[文章编号] 1003－4684(2025)02-0029-07

基于 SE 与 XceptionTime 肌电信号手势识别

刘 聪[1,2]，许婷婷[1]，胡 胜[1]

(1 湖北工业大学电气与电子工程学院，湖北 武汉 430068；
2 武汉华安科技股份有限公司博士后科研工作站，湖北 武汉 430068)

[摘 要] 传统基于卷积神经网络的表面肌电信号手势识别算法普遍识别精度不高，究其原因：一是常规卷积对输入通道做卷积并求和，进而造成 sEMG 信号的通道特征和时序特征混杂一起；二是这些算法对 sEMG 信号中不同通道间的特征提取不够充分。针对这一问题，提出了一种基于 SE 注意力与 XceptionTime 的肌电信号手势识别模型。首先，对原始的表面肌电信号分别进行滤波、归一化和滑动窗口的预处理操作；然后本网络中的 SE-XceptionTime 模块通过使用深度可分离卷积可将通道特征与时序特征分开映射，并设置 3 个不同卷积核大小滤波器提取长或短的时序信息；最后在该卷积后加入 SE 注意力机制，以此来关注通道中的有效特征，弱化无关信息的干扰。实验结果表明，在 NinaproDB1 数据集上 52 类手势识别准确度达到了 88.72%，均高于目前主流的深度学习手势识别网络模型，充分验证了网络模型的有效性。

[关键词] 表面肌电信号；手势识别；深度可分离卷积；注意力机制

[中图分类号] TP391 [文献标识码] A

手势作为人机交互的一种重要手段，在医疗康复[1]、假肢控制[2]、手语翻译[3]等领域有着广泛的应用。在手势识别领域中，表面肌电信号(surface electromyography，sEMG)被广泛的应用于肌肉－计算机界面(muscle-computer interface)中，其中 sEMG 能够很容易地被非侵入性电极采集到，而采集到的信息包含大量的手势运动意图[4]。因此使用表面肌电信号进行手势的识别和分类对于假肢动作控制等领域具有重要意义。

目前基于表面肌电信号的手势识别研究主要分为传统的机器学习和深度学习两类方法[5]。当采用传统的机器学习进行手势识别分类时，需使用时域法、频域法和时频域进行特征提取，结合分类器分类才可获得较优的手势分类性能，但特征提取的缺点很明显，需要依赖大量人工的经验和努力才可选择到合适的特征，同时也会遗漏信号中的重要信息。而深度学习的发展克服了上述问题，它从原始表面肌电信号的数据中自动学习复杂的特征，并在训练过程中考虑输入的所有信息。AtzoriNet[6]是首次提出使用 CNN 网络用于表面肌电信号的手势识别分类，精度达到 66.6%。Ketykó[7]使用权重初始化方法处理 sEMG 信号后输入到 RNN 中，12 种手势的分类精度达到 86%。GengNet[8]则是在表面肌电信号图的基础上构建用于手势识别的 CNN 网络，通过在一个时间窗口内对多个结果进行多数投票，对 52 个手势的识别准确率达到 77.8%。HuNet[9]提出基于注意力机制的 CNN-LSTM 混合网络，可捕获 sEMG 通道和时间特征，从而使 52 种手势的识别精度达到 84.8%；但常规卷积操作是对输入各个通道做卷积，然后对各个通道的卷积结果进行求和，这种操作会使卷积学习到的时序特征和通道特征混合在一起。WeiNet[10]直接输入原始 sEMG 信号，通过 MS-CNN 模型中的每个流独立地提取手势信号的代表特征，之后将提取的所有特征进行融合，与全连接层和 softmax 层相连并识别手势，最后达到 85%手势识别精度。FanNet[11]使用 5 种时域法对原始 sEMG 进行特征提取，将 5 个单独学习到的特征向量进行融合，并结合 BiLSTM 网络进行手势识别分类。YuNet[12]提出一种最佳肌电特征集的卷积神经网络模型，当使用多肌电特征图时手势识别效果最佳；该模型却忽略了不同通道间的重要性，导致网络易受无关信息的干扰。

