DOI:10.16136/j.joel.2023.02.0190

融合双层注意力与多流卷积的肌电手势识别记忆网络

刘 聪[1,2,3*],许婷婷[1,2],马钰同[1,2],刘 粤[1,2],孔祥斌[1,2],胡 胜[1,2]

(1.湖北工业大学 电气与电子工程学院,湖北 武汉 430068;2.太阳能高效利用湖北省协同创新中心,湖北 武汉 430068;3.武汉华安科技有限股份有限公司,博士后科研工作站,湖北 武汉 430068)

摘要:针对表面肌电信号(surface electromyography,sEMG)手势识别使用卷积神经网络(convolutional neural network,CNN)提取特征不够充分,且忽略时序信息而导致识别精度不高的问题,本文创新性地提出了一种融合双层注意力与多流卷积神经网络(multi-stream convolutional neural network,MS-CNN)的 sEMG 手势识别记忆网络模型。首先,利用滑动窗口生成的表面肌电图像作为该模型的输入;然后在 MS-CNN 中嵌入通道注意力层(channel attention module,CAM),弱化无关信息,使网络能够更加专注 sEMG 的有效特征;其次,通过长短期记忆网络(long short term memory network,LSTM)对输入的特征进行时序上的激励,关注更多 sEMG 的时序信息,让网络在时间维度上拥有更强的学习能力;最后,采用时序注意力(time-sequence attention,TSA)层对 LSTM 的状态进行关注,从而更好地学习重要肌肉信息,提高手势识别精度。在 NinaPro 数据集上进行实验测试,结果表明,使用本文提出的网络模型在 DB1 数据集和 DB2 数据集的手势识别精度分别达到了 86.42%和 80.60%,高于大多数主流模型,充分验证了模型的有效性。

关键词:表面肌电信号(sEMG);手势识别;多流卷积神经网络(MS-CNN);长短期记忆网络(LSTM);注意力机制

中图分类号:TP391 **文献标识码**:A **文章编号**:1005-0086(2023)02-0180-10

Incorporating two-layer attention and multi-stream convolutional for sEMG gesture recognition memory networks

LIU Cong[1,2,3*], XU Tingting[1,2], MA Yutong[1,2], LIU Yue[1,2], KONG Xiangbin[1,2], HU Sheng[1,2]

(1. School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei 430068, China; 2. Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Wuhan, Hubei 430068, China; 3. Postdoctoral Workstation, Wuhan Hua′an Science and Technology Co., Ltd., Wuhan, Hubei 430068, China)

Abstract: To solve the problems of insufficient feature extraction and ignoring time series information using convolutional neural network (CNN) to extract features of surface electromyography (sEMG) gesture recognition, resulting in low recognition accuracy, this paper innovatively proposed incorporating two-layer attention and multi-stream convolutional neural network (MS-CNN) for sEMG gesture recognition memory networks model. Firstly, sEMG images generated by sliding windows are used as the input of this model; then the channel attention module (CAM) is embedded in an MS-CNN to weaken irrelevant information and enable the network to focus more on the key features of sEMG; secondly, using the long short term memory network (LSTM) for motivating the input features in time-sequence to pay more attention to the time-sequence information of sEMG, which enables network has a stronger learning ability in the time-dimension; finally, focusing on states of LSTM by time-sequence attention (TSA)

* **E-mail**:20181008@hbut.edu.cn

收稿日期:2022-03-23 **修订日期**:2022-04-29

基金项目:国家自然科学基金(61901165)资助项目

layer to learn important muscle information better for improving gesture recognition accuracy. Performing experimental tests on the NinaPro dataset and the results show that the gesture recognition accuracy in the DB1 dataset and DB2 dataset has reached 86.42% and 80.60% using the network model proposed by this paper, which is higher than most mainstream models, which fully verifies the effectiveness of the model.

Key words: surface electromyography (sEMG); gesture recognition; multi-stream convolutional neural network (MS-CNN); long short term memory network (LSTM); attention mechanism

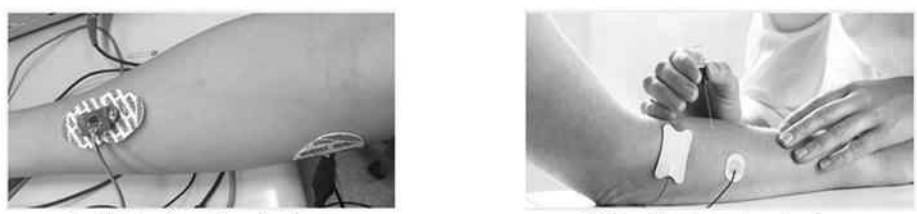
0 引　言

随着科学技术的发展,手势识别在人机交互[1]中被广泛应用,如手语识别[2]、机器人设备控制[3]、虚拟现实游戏[4](virtual reality,VR)和假肢控制[5]等。为了有效地提取各种手势信息,业界提出了诸多方法,其中通过肌电图来捕捉肌肉活动的生物电信号,可更好地获取手势信息。而肌电信号有表面电极(电极贴片)和针式电极两种采集方法。如图 1 所示,表面电极(电极贴片)方式只需在待测区域的皮肤表面放置电极贴片以此来测量肌肉的动作电位,所以采集的信号称为表面肌电信号(surface electromyography,sEMG)。相较于穿透皮肤的侵入式采集方式[6],该方式不会对人体造成伤害。因此,从 sEMG 中获取和识别手势运动意图,如今已成为相关领域研究的热点。

