关键词**: 无线感知;信道状态信息(CSI);深度学习;手势识别;跨域。
**中图分类号**: TP183;TN83 **文献标识码**: A **国家标准学科分类代码**: 520.20.40

# Research on Wi-Fi gesture recognition system based on DSC-SGRU model

He Yulang[1,2] Zhao Zhibiao[1,2] Li Zhen[1,2] Li Shanshan[1,2]
(1. School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China; 2. Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin 300222, China)

**Abstract**: Wi-Fi wireless sensing technology has become a research hotspot in the field of perception, which can realize intelligent perception of human activities and the surrounding environment. The existing wireless sensing models have a large number of parameters, which makes it difficult to sense in real-time in scenarios with limited computing power such as mobile devices. To this end, a classification and recognition model based on a mixture of a lightweight feature extraction module based on depth-separable convolution and a stacked gated recurrent unit is proposed. Firstly, a lightweight feature extraction module based on depth-separable convolution is constructed to capture the spatial features of human gestures and keep the temporal nature of the features unchanged; then the spatio-temporal features of human gestures are learned using a two-layer stacked GRU network; finally, the performance of the model is validated using the open-source dataset Widar, and the BVP features in the CSI information are extracted to improve the recognition of cross-domain scenes accuracy, and a weighted loss function is utilized to solve the sample imbalance problem. The results show that the proposed model achieves an accuracy of 77.6% in cross-domain scenarios with a parameter count of only 236.891 K. Compared with other existing Wi-Fi gesture recognition models, the proposed model greatly reduces the parameters and computational complexity of the model while its performance remains basically unchanged, which lays a foundation for the popularization of the Wi-Fi wireless sensing technology in practical applications.

**Keywords**: wireless sensing; channel state information; deep learning; gesture recognition; cross-domain

收稿日期: 2024-04-13 Received Date: 2024-04-13
*基金项目:国家自然科学基金(62103301)、天津市自然科学基金(22JCQNJC01100)、天津市教育委员会科研项目基金(2020KJ119)和天津市研究生科研创新项目(2022SKYZ296)资助

## 0 引 言

近年来,随着信息技术的发展,人机交互的场景日益增多[1-2]。手势作为人机交互的重要媒介,已成为学术界和工业界研究的热点之一[3]。目前,手势识别主要通过传感器[4-5]、摄像机[6-7]、雷达[8]和 Wi-Fi[9]等设备实现。其中,基于传感器[10-11]的手势识别方案要求用户佩戴传感器,导致其应用场景受限;基于摄像机等视觉传感器识别手势容易受到光照和遮挡的影响导致识别性能大打折扣。此外,摄像机还可能引起用户隐私泄露的问题;尽管使用雷达设备可以获得较高的准确度,但其部署成本较高。相比之下,使用 Wi-Fi 具有诸多优势。首先,Wi-Fi 得到了广泛普及,极大地降低了部署成本。其次,Wi-Fi 信号对于光照具有极强的鲁棒性,可以作为活动感知的稳定信号源。此外,Wi-Fi 已成功应用于环境感知[12]、占用检测[13]、活动识别[14]、室内定位[15]、人群计数[16]、手势识别[17]、人体生命信号检测[18]等方面。

基于 Wi-Fi 的无线感知主要以接收信号强度(received signal strength indicator, RSSI)[19]和信道状态信息(channel state information, CSI)为主要指标。但是目前的研究发现,RSSI 信号受多径效应影响,难以识别人体的细微变化。相比之下,CSI 通过捕获包含正交频分复用(orthogonal frequency-division multiplexing, OFDM)子载波以及每对发射-接收天线之间的信号幅值和相位信息,可以识别更细粒度的动作。目前,CSI 信息可以从配备 Intel 5300 NIC[20]和 Atheros AR9590 NIC[21]的设备或者从 Nexmon 和 OpenFWWF 等设备中获取[22]。

许多传统的机器学习技术已被广泛用于手势识别[23-25]。然而,这些方法通常依赖于专家知识设计的特征[26]。近年来,深度学习技术的发展促进了手势识别技术的革新,与传统机器学习方法相比,深度学习方法具有更高的准确性[27]。然而,基于深度学习的方法面临着一些挑战。随着深度学习理论的不断发展,模型参数量增大,导致模型的推理和部署成本较高,容易产生过拟合问题。此外,无线信号容易受到不同领域因素(包括噪声、系统部署位置、环境和用户)的影响,导致感知模型在新环境的泛化性较差,即模型缺乏跨域感知能力[28]。

针对深度学习模型参数量大的问题,Kabir[29]等提出了一种基于异构深度学习的轻量级手势识别系统 CSI-DeepNet。该系统利用深度可分离卷积神经网络(DS-Conv)的特征注意块(FA)和残差块(RB)提取细粒度特征,并使用较少的模型参数降低复杂度。虽然 CSI-DeepNet 在使用低功耗片上系统(SoC)ESP-32 收集字母数字手势的 CSI 信息方面表现出色,但没有考虑跨域识别时模型性能下降的问题。为了解决场景变换对感知性能的影响,Xiao 等[30]提出了基于 Mean Teacher 的跨域人类活动识别框架 WiTeacher。在此框架中,为了解决源域和目标域之间的特征偏移问题,构建了基于标签平滑的分类损失。通过 StyleGAN 生成类目标样本作为输入数据。为了增强模型的鲁棒性,设计了一种基于样本关系的正则化项,以保持两个样本的距离在有扰动和无扰动的情况下保持不变,利用样本之间的关系来提高识别性能。实验表明,WiTeacher 在不需要目标域任何注释数据的情况下模型性能有明显地提升。然而,WiTeacher 由于设计了标签动态调整相应标签值和基于样本关系的正则化项,使得整个系统较为庞大。Zhang 等[20]提出了 Widar3.0,一种具有一定跨域感知能力的手势识别系统。其关键思想是在 CSI 的多普勒频移(doppler frequency shift, DFS)上推导和提取手势与域无关的特征,使用这些特征训练的通用感知模型 DNN 在实验中表现良好。WiTeacher 和 Widar3.0 都是为了解决模型的跨域问题提出的,但都没有考虑模型的轻量化问题。

基于上述 Wi-Fi 手势识别面临的问题,提出了一种基于深度学习理论的 DSC-SGRU 混合网络模型来构建手势识别系统。首先,描述 CSI 相关理论以及手势如何影响 CSI 的变化。其次,处理了手势数据集的 BVP 信息,并建立 DSC-SGRU 模型。该模型利用由 DSC(depthwise Separable Convolution, DSC)组成的轻量级特征提取模块,捕捉手势 CSI 信息的空间特征而不破坏特征的时序性。提取的特征经过具有时序处理能力的 SGRU(stacked gate recurrent unit, SGRU)模块来处理手势的时序信息。最后,将 DSC-SGRU 与基准模型 GRU、ViT 以及其他组合模型进行对比实验,证明了在更少的参数量(Params)和每秒浮点运算次数(Flops)的情况下仍有较高的识别精度。

## 1 相关理论

CSI 描述了每个子载波上的多径相移和幅值衰减。设 $f$ 为载波频率,$X(f,t)$ 和 $Y(f,t)$ 分别为发送端和接收端的频域信号。它们之间的关系如式(1)所示。

$$Y(f,t) = H(f,t) \times X(f,t) + n(f,t) \tag{1}$$

其中,$n(f,t)$ 表示高斯白噪声,$H(f,t)$ 表示在时刻 $t$ 的 CSI 测量值,其可以通过 $X(f,t)$ 和 $Y(f,t)$ 估计出来。第 $i$ 条子载波的 CSI 测量值 $H_i$ 如式(2)所示。

$$H_i = I_i + jK_i = |H_i| \exp(j\angle H_i) \tag{2}$$

其中,$I_i$、$K_i$ 分别为子载波 $i$ 的同相分量、正交分量,$|H_i|$ 和 $\angle H_i$ 分别为子载波 $i$ 的振幅和相位。考虑到 CSI 信号在室内环境传播时存在多个不同路径而导致到达时

