基于CRNN的CSI动作识别

丁文超，阴庚雷，张俊宝

(中原工学院计算机学院，河南 郑州 4500071)

摘　要：随着Wi-Fi感知技术的发展，出现了大量使用Wi-Fi信道状态信息（Channel State Information, CSI）进行动作识别的应用。然而大多数的方法在数据预处理和训练阶段都依赖于人工构建特征，构建过程耗时耗力并且需要专家的领域知识。针对上述问题本文提出一种基于CRNN（Convolutional Recurrent Neural Network）的CSI动作识别方法。首先，将不同手势的CSI数据做低通滤波处理。接着通过自组织映射(Self Organizing Maps，SOM)聚类的结果选择最佳子载波，并对该子载波上的CSI数据进行扩增。然后，使用格拉姆角求和场 （Gramian Angular Summation Fields，GASF）方法将一维CSI数据转换成二维GASF图像，作为CNN、LSTM构成的CRNN网络的输入数据，训练过程中使用链接时序分类（Connectionist Temporal Classification, CTC）作为损失函数。实验结果表明本文提出的方法能在训练数据较少的情况下达到较高的识别精度，且无需手动构建特征。

关键词：信道状态信息；CRNN；动作识别；自组织映射；格拉姆角场；链接时序分类

中图分类号：TP391.41

CSI action recognition based on CRNN

Ding Wenchao，Yin Genglei，Zhang Junbao

(College of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China)

Abstract: With the development of Wi-Fi sensing technology, a large number of applications have emerged for motion recognition using Channel State Information (CSI) from Wi-Fi. However, most methods rely on manual construction of features during the data pre-processing and training stages, the construction process is often time-consuming and labor-intensive and requires expert domain knowledge. Aiming at these problems, this paper proposes a CSI action recognition method based on Convolutional Recurrent Neural Network (CRNN). First, the CSI data of the different gestures is processed by low-pass filtering. The best subcarriers were then selected by the results of Self Organizing Maps (SOM) clustering, and amplify the CSI data on that subcarrier. Then, the Gramian Angular Summation Fields (GASF) method was used to convert one-dimensional CSI data into two-dimensional GASF images as input data for the CRNN network composed of CNN and LSTM, and Connectionist Temporal Classification (CTC) was used as loss function during the training. The experimental results indicate that the method proposed in this paper can achieve high recognition accuracy with less training data, and there is no need to manually construct features.

Key words: CSI；CRNN；action recognition；SOM；GASF；CTC

1. 引言

随着Wi-Fi技术的发展，及相关硬件价格的下降，生活中各种场所几乎均有Wi-Fi信号的存在。Wi-Fi信号不仅能传输数据，利用其信道状态信息还能进行人体动作识别[1]。传统的动作识别需要借助如加速度计、陀螺仪等接触式传感设备[2]对用户进行监测来识别用户的动作，又或者是借助摄像头等非接触设备的基于视觉[3]的动作识别。相较于传统的识别方式，基于CSI的动作识别利用的是环境中广泛存在的Wi-Fi信号来进行识别，具有不依赖特殊传感器、成本低、易扩展、不存在隐私泄露等优点。

由于采集数据时网卡报告了多组子载波的CSI，且不同的子载波对于环境变化的敏感度不同，所以要选取合适的子载波来提取CSI数据。WiFall[4]将30组子载波上的CSI取平均值作为训练数据来检测摔倒动作的发生。LBR[5]选取CSI方差第二大的子载波作为有效子载波来提取CSI。上述方法均根据某个统计值来选择子载波，需要大量的时间研究数据从而选择合适的统计特征。本文提出一种基于SOM[6]聚类的子载波选择方法。相同的动作，某个子载波上的CSI数据聚类生成多个簇，则说明CSI存在多种变化模式数据间差异较大，数据采集过程中该子载波易受环境干扰。因此对于每个动作本文选取所有子载波聚类后生成簇较少的子载波，作为有效子载波。

在训练特征的选择上，CARM[7]使用离散小波变换从不同时间尺度的不同活动中提取频率分量作为特征。WiSee[8]从手势动作中提取对应的多普勒频移特征来识别手势。两种方法均需手动设计特征。本文提出使用GASF[9]方法将CSI转换成图像从而利用CNN强大的特征提取能力自动提取特征。