早期基于 sEMG 的手势识别研究主要使用传统的机器学习模型,步骤包括信号检测、信号预处理、特征提取和模式分类 4 个阶段。传统的机器学习模型主要对特征提取和模式分类展开研究,目的是通过特征识别来区分 sEMG 继而传输到分类器进行识别。近年来,相关人员研究了一些时域、频域和时频域相结合的特征提取方法,并用 K 近邻(k-nearest neighbor,KNN)、支持向量机(support vector machines,SVM)、线性判别分析(linear discriminant analysis,LDA)和随机森林(random forests,RF)之类的传统分类器进行手势识别。但传统机器学习模型的特征提取设计及选择过程较为复杂,且特征组合的方法也样式繁多,因此导致人工提取的工作量增加。

如今由于深度学习的兴起,为人类手势运动意图的识别提供了一个崭新的方向。相较于传统的机器学习模型,它不需要耗费大量的人工特征提取时间[7],主要是因为它具有较强的学习能力和特征提取能力。文献[8]首次将卷积神经网络(convolutional neural network,CNN)应用于 sEMG 手势识别中,该方法在 NinaPro 数据库上取得了与传统方法相当的性能,减少了人工提取的工作量,但肌电信号在实际采集中含有心电干扰、电磁干扰和肌肉阻抗等噪声信号,虽已进行去噪处理,依然会有少量噪声混入肌电信号中,对 CNN 提取有效的手势信息造成干扰,使网络特征提取不够充分。文献[9]提出了一种新的瞬时 sEMG 图像的 CNN 模型,将 sEMG 瞬时值直接转换成灰度图像,使识别精度取得了较大的提升,却忽略了信号数据的时序信息。文献[10]利用 Myo 臂带采集原始 sEMG 的特征,并采用循环神经网络(recurrent neural network,RNN)从序列数据中提取特征对手语手势进行分类。文献[11]提出了两阶段的领域自适应(2-stage RNN method,2SRNN)的识别模型,此网络采用快速且轻量级的反向传播训练方法,对 12 种手势达到了 86 %的识别精度。但识别手势数较少,同时由于产生 sEMG 信号的运动单元动作电位沿肌纤维传播,而不同手部动作的肌肉状态不一样,随着时间序列的增加不能有效地捕获重要的肌肉激活特征,从而影响使用网络进行手势识别分类。

针对上述问题,本文提出了融合双层注意力与多流卷积神经网络(multi-stream convolutional neuralnetwork,MS-CNN)的肌电手势识别记忆网



(a) Electrode patch collection　　(b) Needle electrode collection

图 1　两种采集方式

Fig. 1　Two collection methods

络框架。以MS-CNN为基准模型，在此加入通道注意力层(channel attention module，CAM)[12]，使每个单流CNN模型在提取特征时能够学习到更加细腻的深层特征，并滤除无用的特征信息，提升了网络的泛化能力。同时，将加入时序注意力层（time-sequence attention，TSA)[13]的长短期记忆网络(long short term memory networks，LSTM)应用至手势识别方法中，以解决CNN模型忽略时序信息的问题，并更好地学习重要肌肉激活特征。本文提出的框架模型能够将深度特征提取和时间序列回归有效地结合起来，在有效捕获手势特征的同时充分地利用了sEMG的时空相关性，使得手势识别准确率得以进一步提升。

1　数据集与预处理

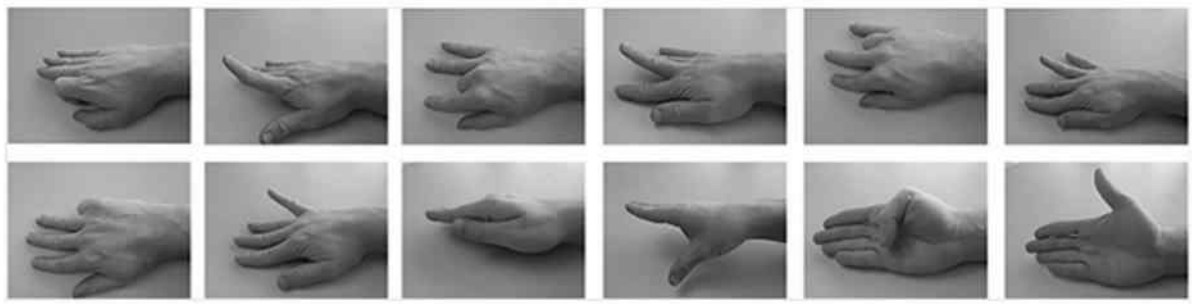
1.1　数据集说明

本文实验选用的是NinaPro表面肌电信号公开数据集内的DB1数据和DB2数据[14]，表1为数据集说明，图2所示为手势动作模型。DB1的手势模型共分为12个基本手指动作、8个等距等张手势、9个手腕基本运动和23个抓握和功能性运动(将日常物体呈现给受试者进行抓取，以模拟日常生活动作)。DB2的手势模型共分为4组：17个手势和手腕运动(由8个等距等张手势和9个手腕基本运动组