间、幅度和相位不同的信号分量,这些不同路径可以由信号在传播过程中反射、折射、衍射等引起。因此 $H_i$ 可以转化如式(3)所示。

$$H_i = \sum_{q=1}^{Q} r_q \cdot e^{-j2\pi f\tau_q} \cdot e^{-j2\pi\Delta f t} \tag{3}$$

其中,$Q$ 是多径分量的个数。$r_q$ 表示第 $q$ 条路径的初始相位偏移和衰减。$\tau_q$ 是第 $q$ 条链路上的传播延迟,发射载波频率和接收载波频率之差 $\Delta f$ 引起的相移用 $e^{-j2\pi\Delta f t}$ 表示。手势的多径效应如图 1 所示。

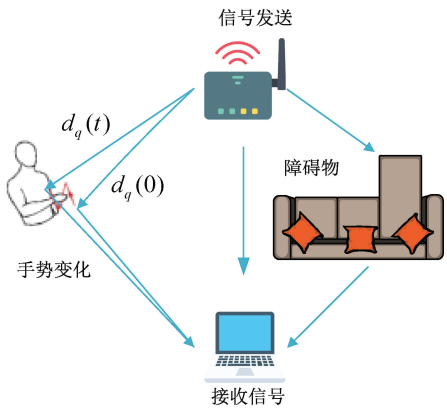



图 1　手势 CSI 的多径效应图

Fig. 1　Map of multipath effects of gesture CSIs

图中的用户正在沿红色箭头方向绘制字母 N。右边是以沙发为代表的环境障碍物。Wi-Fi 发射的无线信号经过手势和环境会被 Wi-Fi 接收端接收。滤除环境的影响,就可得手势与 CSI 信息之间的映射关系。已知手势运动引起的 CSI 变化由第 $q$ 条主要路径表示。当手势从时刻 0 变化到时刻 $t$ 时,第 $q$ 条路径的长度从 $d_q(0)$ 变化到 $d_q(t)$。$\lambda$ 和 $v_q$ 分别表示 CSI 信号波长和手势在短时间内的移动速度。可得 $\lambda = c/f, d_q(t) = d_q(0) + v_q(t)$ 和 $\tau_q = d_q(t)/c$,其中 $c$ 表示 CSI 信号以光速传播。相移 $e^{-j2\pi f\tau_q}$ 可以转化为 $e^{-j2\pi d_q(t)/\lambda}$,这表示 CSI 路径长度改变一个波长,接收到的子载波相位将变换 $2\pi$。CSI 可以分解为动态 CSI 和静态 CSI,分别定义为 $H_d(f,t)$ 和 $H_s(f,t)$。CSI 信号通过环境障碍物等静态物体时产生静态 CSI,通常被视为一个常量。动态 CSI 是受运动物体影响而变化的路径总和,也是分辨不同人体活动的关键,其计算公式如式(4)所示。

$$H_d(f,t) = \sum_{q\in P_d} r_q \cdot e^{-j2\pi d_q(t)/\lambda} \tag{4}$$

其中,$p_d$ 表示受手势影响的动态路径之和。子载波 $i$ 的 CSI 信息 $H_i$ 则转化为如式(5)所示。

$$H_i = (H_d(f,t) + H_s(f,t)) \cdot e^{-j2\pi\Delta f t} = \left(\sum_{q\in P_d} r_q \cdot e^{-j2\pi d_q(t)/\lambda} + H_s\right) \cdot e^{-j2\pi\Delta f t} \tag{5}$$

由于 CSI 的功率 $H_i^2$ 与多径的长度变化密切相关,为解决相位干扰 $e^{-j2\pi\Delta f t}$,使用 CSI 的功率 $H_i^2$ 替代 $H_i$,$H_i^2$ 如式(6)所示。

$$H_i^2 = |H_s(f,t)|^2 + \sum_{q\in P_d} |r_q|^2 + \sum_{q\in P_d} 2|H_s \cdot r_q| \cos\left(\frac{2\pi\nu_q t}{\lambda} + \frac{2\pi d_q(0)}{\lambda} + \phi_{sq}\right) + \sum_{\substack{q,e\in Pd \\ q\neq e}} 2|r_q \cdot r_e| \cos\left(\frac{2\pi(\nu_q - \nu_e)t}{\lambda} + \frac{2\pi(d_q(0) - d_e(0))}{\lambda} + \phi_{qe}\right) \tag{6}$$

其中,$\frac{2\pi d_q(0)}{\lambda} + \phi_{sq}$、$\frac{2\pi(d_q(0) - d_e(0))}{\lambda} + \phi_{qe}$ 分别代表初始相位偏移的常数值。CSI 总功率由一系列正弦信号和一个恒定偏移组成。正弦波的频率是路径长度变化速度的函数,由于人体活动会影响 CSI 路径长度的变化,因此可以得到人体活动与 CSI 信号频率之间的映射关系。通过提取 CSI 信号频率变化的特征就可以反应人体活动的变化,这表明利用 CSI 信号的变化进行手势识别是可行的。

图 2 中的第 1 行和第 2 行分别是 CSI 信息的幅值和相位。这些幅值和相位曲线由 30 条的子载波组成。不同手势的幅值变化较大,而相位的变化则比较细微。每一种手势都具有区别于其他手势的特征。不同的手势具有不同的模式,这为使用深度学习模型进行手势识别提供了可能。

## 2　基于 DSC-SGRU 的手势识别模型

为充分挖掘 CSI 信息中关于手势的时空特征,首先使用由 DSC 组成的轻量级特征提取模块来提取 CSI 信息的空间特征。通过 DSC 的深度卷积 DC 在通道内进行卷积操作,再利用点卷积 PC 按时间维度将空间特征融合。然后将提取的特征传入 SGRU 中,让 SGRU 网络充分学习 CSI 信息中的时空特征,使不同的网络发挥各自的专长做特定的任务。

### 2.1　数据预处理

采用目前已知手势的最大开源数据集 Widar 作为模型的数据,其相关参数见表 1。Widar 分别在教室,大厅和办公室环境下采集多个志愿者的 22 种手势,通过将不同域中的手势数据划分为训练集和测试集来验证模型的域泛化能力。人体反射信号的 ToF、AoA 和衰减等常用指标均与域高度相关,因此采用这些指标难以有效区分不同域下的手势。每种手势在传感空间中有其独特的速度分布,利用速度分布特征可以剥离静态的域信息,因此可以用作区分手势的指标。

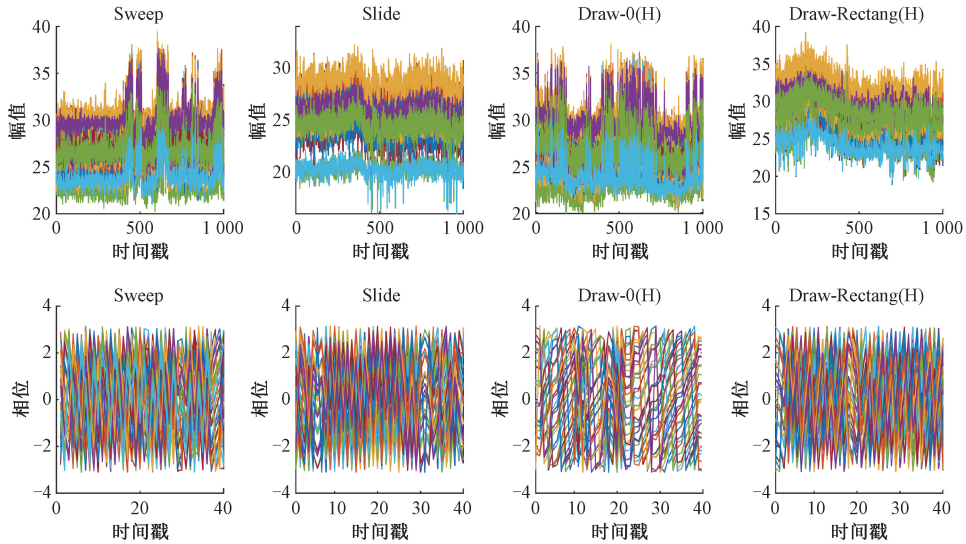



图2　手势的CSI信息

Fig. 2　CSI information of gestures

**表1　Widar数据集相关参数**

**Table 1　Parameters related to the Widar dataset**

| 名称 | Widar |
|---|---|
| 采集工具 | Intel 5300网卡 |
| 手势类别 | 22 |
| 手势名称 | Push&Pull, Sweep, Clap, Slide等22种 |
| 数据类型 | BVP |
| 数据尺寸 | (22,20,20)分别代表(time, x_velocity,y_velocity) |
| 训练样本数 | 34 926 |
| 测试样本数 | 8 726 |