针对上述问题，本文在 XceptionTime 网络[13]的基础上，提出了一种基于 SE 注意力与 Xception-

[收稿日期] 2023－02－08
[基金项目] 湖北省重点研发计划(2021BGD013)；武汉市科技局知识创新专项(2022010801010255)
[第一作者] 刘 聪(1982－)，男，湖北荆州人，湖北工业大学副教授，研究方向为图像处理与深度学习。
[通信作者] 许婷婷(1997－)，女，湖北宜昌人，湖北工业大学硕士研究生，研究方向为图像处理与模式识别。

Time 的肌电信号手势识别网络模型。本方法使用 XceptionTime 模块将通道特征和时序特征进行抽离，然后分开映射，并使用三组深度可分离卷积层去提取短或长时间序列的特征；同时 SE 注意力模块能够捕捉每个通道手势信息的特征，以此降低无关特征干扰；最后在模型中加入 APTx 来代替 ReLU 激活函数，解决了梯度消失和稳定增益的问题，从而使模块能够稳定地提取手势特征。

1 SE 与 XceptionTime 手势识别算法

1.1 手势识别网络框架

本文对 XceptionTime 网络进行了改进，具体结构如图 1 所示。整体的网络框架包括 4 个 SE-XceptionTime 模块，其中滤波器(filters)的参数设置为 16、32、64 和 128。以两个 SE-XceptionTime 模块为一组，分别使用了两个残差块，以此解决网络中存在的退化问题，让模型能够更加有效地提取手势信息。并在残差块中的 1 ×1 卷积层后加入批量归一化，以防止过拟合并加快网络模型的收敛速度。其次将 SE-XceptionTime 模块处理后的 sEMG 手势特征向量与残差网络输出进行合并，再通过 APTx 激活函数来保证网络更充分捕获手势信息，同时增加网络的泛化能力。在经过两轮处理并合并后，将 SE-XceptionTime 模块和与残差连接的总和输出输入到自适应平均池化，降低特征维数同时也避免了数据冗余。之后通过三个 1×1 卷积层，每个卷积层之后均需要进行批量归一化和 APTx 激活函数，得到特征通道数 52 也就是类的数量，再使用自适应平均池化。最后将该自适应平均池化的输出特征输入至 Softmax 分类器，使用 Softmax 分类器进行识别分类，得到 sEMG 的手势分类结果。

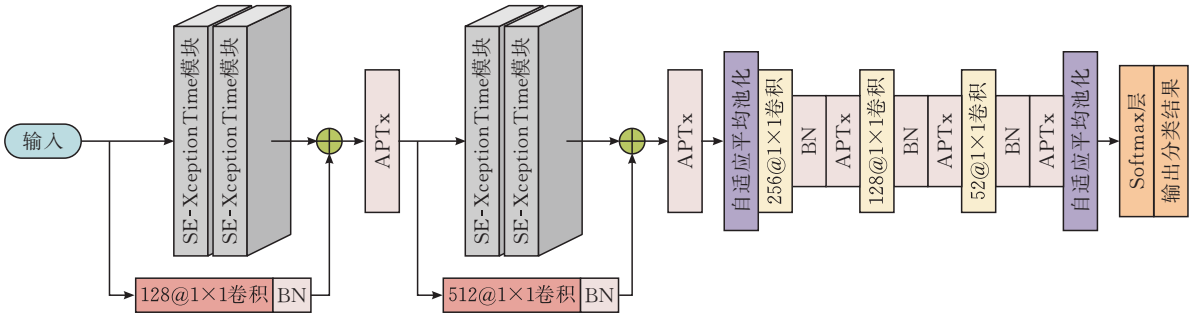


图 1 基于 SE 注意力与 XceptionTime 的肌电信号手势识别网络框架

1.2 SE-XceptionTime 模型

目前，在基于卷积神经网络的 sEMG 手势识别方法研究中，大多数模型是使用通道相关性和时间相关性进行联合映射，并依靠网络模型的深度或分类器来提升识别性能，却忽略了通道特征在手势识别网络中的重要性。为了解决此类问题，本文引入了 XceptionTime 模块，将通道相关性和时间相关性分开映射；并将 SE 注意力模块添加到深度可分离卷积后，从而构成本文所提出的 SE-XceptionTime 模型，具体结构如图 2 所示，以此来捕捉每个通道手势信息的特征，提升手势分类的准确度。