表1　数据说明

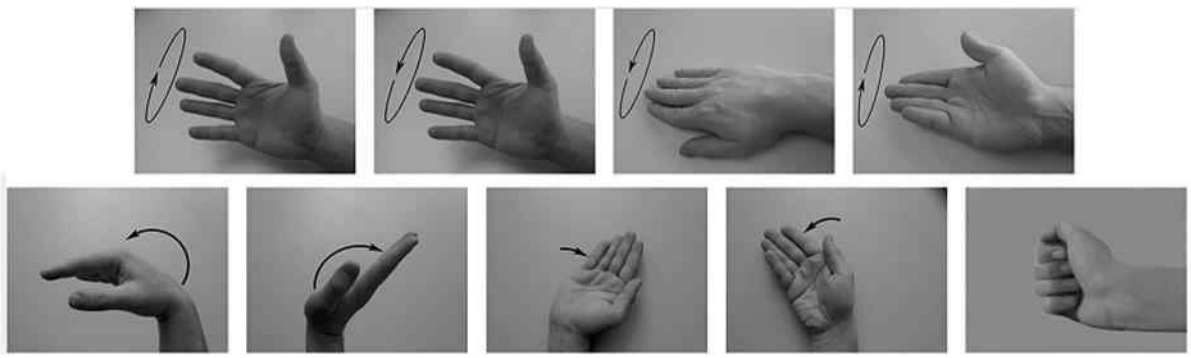
Tab. 1　Datasets information

Parameters	NinaProDB1	NinaProDB2
Number of subjects	27	40
Considered subjects	10	6
Total number of movements	52	50
Avg. years	28±4.6	28±3.1
Sampling frequency (Hz)	100	2000
sEMG electrodes	10 Otto Bock	12 Delsys
Hand kinematics/Dynamics sensors	Cyberglove II	Cyberglove II



(a) 12 basic movements of the fingers

(b) 8 isometric and isotonic hand configurations



(c) 9 basic movements of the wrist

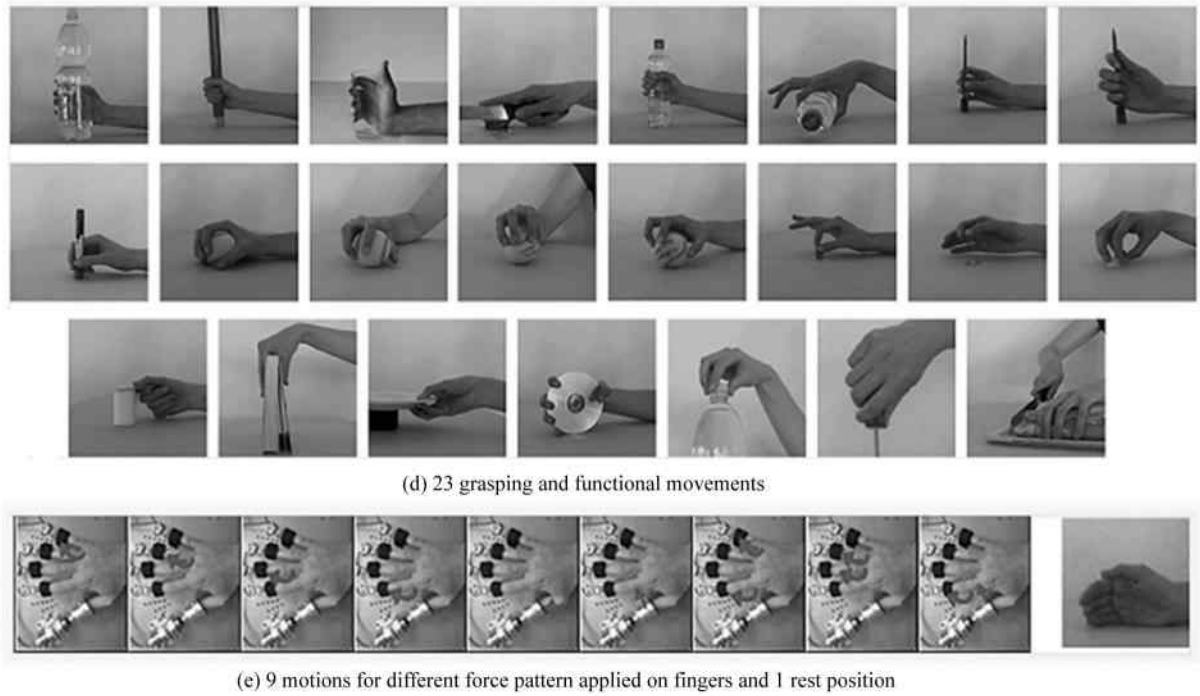


图 2　手势动作

Fig. 2　Movements of fingers

成)、23 个抓握和功能运动、9 个不同力模式手指的运动和 1 个休息动作。

1.2　数据预处理

对 sEMG 进行预处理,可为网络模型提供更多有效的信息,进而提高分类的准确性。sEMG 用于手势分类之前,需要对它进行滤波和归一化处理。滤波就是为了降噪,主要是用于消除污染肌电信号的附加噪声、电路外部或内部因素造成的噪声干扰。本实验的肌电信号通过低通巴特沃斯滤波器进行滤波。之后对过滤后的数据进行归一化处理,使数据量级保持在[−1,1]以内。其中 DB2 数据集需通过巴特沃斯滤波器进行滤波并被下采样至 100 Hz。

本实验数据集最直观的图像表示方法是使用滑动窗口生成 sEMG 图像;总而言之,就是采用滑动窗口对采样信号进行分割。其中信号分割图如图 3 所示。对于实时控制而言,输入延迟是一个需要考虑的重要因素。文献[15]首次提出实时肌电控制系统的控制器延迟应该保持在 300 ms 以下,最近的研究表明实时肌电控制系统的控制器延迟应该保持在 100−250 ms 之间,因此,为了得到合理的样本数,本文只考虑 150 ms 和 200 ms 的时间窗来进行约束,同时选择 50 ms 作为增量窗口。

本文将 L 帧时间窗内 C 通道记录的每个样本记作 X,其中 L 为时间帧数,C 为采集的信号通道数,同时利用滑动窗口将 X 分割成若干子段,每个子段记为 $\{X_1, X_2, \ldots, X_T\}$,输入的肌电图像可以表示为 $X = R^{L\times C}$。图 4 所示为分割的某子段图,S 是滑动窗口宽度,也是图像宽度,C 代表原始 sEMG 的通道数,也代表图像高度或滑动窗口高度。每一流 CNN 网络模型的输入为 $X^j = R^{L\times C}$,其中 j 表示第 j 流网络的输入。对于每个单一的 CNN 流都是根据每个通道记录的 sEMG 图像进行训练的,本文 DB1 数据集使用了 10 个通道,则产生了 10 流。