而DFS体现了大部分速度分布信息。但是DFS还是与人的位置和方向高度相关。如图3展示了手势推拉(Push&Pull)在6个不同信号接收器采集的CSI信息DFS图,观察可得同一类手势在不同Wi-Fi接收端的DFS是不同的。

采用BVP信息作为深度学习模型的输入特征。BVP信息从DFS的主要速度分量中采用压缩感知的方法得到。人做手势时在身体坐标系中物理速度的功率分布仅与手势的特征相关,BVP通过推导出信号功率在身体坐标系中速度分量的分布来消除域的影响。图4表示提取到Push&Pull的BVP特征。由于Push&Pull手势具有14个时间帧,因此有14张特征图。由图发现与域无关的BVP信息特征是高度抽象的。模型通过对这些特征的学习获得了分辨手势的能力。由于每种手势的时序大小不一,为了便于后续模型的训练,将手势的时间戳统一扩充为22帧。训练集和测试集按照8∶2的比例划分。

训练集中各个类别样本的分布数量如图5所示。各

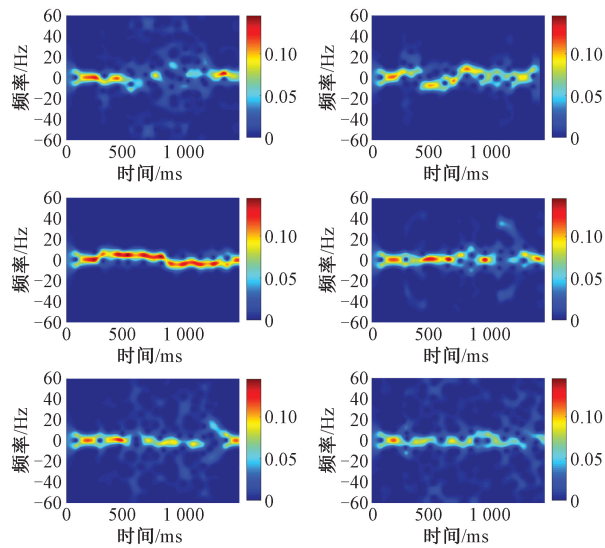



图3　Push&Pull手势6个不同位置的DFS图

Fig. 3　DFS diagram of Push & Pull gesture for 6 different positions

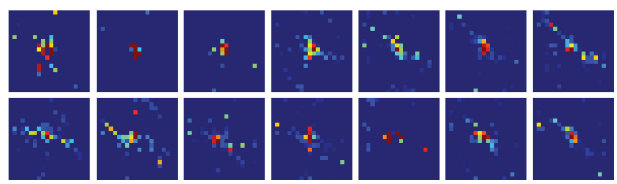


图4　Push&Pull手势BVP特征图

Fig. 4　Push & Pull gesture BVP feature map

个手势样本数存在不均衡的情况。在训练深度学习模型时,样本不均衡可能会导致模型对多数类别的预测效果

更好,而对少数类别的预测效果较差。为了解决这个问题,实验中通过计算不同手势样本数据之间的权重,让损失函数根据权重来平衡样本不均衡的问题,使得模型更加关注少数类别。

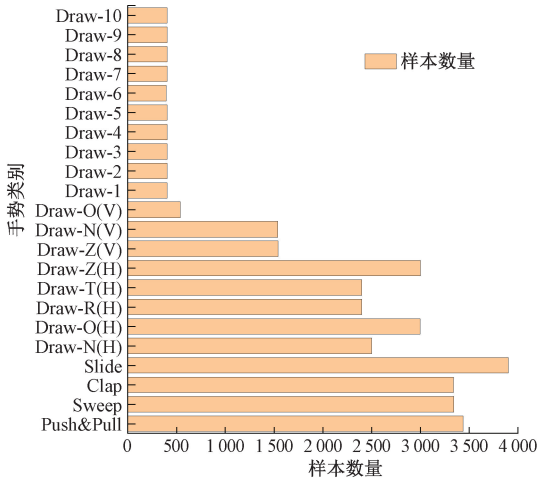



图 5　训练数据各个手势分布

Fig. 5　Distribution of training data for each gesture

## 2.2　DSC-SGRU 混合模型

手势的 BVP 数据首先进入以 DSC 为主要组成部分的轻量级特征提取模块进行手势特征提取。特征提取模块结构如图 6 中 DSC 特征提取模块所示。

它由两个 DSC,ReLu 激活函数,批量归一化层(BN),一个最大池化层和全连接层组成。全连接层由线性层,ReLu 激活函数,BN 层和 Dropout 层组成,网络的相关参数如表 2 所示。

**表 2　网络相关参数**

**Table 2　Network related parameters**

| 参数 | 数值 |
|---|---|
| 学习率 | 0.001 |
| DSC 中的卷积核大小 | 6×6,3×3 |
| 丢弃率 | 50% |
| 训练,测试批次大小 | 1 024,64 |
| 优化器 | Adam |
| 激活函数 | ReLu |

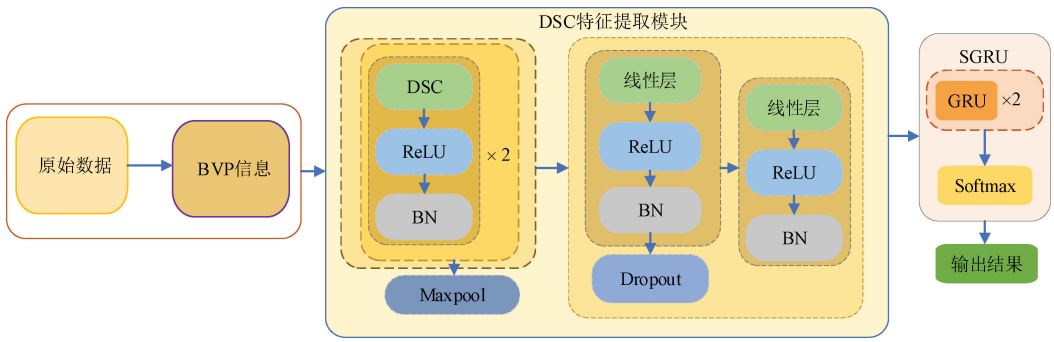



图 6　模型总体流程图

Fig. 6　Overall flow chart of the model

DSC 的详细结构如图 7 所示,由深度卷积和点卷积两部分组成。将 BVP 手势信息的 22 帧时序信息作为通道,即 $M = 22$。保持通道数不变的情况下在通道中进行卷积操作。DC 在逐帧的提取手势空间特征的同时保持了时序信息不变,然后提取的空间特征通过 $N$ 个大小为 1×1 的卷积核进行 PC。将前 M 通道的信息融合,转换为 $N$ 通道输出,点卷积按照时间维度融合,这样处理在弥补了 SGRU 网络特征提取能力不足的同时又最大程度的保留 BVP 数据的时序信息,使 SGRU 仍然可以按时序信息学习手势的特征。相较于传统卷积神经网络 CNN,DSC 拥有更少的参数量[31]。使得模型推理和验证速度都更快,这对于模型的实时性识别和大规模部署都非常有利。此外,DSC 网络结构使它具有更抽象的特征表示,更容易避免过拟合。因此,DSC 可以更好地推广到不同的数据集和任务。这对于日后将模型部署到新域中是十分有利的。由于 DSC 可以分别处理空间特征和通道特征,这使得模型能够更好地捕捉不同尺度和通道之间的相关性,从而提高模型的表征能力。