由于实验对象的运动速度以及采样频率的差异，相同的动作所收集到的CSI数据长度可能不同，但神经网络输入数据的长度需要保持一致。所以要对收集到的CSI数据进行截取，以保证数据长度相同，并且截取到的部分需要包含动作发生的整个过程。WiFinger[10]根据统计特征来检测手势动作的起始点和结束点。文献[11]直接采用了人工截取的方式。上述方法需要大量的人力与计算资源来截取数据。本文的CRNN网络使用了CTC[12]损失函数，从而能够使网络自动对齐输入数据中发生动作的部分，无需对数据进行截取。

深度学习网络需要大量的数据来进行训练从而防止模型过拟合，但是人工采集大量的数据往往是不容易的。因此本文提出通过添加噪音、缩放、滑动平均等方法增加数据量，从而提高模型的鲁棒性。

1. 信道状态信息

CSI代表了信号在传输过程中所产生的时延、幅度衰减以及相位偏移等变化。换句话说，CSI表征了附近环境的无线特性。假设发送端信号为，接收端信号为，那么有[1]



其中为信道频率响应矩阵，CSI就是对的估计。为高斯白噪声。对于单个子载波，CSI的形式如下



其中，和分别代表了CSI的幅度和相位。由于商用网卡接收到的两个连续的CSI间相位差可能很大，因此本文中只使用CSI的振幅进行动作识别。

1. 方法设计

图1所示为方法流程。由信号采集、数据预处理、训练三个阶段组成。其中信号采集阶段将在第3章详细阐述。

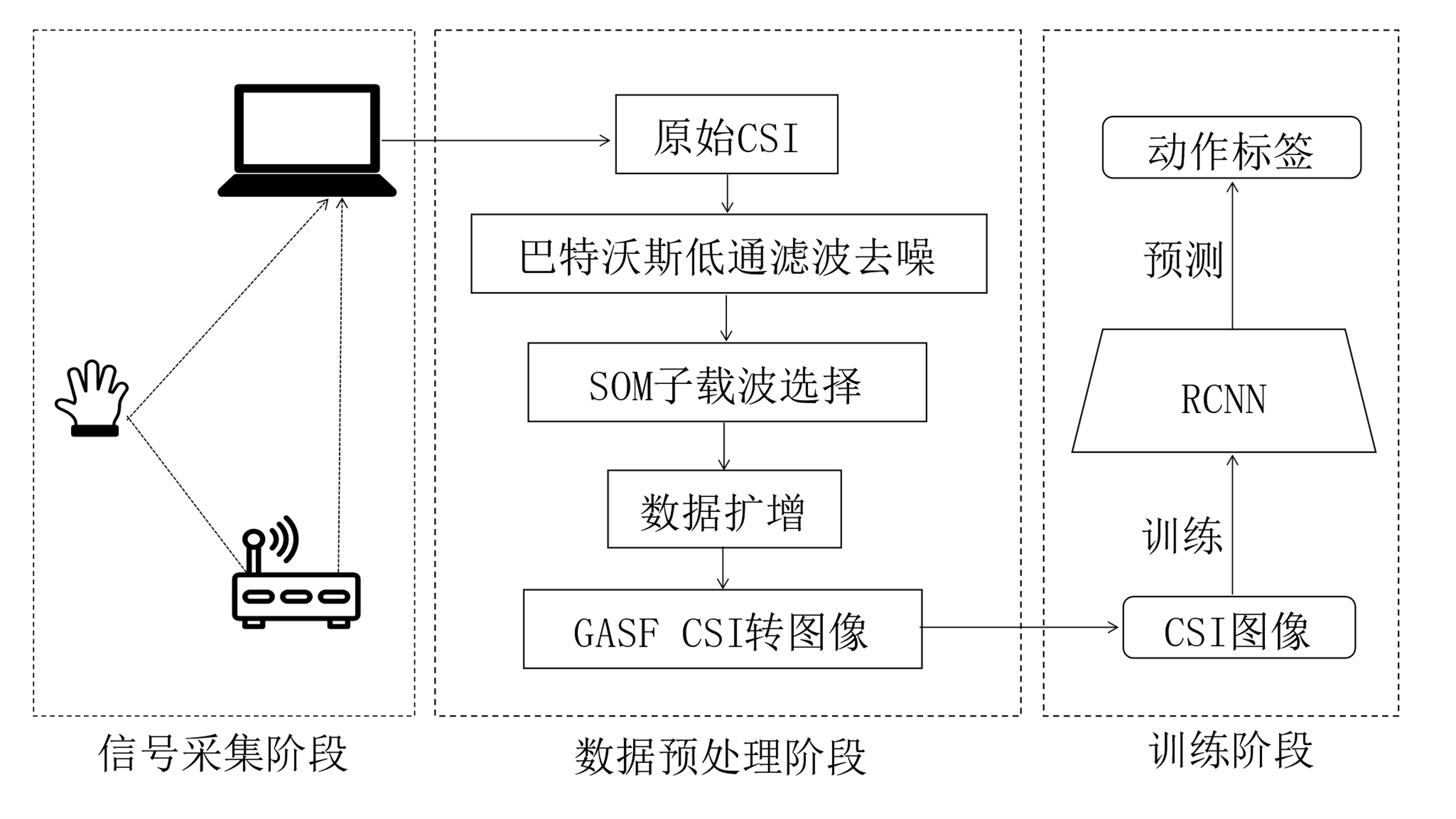


图1　方法流程图

Fig. 1 Method flow chart

* 1. 数据预处理

2.1.1 滤波去噪

由于人体动作的频率较低以及环境中高频噪音存在，所以原始CSI需要滤除高频分量，这里我们使用常用的巴特沃斯低通滤波器进行滤波[1]。巴特沃斯低通滤波器的增益为



其中,表示滤波器的阶数，表示截止频率即功率下降为 -3分贝时的频率，是直流增益。频率低于 的部分会以 的增益通过，而频率高于的就会被抑制。由于手部动作频率范围在0.3 Hz到4.5 Hz之间[13]，因此设置截止频率为5Hz，考虑到滤波效果与计算成本滤波器阶数设置为5。图2所示为滤波前后对比，可以看出滤波后曲线变得平滑高频噪声被抑制。

图2　低通滤波前后CSI幅值对比

Fig. 2 Comparison of CSI amplitude before and after low-pass filtering

2.1.2子载波选择