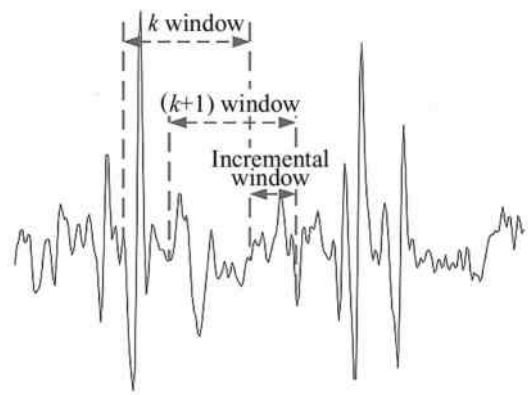


图 3　滑动窗口结构图

Fig. 3　Structure diagram of sliding window

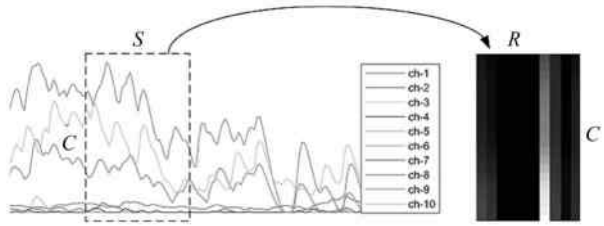


图 4　滑动窗口切割后的图像

Fig. 4　Image produced by sliding window cutting

2　融合双层注意力与 MS-CNN 的 sEMG 手势识别记忆网络

2.1　sEMG 手势识别网络框架

本文将 CAM、TSA 分别与 MS-CNN 和长短期记忆模块相融合，从而构成本文所提出的融合双层注意力与 MS-CNN 的 sEMG 手势识别记忆网络，该网络框架如图 5 所示。

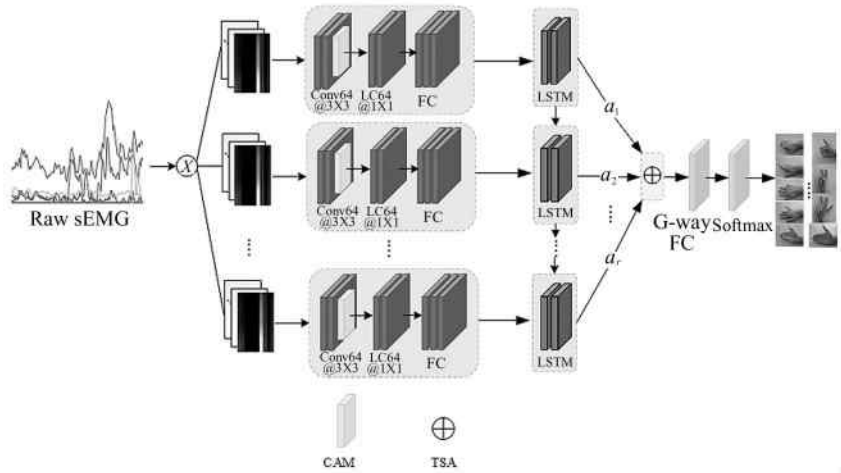


图 5　融合双层注意力与多流卷积神经网络的肌电手势识别记忆网络框图

Fig. 5　Incorporating two-layer attention and MS-CNN for sEMG gesture recognition memory networks

本文的网络模型由 3 部分组成：第 1 部分是利用嵌入 CAM 的 CNN，使网络能对特征进行更充分的提取。CNN 用作特征提取器，将滑动窗口处理 sEMG 生成的$\{X_1, X_2, X_3, \cdots, X_T\}$转换为特征向量$\{F_1, F_2, F_3, \cdots, F_T\}$，并且在 CNN 模型中加入 CAM 能有效地解决特征提取不充分，CAM 将权重值再分配使得 CNN 能够关注有效的特征图，从而学习到更为细化的肌电信号深层特征；第 2 部分是利用 LSTM 提取 sEMG 的时序信息。将 CNN 提取 sEMG 的深度特征向量$\{F_1, F_2, F_3, \cdots, F_T\}$作为 LSTM 的输入序列，参数 T 为特征序列中特征向量的个数，也表示 LSTM 模块的时间步长；第 3 部分是通过 TSA 关注 LSTM 在提取过程中不同手势之间的肌肉信息，并且提高混合 CNN-LSTM 结构的性能。将 LSTM 学习到的特征输送到 TSA 并决定输出权重值。最后连接 G 类手势标签的全连接层，并输送到 Softmax 分类器得到 sEMG 手势分类结果。

2.2　MS-CNN

CNN 是在人工神经网络的基础上开发起来的，它有局部感受野和权值共享的特点。对于单个神经元，只需要局部感知，更高层神经元通过合成局部信息获得更高级的信息。权重共享是指在卷积操作过程中提取的特征参数与位置无关。对于 DB1 数据集，本文利用 10 个单流 CNN 网络组成 MS-CNN 网络模型，能够更高效地提取手势的高级语义信息，提高识别精度。

首先本文利用 MS-CNN 网络，以此对 sEMG 的多个 sEMG 图像进行并行建模。每个单流 CNN 模型共有 7 层，前两层为卷积层，其中每层由 64 个 3×3 的卷积滤波器组成，并在第 2 层后面加入了 CAM。接下来是两个局部连接层，每层由 64 个 1×1 局部连接层来提取 sEMG 图像的局部特征。最后 3 层分别是由 512、512 和 128 个单元组成的全连接层，其中前两层具有 dropout，以此来减少过拟合。每层后都加入了批量归一化（batch-normalization layer，BN）和修正线性单元（rectification linear unit，ReLU），以此来减少内部协变量偏移，从而加速网络收敛，防止梯度消失。MS-CNN 同时也选择自适应矩估计（adaptive moment estimation，Adam）作为网络优化器。最后提取的特征将用作后续时序模型的输入。