GRU[32]是循环神经网络(RNN)中基于两个门向量的一种变体,包括更新门和重置门。重置门用于忘记无用的信息,而更新门则侧重于将必要和有用的信息从前一个时间步传递到当前时间步。

单层的 GRU 是一种浅层模型,表征能力较弱。而堆叠的 GRU(stacked GRU, SGRU)由多个 GRU 单元堆叠而成,表征能力更强。通过增加网络的深度,SGRU 可以捕捉长时间序列信息,从而具有学习复杂序列模式的能力。SGRU 结构如图 8 所示。

第 $i$ 个 GRU 单元的更新门、重置门、历史隐藏状态和最终隐藏状态分别如式(7)~(10)所示。

$$z_t^i = \sigma(W_z^i \cdot [h_{t-1}^i, h_t^{i-1}]) \tag{7}$$

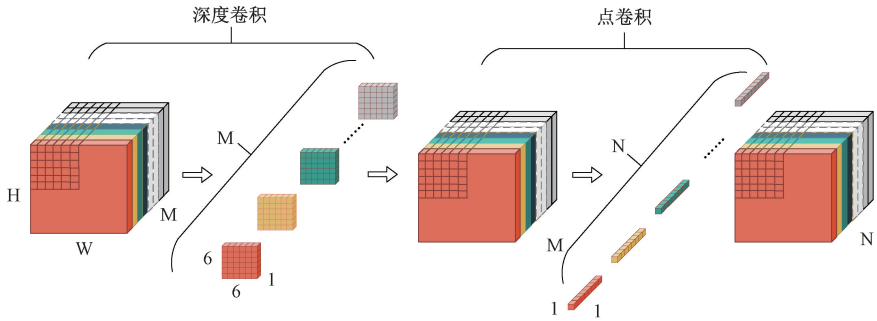



图7　深度可分离卷积示意图

Fig. 7　Schematic diagram of deep separable convolution

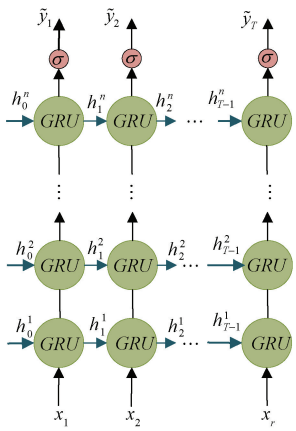



图8　SGRU结构图

Fig. 8　Structure of SGRU

$$r_t^i = \sigma(W_r^i \cdot [h_{t-1}^i, h_t^{i-1}]) \tag{8}$$

$$\tilde{h}_t^i = \tanh(W^i \cdot [r_t^i \cdot h_{t-1}^i, h_t^{i-1}]) \tag{9}$$

$$h_t^i = z_t^i \cdot h_{t-1}^i + (1 - z_t^i) \cdot h_t^{i-1} \tag{10}$$

其中，下标 $t$ 代表时间，SGRU 输出结果 $\tilde{y}_{last}$ 如式(11)所示。

$$\tilde{y}_{last} = \sigma(w_o^n h_o^n + b_o^n) \tag{11}$$

其中，$\tilde{y}_{last}$ 为第 $i$ 个样本的预测标签，$W_o^n$ 为第 $n$ 个 GRU 单元输出层权重，$b_o^n$ 为第 $n$ 个 GRU 单元的偏置。

实验总体流程如图6所示。首先从原始CSI信息的DFS中提取出与域无关的BVP信息。然后进入特征提取模块。该模块由两个DSC模块、一个最大池化层和一个全连接层组成。每个DSC模块包括DSC、ReLU激活函数和BN层。全连接层由两个线性层和一个Dropout层组成。特征提取模块用于提取手势的空间特征。然后特征进入SGRU模块学习手势的时序信息。SGRU设计为两个GRU单元的堆叠，其中第一个GRU单元的隐藏层输出作为第二个GRU单元的输入。这种结构增强了模型的表征能力和对长时间序列信息的推理能力。最后，经过Softmax层输出识别结果，实现人体手势的分类识别。SGRU模块能够对输入数据按照时序信息的进行训练，2层GRU单元也使其能够更好地捕捉复杂序列信息之间的长期依赖性。相较于传统深度学习模型，由于构造轻量化且高效的特征提取模块DSC和SGRU，使得DSC-SGRU模型参数量很少，仅有236.891 K。这样规模的参数量和Flops模型在一般的边缘计算终端，移动设备上可以流畅运行。

## 3　实验与评估

### 3.1　实验环境配置

实验搭建在Ubuntu 18.04平台。该平台配备两张NVIDIA GeForce RTX 3090显卡，共48 G显存。在Pytorch1.12.1深度学习框架下搭建模型，CUDA版本为11.6。

### 3.2　损失函数

DSC-SGRU模型采用交叉熵损失函数(cross entropy loss，CEL)。它是一种适用于多类别分类任务的损失函数，并且能够有效地衡量模型预测与真实标签之间的差异。由于Widar数据集存在样本不均衡问题，采用加权损失函数的方法来解决，即对每个样本的损失乘以其对应类别的权重。假设第 $i$ 个样本的损失为 $L_i$，则整体损失计算如式(12)所示。

$$CEL_{weighted} = \frac{1}{n}\sum_{i}^{n} w_i \times L_i \tag{12}$$

其中，$w_i$ 是第 $i$ 个样本所属类别的权重。将类别权重设置为该类别样本数量与总数量比例的倒数。较少出现的类别将会被赋予较高的权重，从而平衡训练过程中的样本不均衡性。

### 3.3　评价指标

评估实验中采用准确率(accuracy，Acc)、精度(precision，Pr)、F1分数(F1 score)、召回率(recall)、Params和Flops来综合评估模型的性能。其中，准确率

反应了模型的测试结果，Params 表示模型中可学习的权重参数的总数量，而 Flops 则用于评估模型在训练时每秒执行的浮点数运算次数，被用来评估模型的计算复杂度和效率。

### 3.4　实验结果分析

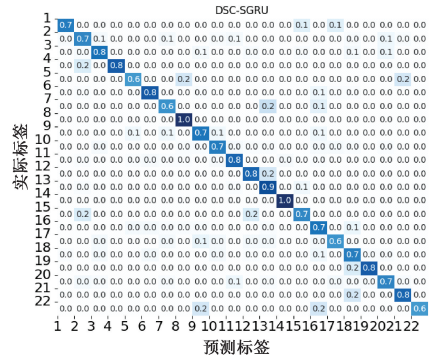
将 LeNet、MLP、GRU、AlexNet、VGG-16、GoogLeNet、ResNet18、ResNet50 和 ViT（vision transformer）等作为基准模型，与 DSC-SGRU 进行对比实验，以验证 DSC-SGRU 模型的性能。AlexNet 包含 8 个卷积层和 3 个全连接层，并且引入了局部响应归一化（local response normalization，LRN）层来抑制过度激活。GoogleNet 使用了 Inception 模块，可以同时处理不同大小的特征图。并且通过 1×1 卷积层降维，以减少计算量和参数数量。VGG-16 有 16 层深，包含 13 个卷积层和 3 个全连接层，所有卷积层都使用小尺寸的 3 × 3 卷积核和 ReLU 激活函数。并且使用了多个池化层来逐渐减少特征图的空间尺寸。ViT 是 transformer 的一种变体，其核心组件包括自注意力机制、多头自注意力机制、前馈神经网络、残差连接和层归一化。ViT 首先将输入数据分割成块（patches），然后将块嵌入成向量利用 Transformer 编码器处理这些嵌入向量，最后通过分类头进行图像分类。混淆矩阵图中数字与具体手势之间的对应关系如表 3 所示。