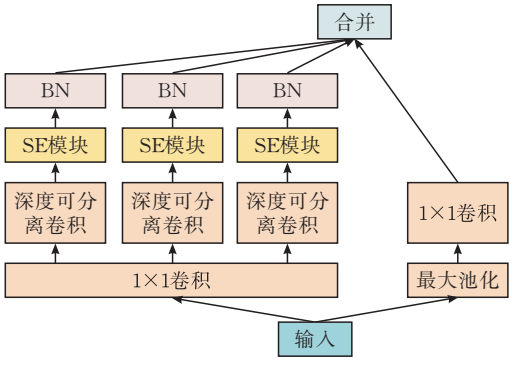


图 2 SE-XceptionTime 模型

1.2.1 XceptionTime 模型

Xception[14] 是 Google 公司在 Inception-v3[15] 基础上改进的模型，其将空间相关性与通道间的相关性分离后再进行处理，于是通过深度可分离卷积代替 Inception-v3 中的原始卷积。如图 3 所示为 Xception 模块图，其中每个 3 ×3 卷积单独的处理一个通道后合并。深度可分离卷积主要分为两个过程：逐通道卷积(depthwise convolution)和逐点卷积(pointwise convolution)[14]。逐通道卷积是将每个输入的通道分别进行独立的空间卷积；点卷积也是 1×1 卷积，通过深度卷积输出的通道作为输入，映射至一个崭新的通道空间。但在 Xception 中先执行 1×1 卷积，之后再执行通道卷积，其顺序差异并不重要，因为这些操作是在堆叠的环境中使用。而 XceptionTime 是由 Elahe 受到 Xception 和 InceptionTime[16] 启发提出来的，其中 InceptionTime 对于时间序列信号的分类能够更准确更快，因此作者将这两种网络性能优势结合设计了一种新的网络架构，图 4 所示为 XceptionTime 中的深度可分离卷积结构图。

本文采用 XceptionTime 模块主要由卷积层、深度可分离卷积层和池化层构成。设置输入样本大小

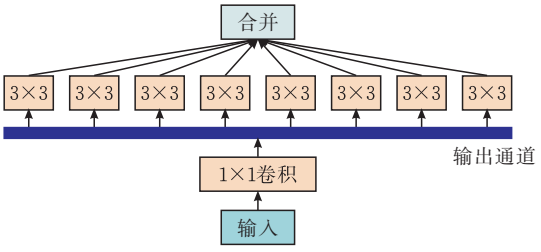


图 3　Xception 模块

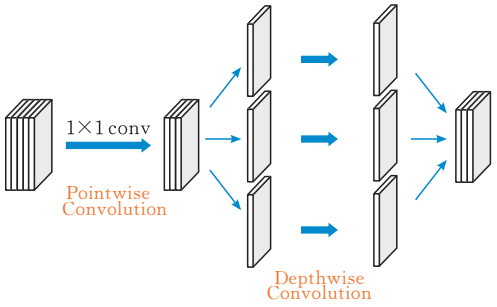


图 4　XceptionTime 中的深度可分离卷积结构

为 $S \times C \times W$，其中 S 为训练批次大小，C 为通道数，W 为宽度即本文设置的滑动窗口长度。将处理后的肌电信号分别输送到两个并行通道，第一条路径经过 1×1 卷积处理后输送到三组深度可分离卷积层，每组卷积的卷积核的大小分别设置为 11、21、41，三组不同核大小的滤波器可以同时提取手势中长时间序列特征和短时间序列特征；另一条路径由最大池化层和 1×1 卷积层组成，最后将两条路径的特征信息进行合并。本文使用 XceptionTime 模块能够在基本不增加网络复杂度的前提下提高模型的分类效果，其核心就是利用卷积核独立处理通道相关性和时间相关性的运算，同时减少了参数量，提升了运行速度，使得特征提取过程更高效。

1.2.2　SE 模块　注意力机制的合理使用对于提高网络模块的性能具有重要作用。SE 模块[17]是注意力机制的一种，它的设计简单，可以直接与现有的先进的体系结构一起使用，如 ResNet[18]、Inception[19]网络等，通过嵌入 SE 模块可以提高网络的性能帮助捕获空间相关性而不需要额外监督的机制，图 5 为 SE 模块的结构图。SE 模块的核心是学习不同通道间的特征并使用，有选择性地关注更具辨别力的特征，并抑制无关信息，显著增强网络的表征能力。