2.3　LSTM

RNN 具有递归反馈机制,使其非常适合处理具有序列特点的数据,但 RNN 又存在梯度消失或梯度爆发等一些缺陷,因此,本文使用了基于 RNN 改进的变体网络—LSTM,使本文的网络模型能够稳定且有效地学习长期依赖信息[16],充分处理网络中 sEMG 的时序问题。本文在 MS-CNN 网络的第 7 层后添加了两个堆叠的 LSTM 层,每个 LSTM 有 128 个单元,同时设置 dropout 的值为 0.5,通过 dropout 方法来抑制过拟合。

LSTM 属于门控 RNN 的范畴,它使用门(Sigmoid 激活函数,然后逐点相乘)来创建每一时刻信息通过的路径,使得输出时就能保证信息不会消失或爆炸。每个 LSTM 的内部结构分别由输入门、输出门、遗忘门和单元状态组成,其模型如图 6 所示。

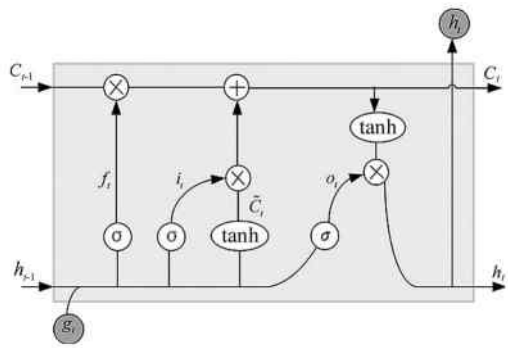


图 6　LSTM 的内部结构

Fig. 6　Inner structure of LSTM unit

它们之间的计算关系如下:

$$f_t = \sigma(W_f[h_{t-1}, g_t] + b_f), \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, g_t] + b_i), \tag{2}$$

$$o_t = \sigma(W_o[h_{t-1}, g_t] + b_o), \tag{2}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, g_t] + b_c), \tag{4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{6}$$

式中,σ 是 Sigmoid 激活函数,i_t、f_t、o_t、C_t 分别是输入门、遗忘门、输出门和单元状态,W_i、W_f、W_c、W_o 是每个门和层的权重向量,b_i、b_f、b_C、b_o 是对应的偏置向量,g_t 是当前节点的输入,h_t 和 C_t 是 LSTM 网络的输出。

2.4　CAM 与 TSA

注意力机制就是一种聚焦于局部信息的机制。通过加入注意力机制能够使其关注所需要的重要特征,并忽略无关特征。注意力机制通常对数据内部进行权值再分配,目前已成功应用于图像识别[17]、语音识别[18]、情感分析[19]等重要领域,现已逐渐应用到肌电信号手势识别中[20]。

本文在 MS-CNN 网络模块中加入了 CAM,以解决 CNN 模型不能专注于关键特征使得提取特征不充分的问题,因此需要通过 CAM 给重要特征图的权值更多,并且抑制非重要特征图的干扰。同时在 LSTM 网络模块中加入了 TSA。当 LSTM 层随着输入时序信息长度的增加,提取的深层特征数量也会增加,由于不能有效地获得重要的肌肉信息,因此本文需要通过引入一个注意力层来学习输入的时间序列信号,给 sEMG 手势识别带来性能上的提高。此注意力层能够为每个末尾状态分配一个权重,然后将它们融合并输出一个新的特征。最后,将注意力机制生成的新特征送到全连接层进行最终推理。

CAM 图如图 7 所示。其具体流程就是输入特征图 F,然后对其空间维度进行压缩,并使用最大池化和平均池化操作处理,从而得到两个一维的矢量。然后将上述两个特征向量发送到共享网络,并产生 CAM 图 $M_c \in R^{l\times c}$。具有隐藏层的多层感知器构成了该 CAM 的共享网络。再将共享网络输出的两个特征继续使用最大池化和平均池化操作后进行融合,最后将融合后的特征通过 Sigmoid 激活函数得到权重系数 M_c。即 CAM 得到的权重系数 M_c 如下式所示:

$$M_c(F) = \sigma(MLP(AvgPool(f)) + MLP(MaxPool(f))) = \sigma(W_1(W_0(F^c_{\text{avg}})) + W_1(W_0(F^c_{\max}))), \tag{7}$$

式中,F^c_{avg} 和 $F^c_{\max}$ 分别表示平均池化向量和最大池化向量,W_0 和 W_1 代表的是多层感知机模型中的两层参数,W_0 和 W_1 之间的特征需要使用 ReLU 作为激活函数去处理。