**表 3　序号对应的具体手势**

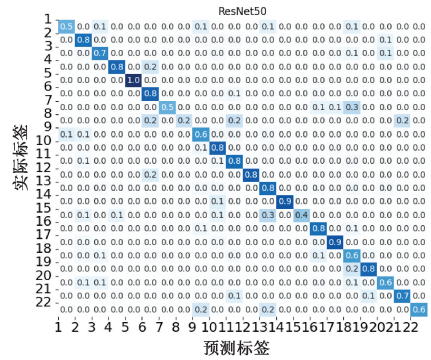
**Table 3　Specific gestures corresponding to serial numbers**

| 编号 | 手势 | 编号 | 手势 |
|---|---|---|---|
| 1 | Draw-3 | 12 | Draw-7 |
| 2 | Draw-N（H） | 13 | Draw-Z（H） |
| 3 | Draw-R（H） | 14 | Draw-1 |
| 4 | Draw-4 | 15 | Draw-2 |
| 5 | Draw-6 | 16 | Slide |
| 6 | Clap | 17 | Draw-N（V） |
| 7 | Draw-O（V） | 18 | Draw-O（H） |
| 8 | Draw-9 | 19 | Draw-10 |
| 9 | Draw-Z（V） | 20 | Draw-T（H） |
| 10 | Sweep | 21 | Draw-5 |
| 11 | Push&Pull | 22 | Draw-8 |

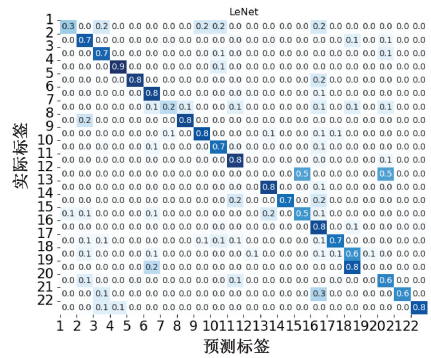
手势括号中的 V（vertical），H（horizontal）分别表示该手势在垂直方向或水平方向绘制。使用 Adam 优化器，它结合了 AdaGrad 和 RMSProp 两种优化器的优点。具有参数更新不受梯度伸缩变换的影响，更新的步长被限制在初始学习率的大致范围内和自动调整学习率等优点。通过多次试验逐次增加模型的训练轮数，当模型的训练精度及测试精度不再上升时，表明模型已充分收敛，终止模型的推理。DSC-SGRU 和 3 种表现较好的基准模型混淆矩阵如图 9 所示。4 种模型的测试准确率从高到底依次是 DSC-SGRU、ResNet50、LeNet、ResNet18。由于每种手势的 CSI 信息复杂度不同，因此导致同一模型对



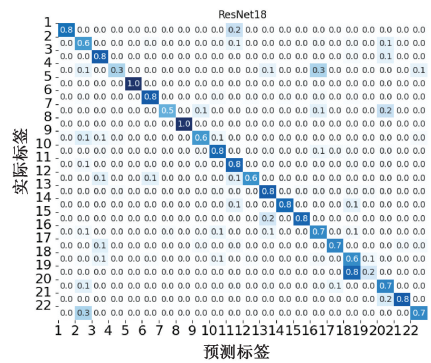

(a) DSC-SGRU模型混淆矩阵
(a) DSC-SGRU model confusion matrix




(b) ResNet50模型混淆矩阵
(b) ResNet50 model confusion matrix




(c) LeNet模型混淆矩阵
(c) LeNet model confusion matrix




(d) ResNet18模型混淆矩阵
(d) ResNet18 model confusion matrix

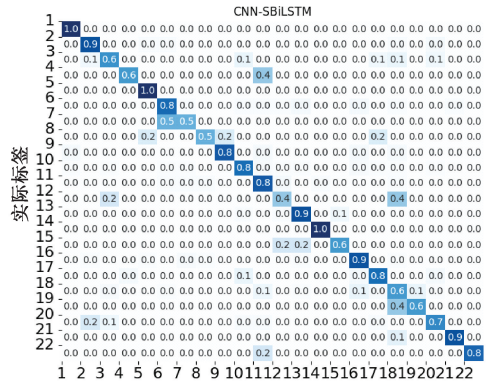
图 9　DSC-SGRU 与典型神经网络结果的对比

Fig. 9　Comparison between DSC-SGRU and typical neural network results

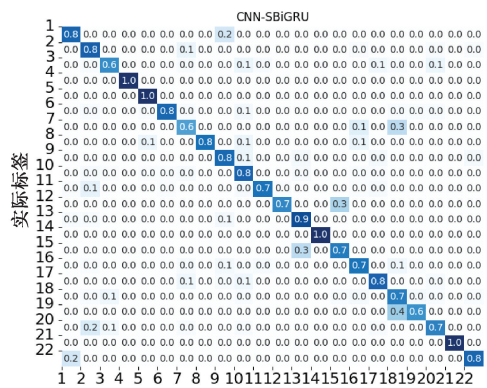
不同手势的识别结果表现各异。ResNet50 实验结果如图 9(b)所示,对 Draw-9 的识别准确率较差,仅有约 20%,对 Draw-2 的识别准确率只有约 40%,且有 30%的概率将 Draw-2 误判为 Draw-Z(H),对手势 Draw-3 的识别准确率约为 50%,对其余手势的识别准确率均在 60%以上。

LeNet 实验结果如图 9(c)所示,对 Draw-7 和 Draw-10 无法识别,并且对 Draw-3 和 Draw-O(V)识别结果极低。而 DSC-SGRU 对 Draw-7、Draw-3 和 Draw-10 识别准确率均在 70%以上,仅对 Draw-O(V)的识别准确率为 20%。ResNet18 实验结果如图 9(d)所示,对 Draw-4 的识别准确率约为 30%,且有 30%的概率将 Draw4 误判为 Slide,有 80%概率将 Draw-10 误判为 Draw-O(H)。而 DSC-SGRU 对手势 Draw-8, Draw-9, Draw-10, Draw-2 和 Draw-Z(H)识别准确率分别为 100%,60%,90%,100%和 90%。对手势 Draw-6 和 Draw-7 识别准确率在 70%以上。对手势 Draw-4, Draw-9, Draw-1, Draw-2, Draw-N(V)和 Draw-5 的识别准确率在 80%以上。其余基准模型如 MLP 由于没有提取时序特征的能力表现较差。GRU 虽然有按照时序信息自动的提取特征能力,但是对于无线信号的手势这类高维抽象的数据略显不足。AlexNet, GoogLeNet, VGG-16 和 ViT 总体表现较差,其中 GoogLeNet,AlexNet 和 VGG-16 的参数量和 Flops 都比较大,不符合构建轻量化模型的理念。ViT 虽然参数量较小,但由于自注意力机制和前馈网络的计算复杂度较高,导致其计算量较大,它的 Flops 是 9.302 M 是 DSC-SGRU 模型 Flops 的 2.7 倍。由于其性能与 CNN、MLP 和 GRU 相似,因此 ViT 不适合 Wi-Fi 传感的任务中。