SE 模块的过程分为三个过程：压缩（f_{sq}）、激励（f_{ex}）与重标定（f_{scale}）[20]。其公式如下所示：

$$f_{sq}(z_k) = \mathrm{GAP}(z_k) \tag{1}$$

$$s = f_{ex}(x, W) = \sigma(W_2 \delta(W_1 x)) \tag{2}$$

$$f_{scale}(z_k, s_k) = z_k \cdot s_k \tag{3}$$

式中：GAP 代表全局平均池化，W_1 和 W_2 代表两个全连接层分类器的参数，δ、σ 分别代表 ReLU 激活函数和 Sigmoid 激活函数。s 代表激励模块获得各通道权重。

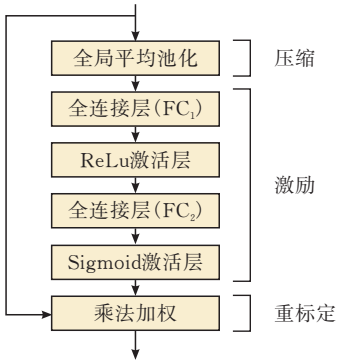


图 5　SE 模块结构

压缩操作是在卷积层之后执行全局平均池化操作，将通道特征信息压缩成实数也就是当前卷积层的全局压缩特征量，从而使感受野延伸至全局范围，以此得到更抽象的手势语义信息；激励操作主要由两个全连接层（FC_1、FC_2）和两个激活函数（ReLU、Sigmoid）组成。此操作主要是为了获取每个特征通道之间的非线性关系，并学习每个特征通道的权重大小，以此来表示该特征通道的重要程度。其中 FC_1 是为了给通道降维，之后采用 ReLU 函数进行非线性激活。FC_2 则是进行升维，使之恢复到原始的通道维度，然后通过 Sigmoid 进行激活。最后重标定模块是通道权重相乘，也就是将 f_{ex} 获得的各通道权重值分别与原始特征对应通道相乘，得到加权后的特征作为输出。

本文通过在深度可分离卷积后中加入 SE 模块，深度可分离卷积能够抽离通道和时序的混杂特征，之后通过 SE 模块直接学习通道特征，计算出不同手势特征信息的权重，使其关注不同通道间的信息，增强有效特征的重要性，从而抑制无效特征的干扰。

1.3　APTx 激活函数

在人工智能领域中寻找最优激活函数一直是一个挑战，早期激活函数已被用来补偿人工神经网络的非线性或改善梯度消失问题。其中 ReLU 的收敛速度和计算速度更快，但当 ReLU 的输入值为负时，会导致神经元不能更新参数，使得神经元失活，进而导致数据多样化的丢失。因此本文将肌电信号手势识别网络框架中的 ReLU 激活函数替换成 APTx 激活函数[21]（alpha plus tanh times）。APTx 激活函数解决了梯度消失和稳定增益的问题，在许多任务中它比 ReLU 激活函数有更好的性能，但它最大优势在训练阶段执行反向传播时的数学运算少，反向传播需要计算每个 epoch 的导数，APTx 激活函数需要更少的数学运算来计算其导数，以此较

少的计算需求加快了网络模型训练的速度，因此也减少了深度学习模型的硬件需求，提升了整个模块的表征能力。APTx 的公式如下：

$$\Phi(x) = (\alpha + \tanh(\beta x)) * \gamma x \tag{4}$$

$$\Phi'(x) = \gamma(\alpha + \tanh(\beta x) + \beta x \operatorname{sech}^2(\beta x)) \tag{5}$$