SOM[6]是一种无监督聚类的神经网络, 通过学习自动对输入进行聚类，且能保持数据的拓扑结构不变。SOM由输入和输出（竞争两层）构成，采用竞争学习的方式进行训练。首先将网络权重初始化，接着每个输入的样本遍历竞争层中的每个节点，计算两者之间的距离（通常为欧式距离），选取距离最小的节点作为获胜节点。

接着用随机梯度下降法更新获胜节点及其邻近节点的权重向量向使它们向输入向量移动。输出节点的权值更新公式为



其中为权重向量，为迭代次数，是当前输入向量，为获胜节点，是下给出和之间距离的邻近函数，用来确定获胜节点对其近邻节点的影响强弱。是一个单调递减的学习率。对每个输入向量这个过程会重复进行直到达到提前设置的循环次数。

本文设置输出层SOM神经元的数量为3进行聚类，图3所示为上下挥手时不同的四个子载波上的CSI聚类后的结果。样本数量为50，可以看到子载波3聚类后只有一个簇，说明数据内在的分布较一致，受环境干扰较小。子载波1、2、4聚类后均生成多个簇，说明同一个子载波上的CSI间差异较大。因此对于上下挥手动作，选择提取子载波3上的CSI作为训练数据。



图3　上下挥手的四个子载波聚类结果

Fig. 3 Clustering results of four subcarriers waving up and down

2.1.3数据扩增

相较传统模型，深度学习的参数非常多，故有很强的拟合能力，从而也就需要更多的数据来避免过拟合的发生。但收集和标记大量的数据是耗时耗力的。通过现有的样本进行数据扩增创建新的样本，能够利用有限的数据有效的提高分类性能[14]。本文采取了多种数据扩增方法包括

1. 添加噪声，CSI序列加上高斯随机噪声序列。
2. 缩放，CSI序列乘以缩放系数。
3. 降采样，设置降采样因子，保留CSI序列中间隔为的采样点，其余采样点舍弃。
4. 滑动平均，采样点的值由滑动窗口内CSI数据的平均值代替。
5. 截取，使用长度为的窗口在CSI序列上截取数据。

图4所示为使用不同方法所产生的新样本。可以看到一个样本能生成多个虚拟样本。且所生成的样本保留了CSI的变化模式。



图4　应用不同数据扩增方法得到的新样本

Fig. 4 New samples obtained by applying different data amplification methods

2.1.4 CSI转换为图像

CNN能够自动高效的从图像中提取特征。因此本文没有手工设计特征，而是使用GASF[9]方法将接收到的CSI编码为不同类型图像，以便使用CNN提取特征。设为滤波后的CSI序列，先通过