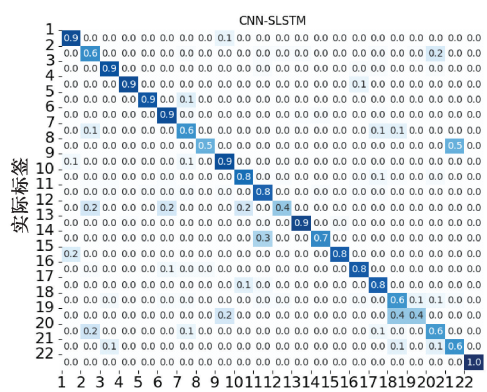
DSC-SGRU 模型,首先由 DSC 组成轻量级特征提取模块,在尽可能保持原始手势信息时序不变的情况下,提取手势特征。经过 DSC 提取的特征再进入 SGRU 网络,按照时序信息学习手势的抽象特征。与现有的手势识别模型相比,DSC-SGRU 在性能近似的情况下,极大的降低了模型的参数量。除了与传统神经网络进行对比实验外,还探究了不同网络模块组合的性能。以堆叠结构组合的模型有 SLSTM 等。以 CNN 组成的特征提取模块和具有时序处理能力的网络(RNN、LSTM、GRU)组合的模型有 CNN-BiLSTM 等。以 DSC 组成的特征提取模块和时序处理模块组合的模型有 DSC-GRU 等。为了对比实验的严谨性,以上对比模型的其余部分均与 DSC-SGRU 模型保持一致,包括训练批次,优化器,损失函数,学习率等。如构建 CNN-SBiLSTM 模型时先使用以 CNN 为主要构件的特征提取模块提取手势空间特征而尽可能保持手势时序特征不发生变化。然后提取的特征进入 SBiLSTM 网络学习特征的时序信息,SBiLSTM 表示两个 BiLSTM 堆叠而成,上一个 BiLSTM 网络隐藏层的输出作为下一个 BiLSTM 网络的输入。CNN-SBiGRU 等都是类似的构建思想。其中 3 种识别性能最好的模型结果如图 10 所示。




(a) CNN-SBiLSTM模型混淆矩阵
(a) CNN-SBiLSTM model confusion matrix




(b) CNN-SBiGRU模型混淆矩阵
(b) CNN-SBiGRU model confusion matrix



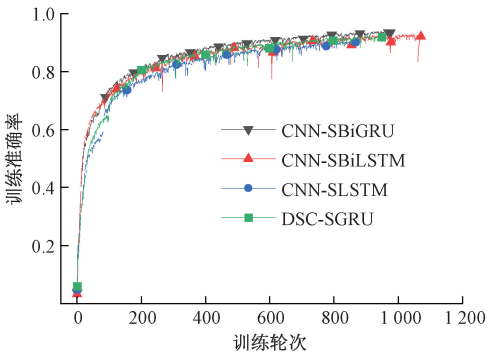

(c) CNN-SLSTM模型混淆矩阵
(c) CNN-SLSTM model confusion matrix

图 10　组合模型分类结果混淆矩阵图

Fig. 10　Combined model classification results

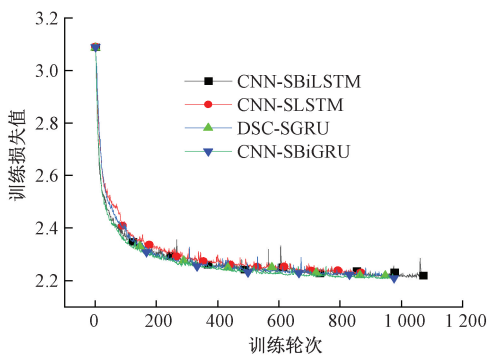
CNN-SBiLSTM 实验结果如图 10(a)所示,有 50%的概率将 Draw-O(V)误识别为 Clap,并且有 40%的概率将 Draw-7 识别为 Draw-O(H)。CNN-SLSTM 实验结果如图 10(c)所示,对 Draw-7 识别结果较差,并且有 40%的概率

会将 Draw-10 误认为 Draw-O(H)。CNN-SBiLSTM 和 CNN-SLSTM 都有将 Draw-10 误认为 Draw-O(H)的情况。这可能是因为绘制 10 的准备手势和在垂直方向绘制 O 的手势有一定的相似性。CNN-SBiGRU 实验结果如图 10(b)所示,对各类手势识别情况总体表现的比较均衡,没有无法识别或者识别率低于 50%的情况,但模型的总体识别准确率低于 CNN-SLSTM 和 CNN-SBiLSTM。4 种性能较好模型的训练曲线和损失曲线如图 12 所示。




(a) 4类最优模型的训练图

(a) Training graphs for 4 classes of optimal models




(b) 4类最优模型的损失图

(b) Loss map for 4 classes of optimal models

图 11　4 类测试准确率最优模型的训练和损失图

Fig. 11　Training and loss diagrams for the optimal model for the 4 classes of test accuracies

图 11(a)是 4 种最优模型的训练图,图 11(b)是 4 种最优模型的损失函数图,它们的训练曲线和损失曲线相似。前 200Epoch 快速的收敛,随后趋于平缓。为了减少单次实验的偶然性偏差对模型精度的影响,对性能最优的 4 种模型进行 3 次实验评估其系统性能。如图 12 所示。

4 种模型的测试准确率均在 77%以上。相较于其他 3 种模型,DSC-SGRU 表现更加稳定,推理结果波动较小。为了验证选取两层 GRU 堆叠的合理性,将 4 层 GRU 堆叠的模型与两层 GRU 堆叠的模型进行对比实验。如图 13 所示,DSC-SGRU-2 是使用两层 GRU 堆叠的实验曲线,DSC-SGRU-4 是使用 4 层 GRU 堆叠的实验曲线。图 13(a)是 2 层 GRU 和 4 层 GRU 模型的训练曲线和损失

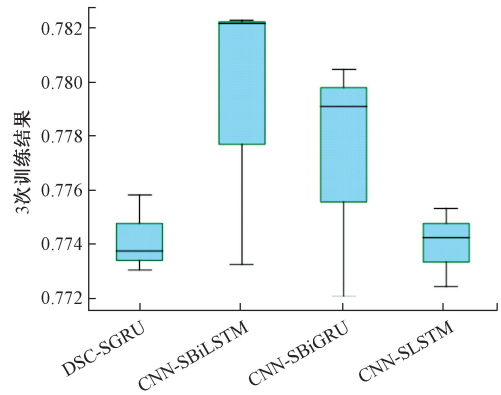
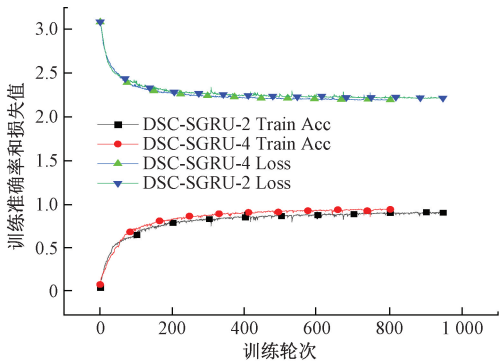



图 12　4 类准确率最高模型 3 次训练结果的箱线图

Fig. 12　Boxplots of the results of 3 training sessions of the models with the highest accuracies in the 4 categories

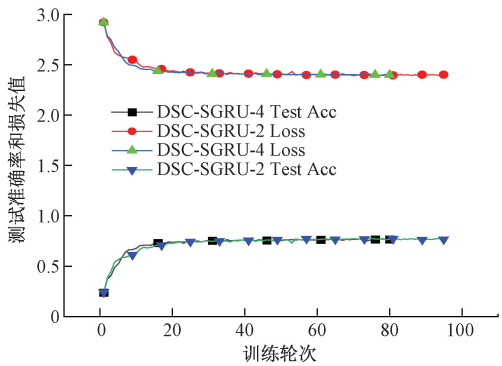
函数曲线,图 13(b)是每训练 10 个 Epoch 就验证一次的测试曲线。在测试准确率近似的情况下,4 层 GRU 的训练准确率明显较高,但是测试准确率却没有明显的上升。这说明选取 2 层 GRU 是合理性的。其余基准模型的实验结果如表 4 所示,模型 DSC-SGRU 在识别能力近似的情况下,具有最低的参数量和运算量。从表 4 中的对比实验结果可知,与传统的模型 MLP、LeNet 相比,DSC-SGRU 的识别结果更高,并且具有更低的参数量。由于使用了特征提取模块,与以 GRU 为代表的时间序列模型相比,DSC-SGRU 的识别准确率也更高。与以残差结构为代表的 ResNet 网络相比,DSC-SGRU 在减少参数量的情况下,保持了较优的识别性能。与 CNN-SBiLSTM、CNN-SBiGRU 和 CNN-SLSTM 等网络模块组合的模型对比中,虽然 DSC-SGRU 的准确率略低于其余 3 种模型,但是 DSC-SGRU 模型的参数量和 Flops 是最低的。DSC-SGRU 与 SGRU 相比 Flops 更小,需要的计算资源更少。与 CNN-SGRU 相比,DSC-SGRU 参数量在增加了约 18%的情况下表现出更优的性能。在对比实验中发现,时序模型的双向结构对识别结果的提升相较于增加的参数量是微乎其微的。例如,DSC-SGRU 与 DSC-SBiGRU 相比,识别结果提升的同时还减少了约 54%的参数量,即网络的双向结构在增加模型参数量的同时并没有大幅提升模型的性能。