本网络中，设 $\alpha = 1$，$\beta = 1$，$\gamma = 1/2$，为了对 APTx 进行可视化分析，APTx 激活函数图表及其导数图表如图 6 所示。

$$\Phi(x) = (1 + \tanh(x)) * \frac{x}{2} \tag{6}$$

$$\Phi'(x) = \frac{(1 + \tanh(x) + x \operatorname{sech}^2(x))}{2} \tag{7}$$

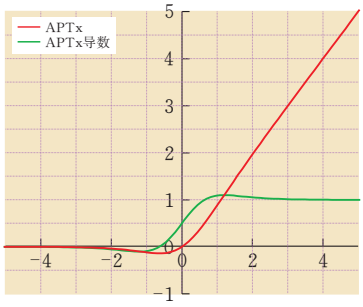


图 6 APTx 函数图及其一阶导数

2 实验结果及分析

2.1 实验数据集

本文实验采用公开的表面肌电信号公开数据集的 NinaPro DB1 数据[22]。该数据集分别由男性受试者 20 名和女性受试者 7 名(右撇子 25 名，左撇子 2 名)组成，共包含 27 名完好受试者的数据。其中，其手势动作共 52 种，手势模型共分为四组：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。在 DB1 的数据集中，sEMG 信号是用 10 个 Otto Bock MyoBock 13E200 进行数据捕获的，采样率为 100 Hz。受试者每个手势重复 10 次，每次持续 5 s，然后休息 3 s。

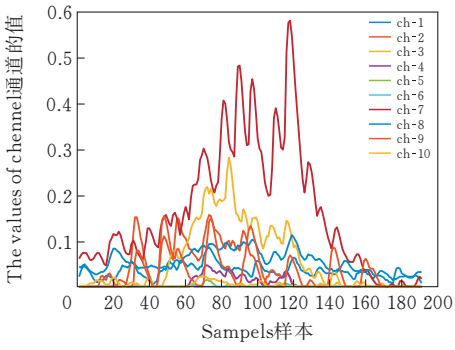
2.2 数据预处理

为了提高网络模型识别手势类别的性能，在构建数据集时，数据预处理是一个重要的步骤。本实验采用低通巴特沃斯滤波器对肌电信号活动进行了降噪，并将处理的肌电信号通过归一化进行处理。本文使用的是基于 μ-law 变换的表面肌电信号非线性方式，称为 μ-law 归一化，这种归一化的方法在语音和通信领域普遍使用，其效果较常用的 min－max 归一化方法更优。μ-law 归一化公式如下：

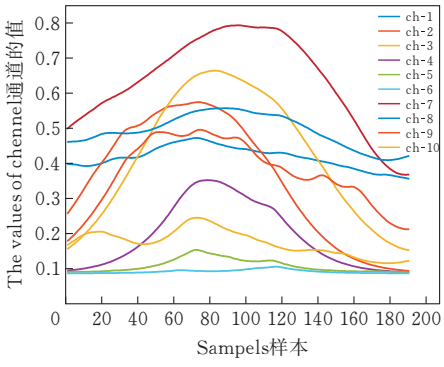
$$F(x_t) = \operatorname{sign}(x_t) \frac{\ln(1 + \mu |x_t|)}{\ln(1 + \mu)} \tag{8}$$

其中，x_t 为归一化的输入标量，μ 为 256。

图 7 为原始的表面肌电信号和使用 μ-law 归一化处理后的表面肌电信号图。



(a)原始的 sEMG 信号



(b) μ-law 归一化处理 sEMG 信号

图 7 未处理的 sEMG 信号和使用归一化处理后的 sEMG 信号

滑动窗口方法也是一项重要的预处理方法，首先它可以增加 sEMG 数据集的训练样本总数，其次通过此类方法能够有效地提高手势分类精度。而对于基于表面肌电信号的实时控制，输入延迟是一个需要考虑的重要因素[23]，因此本文实验选择 150 ms 和 200 ms 滑动窗口对肌电信号进行分割，图 8 所示为滑动窗口分割表面肌电信号。增量窗的大小对手势识别的精度也有影响，因此在本文中使用了四个增量窗来验证本文模型的性能，分别为 25、50、75、100 ms。

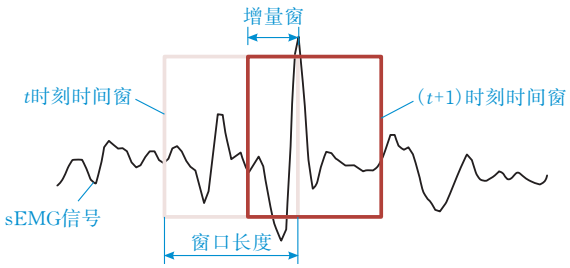


图 8 滑动窗口结构

2.3 实验设置

实验平台以 Ubuntu20.04 操作系统，基于深度学习 pytorch＝1.8.0 框架实现。其中硬件配置为：Inter Xeon(R) Gold 5218 处理器、Nvidia Titan X 显卡和 32GB 显存。将预处理后的 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：其验证集样本由受试对象 s_1、s_5 和 s_7 组