将归一化，此时有，。令将重缩放后的时间序列转换成极坐标系，定义矩阵GASF为



于是有



令可以得到



图5所示为6种手势动作的CSI，通过转换得到的GASF图像，可以看到不同的动作生成的图像有不同的纹理特征。



图5　不同手势的CSI转化得到的图像

Fig. 5 CSI images of different gestures

* 1. 训练

本文中的CRNN由CNN、RNN构成，训练时使用CTC作为损失函数。网络结构如图5所示。

2.2.1 卷积神经网络

首先我们使用卷积神经网络[15]来对图像提取特征，卷积网络部分由四个卷积模块组成，每个卷积模块由两个二维卷积层加上一个池化层构成。卷积层的卷积计算公式为



其中是输入张量，是维度为的卷积核，为该卷积层的输出张量，是张量的坐标值，，是卷积核的坐标值。本文所使用卷积核大小为，并进行了零填充以保持卷积后的输出特征图与输入的分辨率相同。

池卷积层的池化操作能够减少模型的计算量和参数个数防止过拟合。本文使用了最大池化，其公式为



是池化后的结果。是特征图张量上位置上的元素。本文使用的池化核大小为。

2.2.2 长短期记忆网络

由于CSI是时间序列数据，所以接下来使用两个长短期记忆（Long Short-Term Memory，LSTM）层对CSI序列时间上的变化进行建模。LSTM[16]是一种拥有三个“门”的特殊网络结构，包括遗忘门、输入门、输出门。输入门选择保留的信息，遗忘门选择被遗忘的信息，输出门决定哪些状态信息需要输出。如公式（11）所示



，，，分别表示输入门、遗忘门、输出门的输出以及输入的单元状态。，分别为sigmoid、tanh激活函数，，分别为上一时刻的输出和当前时刻的输入。当前时刻的单元状态为：



为上一时刻的单元状态，表示元素乘。由当前时刻的输出门输出和单元状态可得当前时刻的输出为



2.2.3 CTC损失函数

由于CSI序列长度不一致，那么训练前每个样本都需要标记出发生动作的片段在整个CSI序列中的位置，从而截取得到长度一致的输入数据，本文使用CTC[12]损失函数不需要对数据进行截取。LSTM提取完特征后由两个全连接层和Softmax层进行分类，分类结果作为CTC损失函数的输入，此外还需要提供输入长度、标签、标签长度。

CTC算法的不同之处在于引入了一个空白（\_）标签表示无动作发生，设给定长度为的输入序列，神经网络对应的输出序列为，其中为神经网络提取到的特征图的长度，为第帧的条件概率分布，为动作标签和空白标签的个数。CTC算法的输出序列路径为的概率为