DSC-SGRU 模型在识别准确率近似的情况下,模型的 Params 和 Flops 更少。Flops 和参数量的减少意味着该模型在同样的硬件环境下可以更快地推理和验证。从而节省计算资源。对于资源有限的环境,如移动设备等,该模型可以有效的降低计算和存储成本。参数量较少的模型也具有更低的过拟合风险,能够更好地泛化到新的数据集上。这对于提升模型在不同环境下的泛化能力和稳健性是至关重要的。




(a) DSC-SGRU2层和4层训练准确率和损失

(a) DSC-SGRU 2-and 4-layer training accuracies and loses




(b) DSC-SGRU2层和4层测试准确率和损失

(b) DSC-SGRU Layer 2 and 4 Test Accuracy and Losses

图 13 DSC-SGRU 模型 2 层 GRU 和 4 层 GRU 的对比结果

Fig. 13 Comparison results of 2-layer GRU and 4-layer GRU for the DSC-SGRU model

## 4 结 论

为了解决现有 Wi-Fi 感知模型参数量大和计算复杂度高的问题,提出了一种新的基于 Wi-Fi CSI 信息的手势识别模型 DSC-SGRU,采用由 DSC 组成的轻量级特征提取模块提取手势特征,再由 SGRU 学习手势的时空特征。实验表明,由 DSC 组成的轻量化特征提取模块能够有效降低模型参数量,同时能够更好地捕捉不同尺度和通道之间的相关性,从而提高模型的表征能力。此外,双层堆叠的 SGRU,在模型参数量和精度之间获得了良好的平衡。DSC-SGRU 模型的整体性能不亚于其他先进的手势识别模型,并且在推理、计算等方面都更加高效。由于使用多域的 Widar 数据集提取的 BVP 特征来推理模型,使模型具有一定的跨域感知能力。DSC-SGRU 轻量化特性为其能在边缘计算终端、移动设备等计算能力有限的场景下实现实时感知奠定了基础。未来将参考其他减少模型参数量的方法,如剪枝算法等,或更有效的特征提取算法如宽度可分离卷积进一步压缩 DSC-SGRU 模型的参数,并且采集更多不同域中的 CSI 信息以全面提升模型的跨域感知能力和实际应用的广泛性。同时探索将 DSC-SGRU 模型部署在实际环境中,验证模型在线实时识别能力。

表 4 各类模型的各项评价指标对比

Table 4 Comparison of various evaluation indicators for each type of model

| 模型 | 准确率/% | 精度/% | 召回率/% | F1 分数/% | 每秒浮点计算(Flops) | 参数量(Params) |
|---|---|---|---|---|---|---|
| MLP | 64.2 | 65.2 | 62.9 | 63.0 | 9.145 M | 9.146 M |
| LeNet | 71.5 | 70.3[4] | 68.8 | 68.8 | 3.370 M | 298.838 K |
| RNN | 43.2 | 47.9 | 46.5 | 46.4 | 658.944 K | 31.254 K |
| LSTM | 59.0 | 67.1 | 65.2 | 65.4 | 2.637 M | 120.726 K |
| BiLSTM | 57.3 | 58.4 | 55.6 | 56.3 | 5.273 M | 240.022 K |
| GRU | 58.2 | 59.6 | 59.4 | 58.8 | 1.980 M | 90.902 K |
| BiGRU | 54.6 | 59.3 | 57.6 | 57.2 | 3.958 M | 180.374 K |
| SRNN | 58.7 | 46.1 | 56.3 | 50.2 | 4.168 M | 190.486 K |
| SLSTM | 73.3 | 72.3 | 71.9 | 71.9 | 16.695 M | 757.654 K |
| SGRU | 70.4 | 71.9 | 70.9 | 70.8 | 12.528 M | 568.598 K |
| SBiGRU | 71.3 | 71.8 | 71.2 | 70.9 | 15.558 M | 706.326 K |
| ResNet18 | 71.3 | 69.7 | 71.1 | 69.7 | 50.524 M | 11.192 M |
| ResNet50 | 72.9 | 69.5 | 68.6 | 68.6 | 87.005 M | 23.583 M |
| ResNet101 | 67.8 | 69.1 | 67.4 | 67.6 | 163.183 M | 42.602 M |
| CNN-RNN | 42.0 | 28.3 | 43.4 | 33.5 | 2.163 M | 42.574 K |
| CNN-LSTM | 76.1 | 76.0 | 74.2 | 74.5 | 3.928 M | 117.374 K |
| CNN-GRU | 73.5 | 71.8 | 69.7 | 69.9 | 3.379 M | 92.542 K |
| CNN-BiLSTM | 76.2 | 77.3 | 75.8 | 76.2 | 6.040 M | 216.654 K |
| CNN-BiGRU | 73.7 | 76.5 | 74.2 | 74.6 | 5.038 M | 167.038 K |
| CNN-SLSTM | 78.7[2] | 80.8[1] | 78.9[1] | 79.4[1] | 4.062 M[15] | 249.470 K[19] |
| CNN-SGRU | 76.7 | 76.2 | 74.8 | 75.1 | 5.579 M | 191.614 K |

续表

| 模型 | 准确率/% | 精度/% | 召回率/% | F1 分数/% | 每秒浮点计算(Flops) | 参数量(Params) |
|---|---|---|---|---|---|---|
| CNN-SBiLSTM | 78.8[1] | 75.9[11] | 74.4[10] | 74.4[10] | 14.877 M[26] | 611.966 K[24] |
| CNN-SBiGRU | 78.5[3] | 73.7[14] | 76.9[3] | 77.1[3] | 5.534 M[20] | 365.230 K[21] |
| DSC-RNN | 54.0 | 63.4 | 64.8 | 62.8 | 2.175 M | 91.051 K |
| DSC-LSTM | 74.7 | 77.5 | 75.0 | 75.7 | 2.831 M | 116.187 K |
| DSC-GRU | 72.4 | 72.0 | 69.0 | 69.4 | 2.282 M | 91.355 K |
| DSC-BiLSTM | 74.1 | 75.2 | 72.9 | 73.5 | 5.039 M | 215.515 K |
| DSC-BiGRU | 72.4 | 75.3 | 71.3 | 72.7 | 3.940 M | 165.851 K |
| DSC-SLSTM | 76.5 | 76.9 | 75.8 | 76.0 | 13.779 M | 610.779 K |
| **DSC-SGRU** | **77.6[4]** | **76.7[7]** | **75.1[7]** | **75.5[9]** | **3.407 M[10]** | **236.891 K[18]** |
| DSC-SBiLSTM | 73.7 | 77.8 | 77.2 | 77.2 | 14.877 M | 611.966 K |
| DSC-SBiGRU | 76.1 | 77.2 | 76.6 | 76.4 | 5.534 M | 365.230 K |
| ViT | 63.0 | 63.0 | 59.2 | 59.5 | 9.302 M | 86.262 K |
| AlexNet | 57.4 | 59.0 | 57.4 | 57.4 | 228.206 M | 6.921 M |
| Vgg-16 | 43.2 | 45.2 | 43.2 | 43.3 | 229.689 M | 5.369 M |
| GoogLeNet | 60.8 | 62.0 | 61.9 | 61.5 | 34.946 M | 6.010 M |

## 参考文献


[1] ZHOU Y, JIANG G, LIN Y. A novel finger and hand pose estimation technique for real-time hand gesture recognition [J]. Pattern Recognition, 2016, 49: 102-114.