同时本文还使用了 TSA,注意力层将 LSTM 网络输出的 h_t 输入到一个单层的 MLP 中得到的 u_t,其中 u_t 也可作为 h_t 的隐含表示。权重向量 w_t 是随机初始化得到的,权重向量 w_t 和 u_t 经过 Softmax 函数得到注意力权重 α_t,向量 r_t 则是时间序列加权求和。其注意力层的公式表示如下:

$$u_t = \tanh(W_\omega h_t + b_\omega), \tag{8}$$

$$\alpha_t = \text{Softmax}(w_t u_t), \tag{9}$$

$$r_t = \sum_{t=1}^{T} \alpha_t h_t, \tag{10}$$

式中,h_t 是 LSTM 模块的第 t 个隐藏单元的输出,W_ω 和 w_t 是注意力层加权权重,α_t 是第 t 个注意力权

重，r_t 是注意力模块的输出。

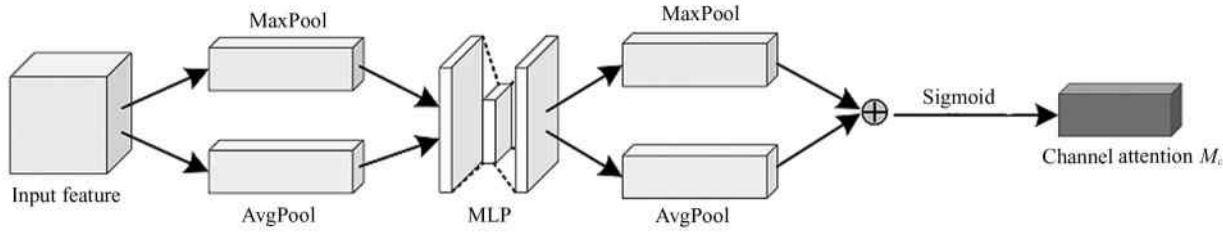


图 7　通道注意力层

Fig. 7　Channel attention module

3　实验与分析

3.1　实验设置

本实验主要对 NinaPro 数据库进行了评估。实验平台以 Windows 操作系统，基于深度学习 Tensorflow 框架实现的。其中硬件配置为：Inter Xeon (R) Gold 5218 处理器、Nvidia Titan X 显卡和 32 GB 显存。将处理后的 sEMG 数据集划分为训练集和验证集，为了避免样本重复，按照以下规则划分：其训练集样本占所有的 80%，验证集占所有的 20%。本文的训练批次大小(batch-size)为 128，训练 200 个 epoch，并将学习率设置为 0.001。根据验证集的识别结果，计算数据库的分类精度，设平均分类准确率为 Acc，正确分类手势数量为 N，手势样本总数为 M，如下式：

$$Acc = \frac{N}{M} \cdot 100\%, \tag{11}$$

3.2　实验结果分析

本文将与目前最新的基于 sEMG 的深度学习手势识别方法进行性能比较，以验证本文提出的网络框架的优势。结果如表 2 所示。

文献[14]是开发 NinaPro 数据库的作者，作者使用传统机器学习的方法进行手势识别，利用均方根、时域特征、直方图和边缘离散小波变换进行特征信号的提取，使用 RF 分类器进行手势动作识别效果最佳，DB1 和 DB2 的分类精度分别达到了 75.30% 和 75.27%；之后又初次将传统的 CNN[8] 应用于 sEMG 中，其中包含 3 个卷积层和两个池化层，最终分别达到了(66.60±6.40)%和(66.28%±7.70)%的分类精度，由于是首次使用 CNN 网络，性能并没有传统学习方法的好。GengNet[9] 则是采用瞬时 sEMG 图像作为输入，使用 CNN 模型后采用多数投票的方法，DB1 数据集的分类准确率达到了 77.80%。ChengNet[21] 提出基于 sEMG 特征图像的 CNN 模型，对 52 个手势动作进行识别，平均准确率达到了 82.54%。WeiNet[22] 提出了 MS-CNN 框架，对每个流的 CNN 进行独立训练之后进行融合，52 种手势分类结果达到了 85.00%的识别准确率。ZhaiNet[23] 则是使用短延迟降维 sEMG 频谱图作为 CNN 网络的输入，最终 DB2 数据集的 50 种手势识别准确率达到了 78.71%。

本文所提出的融合双层注意力与 MS-CNN 的 sEMG 手势识别记忆网络模型和其他模型对比可发现，对于 DB1 数据集，当窗口大小为 150 ms 时，相较于 AtzoriNet 模型提升了 18.66%，而相较于 WeiNet 网络仅提升了 0.86%；当窗口大小选择 200 ms 时，相较于 Atzori-RF，此类传统机器学习方法提升了 11.12%。对于 DB2 数据集，当窗口大小为 150 ms 时，相较于 AtzoriNet 模型提升了 13.13%；当窗口大小选择 200 ms 时，相较于 ZhaiNet 方法提升了 1.89%。本模型获得的手势识别准确率都高于其他的手势识别方法，因此证明本网络模型性能更好。

本模型之所以能取得如此高的识别率，得益于本模型直接通过融合 CAM 的 MS-CNN 网络进行深度特性的提取，CAM 能有效地识别并丢弃原始数据中的无用信号，弥补了深度特征挖掘不够的问题；相较于传统的 CNN 网络，本模型又加入了 LSTM 和 TSA 层，它能够解决 CNN 网络对时序信息问题的忽略；LSTM 能够对提取的深度特征进行时序激励，加入的注意力层通过权重的分配能够解决模型时间序列里的重要特征的差异性问题，因此使得本网络模型的性能优于上述所提到的主流手势识别模型。

为了验证此模型对 sEMG 信号手势识别的性能，分别采用了融合双层注意力与 MS-CNN 的 sEMG 手势识别记忆网络与经典分类器进行手势分类对比实验，选取 NinaProDB1 数据集中任意 2 种手

势、4 种手势、8 种手势和 16 种手势进行实验测试。此处的经典分类器包括 KNN、SVM 和 RF,特征提取方法采用文献[16]中的方法。手势分类精度对比结果如图 8 所示,可以看出,本文提出的网络模型均高于其他分类器的手势分类精度,特别当手势数增加时,本模型的优势便凸显出来。此实验也证明了将两种注意力层分别作用于 MS-CNN 与 LSTM 网络进行手势分类的有效性。