用表示映射函数，消除重复的标签和空白标签例如



因此输出序列路径为的概率可以表示为经过映射后为的所有的概率之和，即



CTC损失函数定义如下



CTC算法通过最大路径解码所解得的概率最大的输出序列路径为



其中为Ｔ帧输入序列的后验概率最大的CTC输出序列路径。

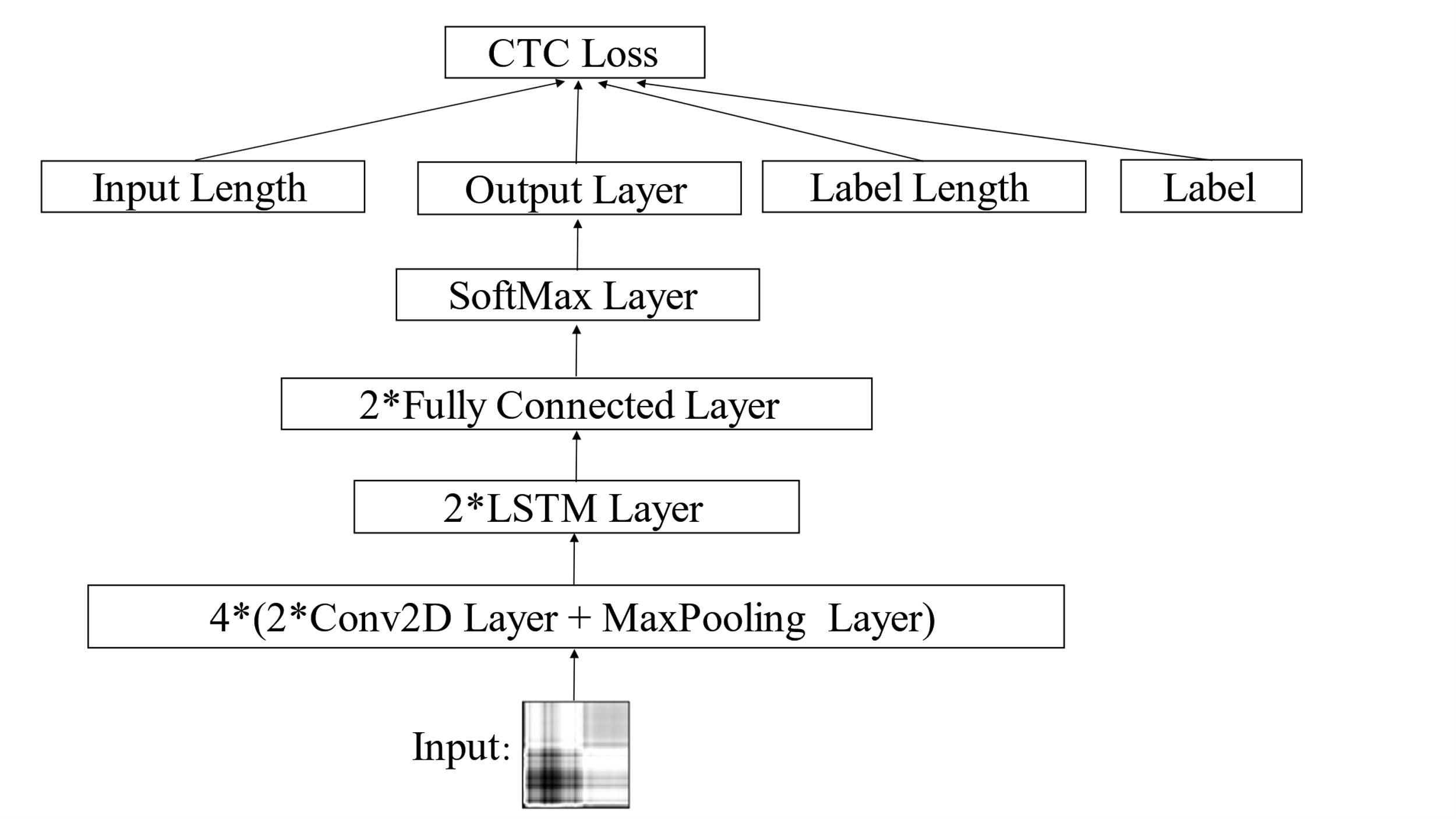


图5 网络结构

Fig. 5 Network structure

1. 实验设计
   1. 数据采集

将一台装有英特尔5300网卡，系统内核版本为4.1.10的HP Pavilion15笔记本电脑作为接收端接收信号。一个支持802.11n 协议的MERCURY MW305R路由器为发射端发送信号。使用Halperin[17]发布的CSI工具包来修改网卡驱动。设置路由器的工作频率为2.4G，发包频率为20HZ。由于发射端有两根天线，接收端有三根天线，每个天线对上有30个子载波。因此接收到的CSI数据为2\*3\*30 的复数矩阵。

图6 所示为实验环境。实验对象站在笔记本和路由器之间的桌子前做如图7所示的6种手势动作，每个手势采集50组数据，共计300组数据。使用240组数据进行训练，60组数据进行测试。