[2] 陈仁钧, 费敏锐, 杨傲雷. 面向人机交互的手势指向估计方法 [J]. 仪器仪表学报, 2023, 44(3): 200-208.
CHEN R J, FEI M R, YANG A L. Estimation of gesture pointing for human-robot interaction [J]. Chinese Journal of Scientific Instrument, 2023, 44(3):200-208.

[3] VENKATNARAYAN R H, PAGE G, SHAHZAD M. Multi-user gesture recognition using WiFi [C]. proceedings of the Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, 2018.

[4] BOTROS F S, PHINYOMARK A, SCHEME E J. Electromyography-based gesture recognition: Is it time to change focus from the forearm to the Wrist? [J]. IEEE Transactions on Industrial Informatics, 2022, 18(1): 174-184.

[5] 赵世昊, 周建华, 伏云发. 注意力机制 CNN 结合肌电特征矩阵的手势识别研究 [J]. 电子测量与仪器学报, 2023, 37(6): 59-67.
ZHAO SH H, ZHOU J H, FU Y F. Investigation of gesture recognition using attention mechanism CNN combined electromyography feature matrix [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(6):59-67.

[6] HUU P N, NGOC T L, MINH Q T. Proposing gesture recognition algorithm using two-stream convolutional network and LSTM[C]. Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), 2021.

[7] AMIRTHA VARSHINI A, BHAVANI G, VITHYA, et al. Real-time hand gesture recognition for robotic arm and home automation [C]. Proceedings of the 2021 International Symposium on Electrical, Electronics and Information Engineering, 2021.

[8] LIEN J, GILLIAN N, KARAGOZLER M E, et al. Soli: Ubiquitous gesture sensing with millimeter wave radar [J]. ACM Transactions on Graphics (TOG), 2016, 35(4): 1-19.

[9] YOUSEFI S, NARUI H, DAYAL S, et al. A survey on behavior recognition using WiFi channel state information [J]. IEEE Communications Magazine, 2017, 55(10): 98-104.

[10] CHU X ZH, LIU J, SHIMAMOTO S. A sensor-based hand gesture recognition system for Japanese sign language[C]. proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), 2021.

[11] 牛群峰, 石磊, 贾昆明, 等. 基于改进 ResNet50 的表面肌电信号手势识别 [J]. 国外电子测量技术, 2024, 43(4): 181-189.
NIU Q F, SHI L, JIA K M, et, al. SEMG gesture recognition based on improved ResNet50 [J]. Foreign Electronic Measurement Technology, 2024, 43(4): 181-189.

[12] HUANG J CH, DUAN N, JI P, et al. A crowdsource-based sensing system for monitoring fine-grained air quality in urban environments [J]. IEEE Internet of Things Journal, 2019, 6(2): 3240-3247.

[13] WU CH SH, WANG B B, AU O C, et al. Wi-Fi can do more: Toward ubiquitous wireless sensing [J]. IEEE Communications Standards Magazine, 2022, 6(2): 42-49.

[14] 刘苗苗, 樊春玲. 基于 WiFi 信号的老年人家居行为

识别算法[J]. 电子测量技术, 2023, 46(6): 185-192.
LIU M M, FAN CH L. Human activity recognition algorithm for elderly home based on WiFi signal [J]. Electronic Measurement Technology, 2023, 46(6): 185-192.
[15] RUAN Y L, CHEN L, ZHOU X, et al. iPos-5G: Indoor positioning via commercial 5G NR CSI [J]. IEEE Internet of Things Journal, 2023, 10(10): 8718-8733.
[16] CHOI H, FUJIMOTO M, MATSUI T, et al. Wi-CaL: WiFi sensing and machine learning based device-free crowd counting and localization [J]. IEEE Access, 2022, 10: 24395-24410.
[17] MENG W, CHEN X C, CUI W, et al. WiHGR: A robust WiFi-based human gesture recognition system via sparse recovery and modified attention-based BGRU [J]. IEEE Internet of Things Journal, 2022, 9(12): 10272-10282.
[18] JAEHYUN P, YEOJIN J, GAEUN L, et al. 915-MHz continuous-wave doppler radar sensor for detection of vital signs [J]. Electronics, 2019, 8(5): 561.
[19] ABDELNASSER H, YOUSSEF M, HARRAS K A. WiGest: A ubiquitous WiFi-based gesture recognition system[C]. proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM), 2015.
[20] 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, 2022, 44(11): 8671-8688.
[21] YANG J, CHEN X, ZOU H, et al. SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing [J]. Patterns, 2023, 4(3): 100703.
[22] GRINGOLI F, SCHULZ M, LINK J, et al. Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets [C]. Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, 2019.
[23] WANG W, LIU A X, SHAHZAD M, et al. Device-free human activity recognition using commercial WiFi devices [J]. IEEE Journal on Selected Areas in Communications, 2017, 35(5): 1118-1131.
[24] FENG C, ARSHAD S, LIU Y. Mais: Multiple activity identification system using channel state information of wifi signals[C]. proceedings of the Wireless Algorithms, Systems, and Applications: 12th International Conference, 2017.
[25] FU ZH J, XU J SH, ZHU ZH D, et al. Writing in the air with WiFi signals for virtual reality devices [J]. IEEE Transactions on Mobile Computing, 2018, 18(2): 473-484.
[26] MA Y S, ZHOU G, WANG SH G. WiFi sensing with channel state information: A survey [J]. ACM Computing Surveys (CSUR), 2019, 52(3): 1-36.
[27] CHEN Z, ZHANG L, JIANG C, et al. WiFi CSI based passive human activity recognition using attention based BLSTM [J]. IEEE Transactions on Mobile Computing, 2018, 18(11): 2714-2724.
[28] WANG F, GONG W, LIU J, et al. Channel selective activity recognition with WiFi: A deep learning approach exploring wideband information [J]. IEEE Transactions on Network Science and Engineering, 2020, 7(1): 181-192.
[29] KABIR M H, HASAN M A, SHIN W. CSI-DeepNet: A lightweight deep convolutional neural network based hand gesture recognition system using Wi-Fi CSI signal [J]. IEEE Access, 2022,10:114787-114801.
[30] XIAO CH J, LEI Y, LIU CH, et al. Mean teacher-based cross-domain activity recognition using WiFi signals [J]. IEEE Internet of Things Journal, 2023, 10(14): 12787-12797.
[31] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[32] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. ArXiv preprint arXiv:14123555, 2014.

## 作者简介

**何育浪**,2021年于陇东学院获得学士学位,现为天津职业技术师范大学研究生,主要研究方向为无线感知。
E-mail: 0521221006@tute.edu.cn

**He Yulang** received his B. Sc. degree from Longdong university in 2021. Now he is a M. Sc. candidate in Tianjin University of Technology and Education. His main research interest includes wireless sensing.

**赵志彪**(通信作者),2012年于燕山大学获得学士学位,2019年于燕山大学获得博士学位,现为天津职业技术师范大学讲师,主要研究方向为工业智能控制、机器学习。
E-mail: zhaozhibiao@tute.edu.cn

**Zhao Zhibiao** (Corresponding author) received his B. Sc. degree from Yanshan university in 2012 and received Ph. D. degree from Yanshan university in 2019, respectively. Now he is tutors in Tianjin University of Technology and Education. His main research interests include industrial intelligent control and machine learning.