成，其余的均为训练集样本。本文的训练批次大小(batch-size)为 32，训练 100 个 epoch，使用 adam 优化器进行训练并将学习率设置为 0.001。根据 DB1 数据库 sEMG 手势的识别结果，计算 M 个人手势分类精度，公式定义如下：

$$\text{平均手势分类准确率} = \frac{1}{M}\sum_{M}\frac{\text{正确分类手势数量}}{\text{手势分类总数}} \quad (9)$$

2.4　实验分析

增量窗大小对网络模型识别手势准确率的影响如表 1 所示，本文采用 200 ms 的滑动窗口，并设置不同的增量窗口(25 ms、50 ms、75 ms 和 100 ms)对表面肌电信号进行分割，以此来测试不同的增量窗口对本文框架模型的性能影响。增量窗大小会影响分割 sEMG 样本数量，增量窗越大则样本数越多，反之则越少。表 1 对受试对象 s_1、s_5 和 s_7 的 52 种手势进行了测试，从表 1 中得出受试对象 s_1 和 s_5 使用 50 ms 的增量窗均高于其他受试对象，虽受试对象 s_7 使用 25 ms 增量窗的手势识别率最高，但总体来说还是增量窗为 50 ms 的更佳，因此本文使用 50 ms 的增量窗口来提高手势识别的实时性。

表 1　增量窗大小对手势识别准确度的影响

窗口大小/ms	Acc/%		
	s_1	s_5	s_7
25	92.47	81.54	85.03
50	**93.97**	**83.90**	**84.30**
75	89.37	79.80	81.93
100	88.16	78.91	79.88

为证明在整体框架中使用 APTx 激活函数的合理性，本文分别使用 ReLU 和 APTx 两种激活函数进行验证实验。数据预处理采用 50 ms 的增量窗，结果如表 2 所示。实验结果发现，使用 APTx 代替原始模块中的 ReLU 激活函数后，受试对象 s_1、s_5 和 s_7 的 52 种手势识别准确率均有所提升。其中受试者 s_1 使用 150 ms 的滑动窗口时，APTx 相较于 ReLU 激活函数仅提升了 0.7%；而 s_5 设置滑动窗口为 200 ms 时，提升了 1.58%，手势识别准确率提升效果最明显，也充分说明了 APTx 激活函数的性能更优。

表 2　激活函数的验证实验

方法	滑动窗大小/ms	Acc/%		
		s_1	s_5	s_7
ReLU	150	91.84	81.84	81.91
	200	92.39	82.18	82.58
APTx	150	**92.54**	**82.71**	**82.79**
	200	**93.97**	**83.90**	**83.30**

表 3 为本模型在 NinaproDB1 数据集上与近年来其他主流模型的手势识别分类精度的对比结果。从表中对比结果发现，本文提出的 SE-XceptionTime 网络模型的手势识别精度均高于其他网络模型。当网络模型的输入为原始 sEMG 信号时，文献 AtzoriNet、HuNet、WeiNet 分别使用的简单的 CNN 网络模型、基于注意力的混合 CNN-LSTM 网络模型和多流 CNN 模型去识别分类，其中预处理的滑动窗口设置为 150 ms，最终手势分类结果分别达到了 66.60%、83.7% 和 84.4%，而本文所使用的 SE-XceptionTime 网络模型相较于它们提升了 20.66%、4.56% 和 2.86%。文献 YuNet 和 FanNet 则使用 CNN＋SVM(支持向量机)和 BiLSTM(双向长短期记忆网络)进行手势识别分类，分类精度达到了 86.19% 和 88.50%，手势识别率均低于本网络模型，由此说明了本网络模型的优越性。GengNet 使用 sEMG 图像作为输入，也就是将原始的 sEMG 信号处理为二维灰度图像，之后使用 CNN 网络和简单投票机制结合的方法，设置滑动窗口为 200 ms，52 种手势识别精度达到了 77.8%，而本模型相较于它提升了 10 个百分点。与现有的主流模型相比，本模型设置滑动窗口无论是 150 ms 还是 200 ms，其手势分类精度都有所提高，也证明了选用该模型识别 sEMG 信号进行 52 种手势分类的有效性。其中图 9 为 27 位受试对象的 52 种手势分类准确率。