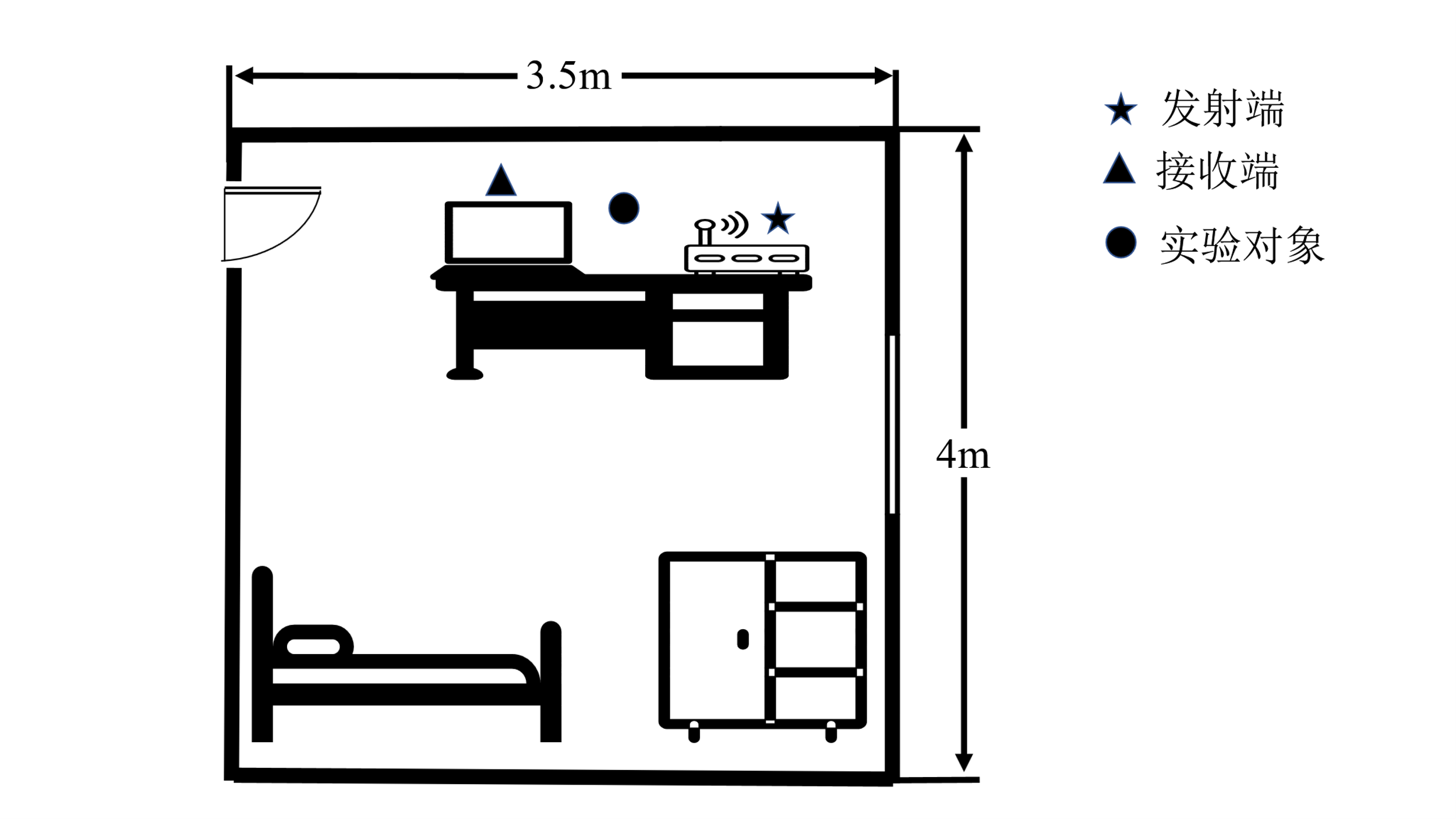


图6 实验环境平面图

Fig. 6 Experimental environment plan

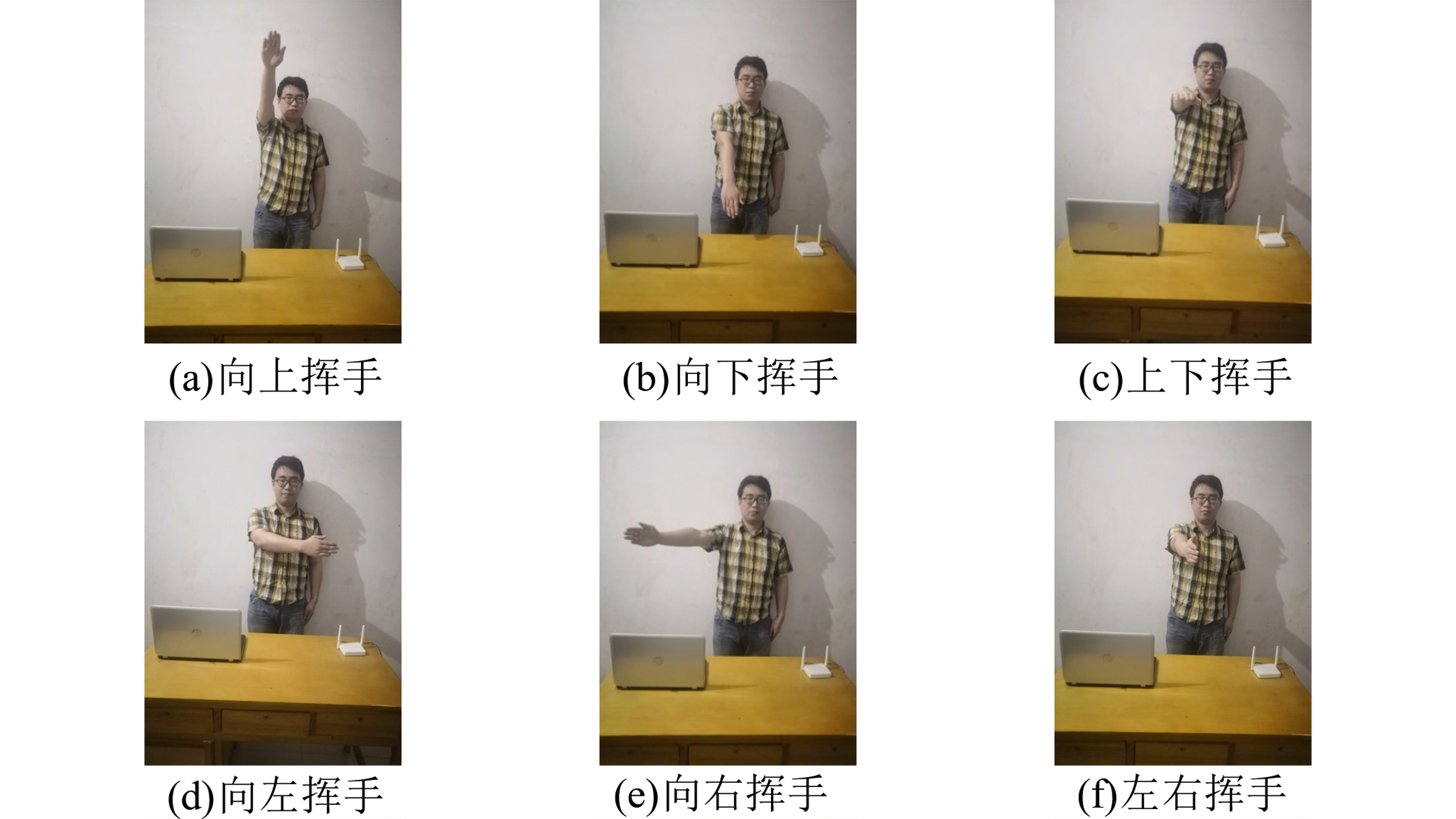


图7 6种手势

Fig. 7 6 Gestures

* 1. 对比分析

本文做了一系列对照实验来对比分析所使用的子载波选择方法和分类算法的有效性。

3.2.1不同的子载波选择方法

为了验证本文提出的，基于SOM聚类的子载波选择方法的有效性。对照实验使用了WiFall[4]的所有子载波求平均、LBR[5]的选择 CSI方差第二大的子载波，两种子载波选择方法作为对比。仅改变子载波选择方法，其他步骤相同，实验结果如表1所示。

表1　不同的子载波选择方法结果对比

Table 1　 Comparison of different subcarrier selection methods

|  |  |
| --- | --- |
| 方法 | 准确率/% |
| 30组子载波上的CSI求平均  选取方差第二大的子载波  基于SOM聚类的子载波选择 | 75  78.3  93.3 |