表 2　本文模型与其他方法的分类精度对比

Tab. 2　Comparison of classification accuracy of our model and other methods

Methods	Dataset	Number of gestures	The length of voting window/ms	
			150	200
Atzori-RF[14]	NinaProDB1	50	—	75.30%
AtzoriNet[8]	NinaProDB1	50	66.60%±6.40%	—
GengNet[9]	NinaProDB1	52	—	77.80%
ChengNet[21]	NinaProDB1	52	—	82.54%
WeiNet[22]	NinaProDB1	52	84.40%	85.00%
Ours	NinaProDB1	**52**	**85.26%**	**86.42%**
Atzori-RF[14]	NinaProDB2	50	—	75.27%
AtzoriNet[8]	NinaProDB2	50	66.28%±7.70%	—
ZhaiNet[23]	NinaProDB2	50	—	78.71%
Ours	NinaProDB2	**50**	**79.41%**	**80.60%**

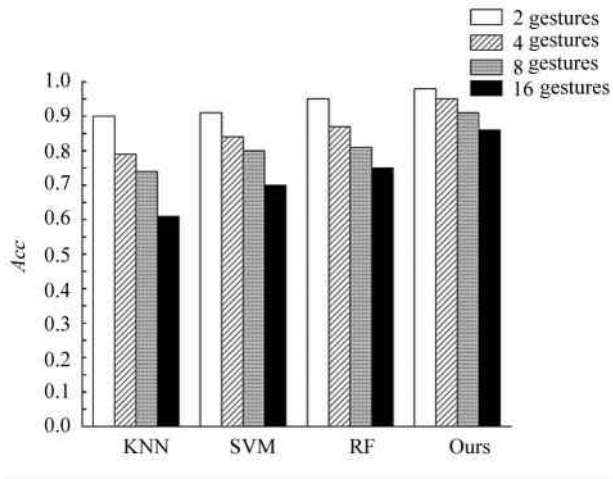


图 8　本文模型与经典分类器的手势分类精度对比

Fig. 8　Comparison of gesture classification accuracy between ours model and classic classifiers

3.3　消融实验

为了进一步验证本文所提出的网络模型的各个模块的有效性,使用 MS-CNN 作为基准模型(Baseline),选取 NinaProDB1 作为本实验的数据集,原始肌电信号以滑动窗口大小为 200 ms 进行分割,再使用 CAM、LSTM 和 TSA 3 个模块在数据集上进行消融实验。手势识别准确率的消融实验结果如表 3 所示。本实验在 Baseline 上加入 CAM 后使模型的准确率由 82.32% 提升至 83.69%。之后再加入 LSTM 网络,相较于只有 Baseline+CAM 模型提升了 2.08%。最后在上述基础上加入 TSA 模块,使网络的识别性能达到了 86.42%,相较于 Baseline+CAM+LSTM 模型提高了 0.65%。通过实验可以发现,在 Baseline 上依次加入 CAM、LSTM 和 TSA 层后,其评价指标 *Acc* 皆有较为明显的提升。由此可说明,在 MS-CNN 中加入本文所使用的 3 个模块可有效提升识别分类准确率。

表 3　网络模型的消融实验

Tab. 3　Ablation experiment of network model

	Baseline	CAM	LSTM	TSA	*Acc*/%
1	√	—	—	—	82.32
2	√	√	—	—	83.69
3	√	√	√	—	85.77
4	√	√	√	√	86.42

图 9 为本模型对 52 种手势识别的混淆矩阵,其中每一行为本模型的预测手势类别,每一列为本模型的真实手势类别,斜对角为各类手势的正确识别结果,颜色越深,则说明识别精确度越高,模型的效果越好。从混淆矩阵图可以看出,大部分手势动作的识别精确度都比较高,集中在斜对角线上,但是也存在部分手势动作识别准确率不高的情况,比如说手势 1 和手势 2,手势 6、手势 8 和手势 20,手势 9、手势 10、手势 11 和手势 12,手势 37 和手势 38 等存在的混淆较为明显,其中有部分手势被识别为其他的手势动作而造成识别准确率较低。经过分析之后发

现，是由于手指的发力点或手指运动方向较为相似，因此部分手势识别会受到相似手势的影响，进而造成手势识别效果不佳。

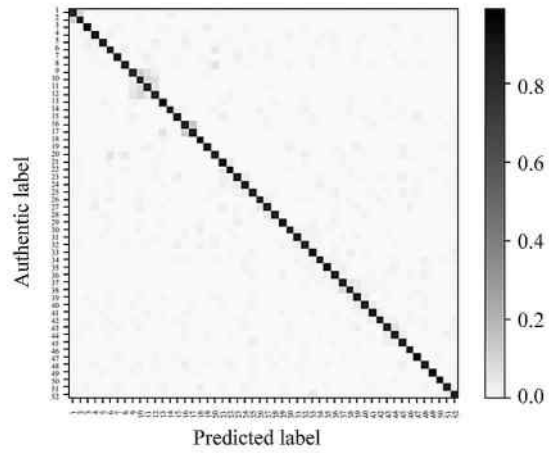


图 9　52 种手势识别混淆矩阵

Fig. 9　Confusion matric of 52 kinds of gesture

4　结　论

本文提出了一种融合双层注意力与 MS-CNN 的 sEMG 手势识别记忆网络模型，该模型通过 MS-CNN 快速有效地对肌电图进行特征提取，解决人工提取特征繁琐的问题；并在 MS-CNN 中融入 CAM，使网络能够更充分地捕获手势的关键语义信息；同时使用融入 TSM 的 LSTM，来利用 sEMG 的时空信息，保障 sEMG 的时变性，并且更好地获得关键肌肉激活特征，让本网络模型的手势识别性能得以较大提升。各项实验结果表明，该网络模型的手势识别精度高于大多主流模型，充分验证了所提模型的有效性。在今后的研究中，需重点关注相似手势的研究，以此对手势进行更好的区分和识别。

参考文献：

[1] DING Q C，XIONG A B，ZHAO X G，et al. Review of research and application of motion intention recognition method based on surface electromyography [J]. Journal of Automation，2016，42(1)：13-25.
丁其川，熊安斌，赵新刚，等. 基于表面肌电的运动意图识别方法研究及应用综述[J]. 自动化学报，2016，42(1)：13-25.