表 3　本文模型与其他方法的分类精度对比

一	方法	输入	手势数	滑动窗大小/ms	Acc/%
AtzoriNet[6]	CNN	raw sEMG	50	150	66.60
GengNet[8]	CNN＋ majority voting	sEMG image	52	200	77.80
HuNet[9]	Attention-base hybrid CNN-LSTM	raw sEMG	52	150	83.70
				200	84.80
WeiNet[10]	MS-CNN	raw sEMG	52	150	84.40
				200	85.00
YuNet[12]	CNN＋SVM	raw sEMG	52	200	86.19
FanNet[11]	BiLSTM	raw sEMG	52	200	88.50
Ours	SE-XceptionTime	raw sEMG	52	150	**87.26**
				200	**88.72**

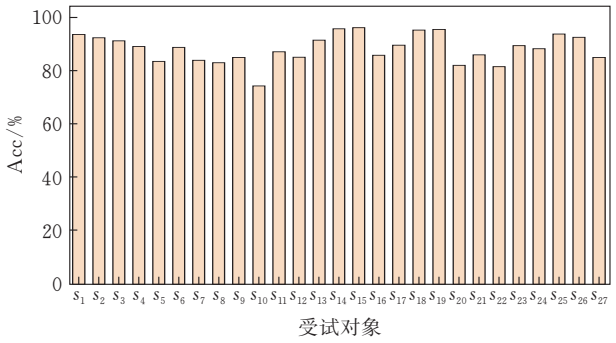


图 9 受试对象的 52 种手势分类准确率

为了验证本文所提出的网络架构中各个模块的性能，以 XceptionTime 网络作为基准模型（Baseline）在数据集上进行验证，原始的表面肌电信号进行滤波和 μ-law 归一化处理后，使用 200 ms 的滑动窗口进行分割，其中增量窗设置为 50 ms，再对三组深度可分离卷积后中加入的 SE 模块和 APTx 激活函数进行消融对比实验，结果见表 4。相较于 Baseline，本文所提出的网络模型极大地提高了手势识别精度，手势识别准确率由 84.15% 提升至 88.72%。本实验在 Baseline 上分别加入 SE 模块和 APTx 激活函数，网络的识别性能分别达到了 86.97% 和 85.53%，与 Baseline 相较，手势识别的准确度均有所提升。由此验证了在 XceptionTime 网络中加入的两个模块可以让网络有效地选择手势特征中的关键信息，从而提升了网络的识别精度。

表 4 不同模块的消融实验

	方法	Acc/%
1	Baseline	84.15
2	＋SE 模块	86.97
3	＋APTx 激活函数	85.53
4	＋ SE 模块＋APTx 激活函数	88.72

3 结论

为了抽离表面肌电信号中时序、通道的混杂特征，并充分挖掘通道中的有效特征，本文提出了基于 SE 注意力与 XceptionTime 的肌电信号手势识别网络方法。该模型中的深度可分离卷积通过设置不同大小的滤波器来提取时序特征，以此将时序与通道特征分开学习。并通过加入的 SE 注意力使其关注不同通道的手势特征，达到充分提取手势信息的目的；同时使用自适应平均池化，降低了输出特征维数，避免过拟合，且配合 APTx 激活函数的使用，让模型的表征能力得到提升，从而使本网络方法对肌电信号手势进行识别的能力有了明显增长。为了验证模型的有效性，与现有的深度学习模型的手势识别进行对比，实验结果表明该网络模型手势识别准确率达到了 88.72%，显著高于大多主流模型。

[参 考 文 献]

[1] FANG C, HE B, WANG Y, at al. EMG-centered multisensory based technologies for pattern recognitionIn rehabilitation: State of the Art and Challenges [J]. Biosensors, 2020, 10(08): 85.