结果表明直接对所有子载波上的CSI求平均，结果准确率较低，原因可能是不同的子载波对动作和环境的敏感度不同，直接求平均值将各种环境因素的影响叠加到一起，使得CSI中的有效信息被掩盖。选取方差第二大的子载波，准确率虽然有所提高，但是由于CSI易受环境噪音影响，较大的方差可能是由于环境噪音变化引起的，而且不能保证相同动作的CSI变化模式的一致性。通过聚类选择子载波能够保证相同动作的CSI数据内部的一致性，不会出现异常数据，准确率较高。

3.2.2不同的网络结构

将SignFi[18]使用的CNN及文献[19]的所使用LSTM作为对照，并把使用交叉熵损失函数的CRNN也作为一种不同的网络进行对比，结果如表2所示。

表2　不同的网络结构结果对比

Table 2 Compared with the results of the network structure

|  |  |
| --- | --- |
| 方法 | 准确率/% |
| CNN  LSTM  使用交叉熵损失函数的CRNN  使用CTC损失函数的CRNN | 86.7  50  64.9  93.3 |

可以看到仅使用LSTM由于模型参数数量较少因此拟合能力较弱。使用CNN能够达到相对较高的准确率。相较于仅使用CNN或LSTM，同时使用CNN、LSTM能综合利用两种网络结构的优势，准确率较高。使用CTC损失函数不仅不需要对CSI进行截取，而且能够有效的识别发生动作的CSI序列片段，从而提高模型的分类准确率。