[2] MING J，OMAR Z，JAWARD M H. A review of hand gesture and sign language recognition techniques[J]. International Journal of Machine Learning and Cybernetics，2019，10(1-3)：1-23.

[3] JACQUES R. Tuning control parameters of vibration reduction and motion control systems for fabrication equipment and robotic systems[J]. Acoustical Society of America Journal，2007，121(6)：3263.

[4] GORZKOWSKI S，SARWAS G. Exploitation of EMG signals for video game control[C]//2019 20th International Carpathian Control Conference (ICCC)，May 26-29，2019，Krakow-Wieliczka，Poland. New York：IEEE，2019：1-6.

[5] PARAJULI N，SREENIVASAN N，BIFULCO P，et al. Real-time EMG based pattern recognition control for hand prostheses：a review on existing methods，challenges and future implementation[J]. Sensors，2019，19(20)：4596.

[6] ZHANG X. Body gesture recognition and interaction based on surface electromyogram[D]. Hefei：University of Science and Technology of China，2010.
张旭. 基于表面肌电信号的人体动作识别与交互[D]. 合肥：中国科学技术大学，2010.

[7] XU L K，ZHANG K Q，XU Z H，et al. Convolutional neural network human gesture recognition algorithm based on surface EMG signal energy Kernel phase diagram [J]. Journal of Biomedical Engineering，2021，38(4)：621-629.
许留凯，张克勤，徐兆红，等. 基于表面肌电信号能量核相图的卷积神经网络人体手势识别算法[J]. 生物医学工程学杂志，2021，38(4)：621-629.

[8] ATZORI M，COGNOLATO M，MüLLER H. Deep learning with convolutional neural networks applied to electromyography data：A resource for the classification of movements for prosthetic hands[J]. Frontiers in Neurorobotics，2016，10：9.

[9] GENG W D，DU Y，JIN W G，et al. Gesture recognition by instantaneous surface EMG images[J]. Scientific Reports，2016，6(1)：1-8.

[10] AMOR A，GHOUL O E，JEMNI M. Toward sign language handshapes recognition using Myo armband[C]//International Conference on Information & Communication Technology & Accessibility，December 19-21，2017，Muscat，Oman. New York：IEEE，2017：1-6.

[11] KETYKó I，KOVáCS F，VARGA K Z. Domain adaptation for sEMG-based gesture recognition with recurrent neural networks [C]//International Joint Conference on Neural Networks (IJCNN)，July 14-19，2019，Budapet，Hungary. New York：IEEE，2019：1-7.

[12] WOO S，PARK J，LEE J Y，et al. CBAM：Convolutional block attention module[C]//European Conference on Computer Vision (ECCV)，September 8-14，2018，Munich，Germany. Cham：Springer，2018：3-19.

[13] YANG Z，YANG D，DYER C，et al. Hierarchical attention networks for document classification[C]//2016 Confer-

ence of the North American Chapter of the Association for Computational Linguistics:Human Language Technologies,June 12-17,2016,San Diego,California. Stroudsburg, PA: Association for Computational Linguistics, 2016:1480-1489.

[14] ATZORI M,GIJSBERTS A,CASTELLINI C,et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses[J]. Scientific Data,2014,1(1):1-13.

[15] HUDGINS B,PARKER P,SCOTT R N. A new strategy for multifunction myoelectric control[J]. IEEE Transactions on Biomedical Engineering, 1993,40(1):82-94.

[16] GRAVES A,JAITLY N,MOHAMED A. Hybrid speech recognition with deep bidirectional LSTM[C]//2013 IEEE Workshop on Automatic Speech Recognition and Understanding,December 08-12,2013,Olomouc,Czech Republic. New York:IEEE,2013:273-278.

[17] XU K,BA J,KIROS R,et al. Show,attend and tell: Neural image caption generation with visual attention[EB/OL]. (2016-04-19)[2022-03-23]. https://arxiv. org/abs/1502.03044v3.

[18] CHOROWSKI J,BAHDANAU D,SERDYUK D,et al. Attention-based models for speech recognition[EB/OL]. (2015-06-24)[2022-03-23]. https://arxiv. org/abs/1506.07503.

[19] BAZIOTIS C,PELEKIS N,DOULKERIDIS C. Datastories at semeval-2017 task 4:Deep LSTM with attention for message-level and topic-based sentiment analysis[C]//11th International Workshop on Semantic Evaluation (SemEval-2017), August 3-4, 2017, Vancouver, Canada. Strouclsburg,PA:Association for Computational Linguistics,2017:747-754.

[20] HU Y,WONG Y,WEI W,et al. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition[J]. PloS One,2018,13(10): e0206049.

[21] CHENG Y,LI G,YU M,et al. Gesture recognition based on surface electromyography-feature image[J]. Concurrency and Computation: Practice and Experience,2021,33(6):e6051.

[22] WEI W,WONG Y,DU Y,et al. A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface[J]. Pattern Recognition Letters,2019,119:131-138.

[23] ZHAI X,JELFS B,CHAN R H,et al. Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network[J]. Frontiers in Neuroscience,2017,11:379.

作者简介:

刘　聪　(1982—),男,博士,副教授,硕士生导师,主要从事数字图像处理和模式识别方面的研究.