[2] PARAJULI N, SREENIVASAN N, BIFULCO P, at 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.

[3] 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.

[4] 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(01):13-25.

[5] 曹梦琳，陈宇豪，王钰，等. 基于表面肌电图的人体运动意图识别研究进展[J].中国康复理论与实践 ，2021，27(05):595-603.

[6] AMOR A, GHOUL O E, JEMNI M . Toward sign language handshapes recognition using myo armband [C]// International Conference on Information & Communication Technology & Accessibility. 2017.1-6.

[7] KETYKÓ I, KOVÁCS F, VARGA K Z. Domain adaptation for semg-based gesture recognition with recurrent neural networks[C]// Proc Of The International Joint Conference On Neural Networks. IEEE, 2019: 1-7.

[8] GENG W D, DU Y, JIN W G, et al. Gesture Recognition By Instantaneous surfaceemg Images[J]. Scientific Reports, 2016, 6(01): 1-8.

[9] 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.

[10] 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.

[11] 范秀琴，喻洪流，李素姣. 基于 BiLSTM 的表面肌电图手势识别算法[J]. 生物医学工程学进展，2021，42(02):92-96.

[12] 余明超. 基于深度学习和肌电特征选取的手部动作识别研究[D].武汉：武汉科技大学，2020.

[13] ELAHE R, SOHEIL Z, SEYED F A, et al. Xceptiontime:a novel deep architecture based on depthwise sep-

arable convolutions for hand gesture classification[EB/OL]. (2019-11-09). https://arxiv.org/abs/1911.03803.

[14] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251-1258.

[15] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818-2826.

[16] HASSAN I F, BENJAMIN L, GERMAIN F , et al. Inceptiontime: finding alexnet for time series classification[EB/OL]. (2019-09-11). https://arxiv.org/abs/1909.04939.

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of The IEEE Conference on Computer Vision And Pattern Recognition, 2018: 7132-7141.

[18] KAIMING H, XIANGYU Z, SHAOQING REN, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[19] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]// Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[20] CHEN H, ZHANG Y, ZHOU D, et al. Improving gesture recognition by bidirectional temporal convolutional netwoks[C]// International Conference on Robotics and Rehabilitation Intelligence. Springer, Singapore, 2020: 413-424.

[21] KUMAR R. APTx: better activation function than mish, swish, and relu's variants used in deep learning [EB/OL]. (2022-09-23). https://arxiv.org/abs/2209.06119.

[22] 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.

[23] HUDGINS B, PARKER P, SCOTT R N. A new strategy for multifunction myoelectric control[J]. IEEE Transactions On Biomedical Engineering, 1993, 40 (01): 82-94.

sEMG Gesture Recognition based on SE Attention and Xception Time

LIU Cong[1,2], XU Tingting[1], HU Sheng[1]

(1 *School of Electrical and Electronic Engin., Hubei Univ. of Tech., Wuhan* 430068, *China*;
2 *Wuhan Hua'an Science and Tech. Co., Ltd., Postdoctoral Workstation, Wuhan* 430068, *China*)

Abstract: Conventional convolutional neural network—based surface electromyography signal gesture recognition algorithms generally have poor recognition accuracy. The reasons for this are as follows: first conventional convolution does convolution and summation on the input channels, which in turn causes channel features and timing features of the sEMG signal to be mixed; second, these algorithms do not sufficiently extract the features between different channels in the sEMG signal. To address this problem, this paper innovatively proposes sEMG gesture recognition model based on SE attention and XceptionTime. First, the original sEMG signal is pre—processed with filtering, normalisation and sliding window operations respectively; Then the SE-XceptionTime module in the network then maps channel features to temporal features separately by using depthwise separable convolutions and sets three different convolution kernel size filters to extract long or short temporal information; finally, the SE attention mechanism is added after this convolution, a SE attention mechanism is added after this convolution to focus on the effective features in the channel and weaken the interference of irrelevant information. The experimental results show that the accuracy of 52 types of gesture recognition on the NinaproDB1 dataset reaches 88.72%, which is higher than the current mainstream deep learning gesture recognition network models, fully validating the effectiveness of the network model in this paper.

Keywords: surface electromyography signals; gesture recognition; depthwise separable convolution; attentional mechanisms

[责任编校: 闫　品]