1. 结束语

本文提出了一种基于SOM聚类的子载波选择方法，并将CSI数据序列转换为图像，通过CNN、LSTM构成的网络使用CTC损失函数进行训练分类。在数据预处理和模型训练中，采用聚类和深度学习的方法代替传统的基于手工设计特征的方法，并利用现有样本进行了数据扩增。实验结果表明，本文提出的方法识别准确率较高，且无需手工设计特征。下一步将会研究提高模型的泛化能力。

参考文献

[1] 鲁勇, 吕绍和, 王晓东, 等. 基于WiFi信号的人体行为感知技术研究综述[J]. 计算机学报, 2019, 42(02): 3-23.(Lu Yong, Lu Shaohe, Wang Xiaodong, *et al*. A Survey on WiFi Based Human Behavior Analysis Technology [J]. Chinese Journal of Computers, 2019, 42(02): 3-23)

[2] Mantyjarvi J, Himberg J, Seppanen T. Recognizing human motion with multiple acceleration sensors [C]// 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236). Piscataway, NJ: IEEE Press, 2001: 747-752 vol.2.

[3] 胡琼, 秦磊, 黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报, 2013, 36(12): 2512-2524.(Hu Qiong, Qin Lei, Huang Qingming. A Survey on Visual Human Action Recognition [J]. Chinese Journal of Computers, 2013, 36(12): 2512-2524)

[4] Wang Yuxi, Wu Kaishun, Ni L. WiFall: Device-Free Fall Detection by Wireless Networks[J]. IEEE Transactions on Mobile Computing, 2017, 16(2): 581-594.

[5] 姚青桦, 崔然. 一种基于CSI的轻量级行为识别方法[J]. 计算机应用研究, 2018, 35(11): 3397-3399,3404.(Yao Qinghua, Cui ran. Lightweight behavior recognition method based on CSI [J]. Application Research of Computers, 2018, 35(11): 3397-3399,3404.)

[6] Termini S. T. Kohonen,self-organizing maps[J]. Rendiconti del Circolo Matematico di Palermo, 1995, 44(3): 506-506.

[7] Wang Wei, Liu A X, Shahzad M, *et al*. Understanding and modeling of wifi signal based human activity recognition [C]// Proc of the 21st Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2015: 65-76.

[8] Pu Qifan, Jiang Siyu, Gollakota S. Whole-home gesture recognition using wireless signals[J]. Acm Sigcomm Computer Communication Review, 2013, 43(4): 485-486.

[9] Wang Zhiguang, Oates T. Imaging time-series to improve classification and imputation [C]// Proc of the 24th International Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2015: 3939–3945.

[10] Li Hong, Yang Wei, Wang Jianxin, *et al*. WiFinger: talk to your smart devices with finger-grained gesture [C]// Proc of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York: ACM Press, 2016: 250-261.

[11] Wang Fei, Feng Jianwei, Zhao Yinliang, *et al*. Joint Activity Recognition and Indoor Localization With WiFi Fingerprints[J]. IEEE Access, 2019, 7: 80058-80068.

[12] Graves A, Fernández S, Gomez F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks [C]// Proc of the 23rd International Conference on Machine Learning. New York: ACM Press, 2006: 369-376.

[13] Xiong Yingen, Quek F. Hand Motion Gesture Frequency Properties and Multimodal Discourse Analysis[J]. International Journal of Computer Vision, 2006, 69(3): 353-371.

[14] Um T, Pfister F, Pichler D C, *et al*. Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks [C]// Proc of the 19th ACM International Conference on Multimodal Interaction. New York: ACM Press, 2017: 216-220.

[15] Lecun Y, Bottou L, Bengio Y, *et al*. Gradient-Based Learning Applied to Document Recognition[J]. Proc of the IEEE, 1998, 86(11): 2278-2324.

[16] Hochreiter S, Schmidhuber J. Long Short-term Memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[17] Halperin D, Hu W, Sheth A, *et al*. Tool Release: Gathering 802.11n Traces with Channel State Information[J]. Computer Communication Review, 2011, 41(1): 53-53.

[18] Ma Yongsen, Zhou Gang, Wang Shuangquan, *et al*. Signfi: Sign language recognition using wifi[J]. Proc of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 1-21.

[19] 刘佳慧, 王昱洁, 雷艺. 基于LSTM的CSI手势识别方法[J]. 计算机科学, 2019(S2): 283-288.(Liu Jiahui, Wang Yujie, Lei Yi. CSI Gesture Recognition Method Based on LSTM [J]. Computer science, 2019(S2): 283-288)