

基于短时动态卷积的吞咽动作识别研究

周子森¹, 杨建^{2*}

¹五邑大学电子与信息工程学院, 广东 江门

²五邑大学机械与自动化工程学院, 广东 江门

收稿日期: 2026年3月24日; 录用日期: 2026年4月22日; 发布日期: 2026年4月30日

摘要

吞咽困难或吞咽功能障碍严重影响人类健康、尤其影响老年人群体, 而目前现有的吞咽识别模型存在精度不足、模型参数冗余等问题。针对上述问题, 本研究提出采用自主研发的柔性压电聚乳酸传感器作为数据采集终端, 通过一种短时动态卷积(STDyConv)方法, 用于分析受试者进行吞咽动作时采集的生理信号。该方法是由一个短时频率自适应卷积模块和一个全局传递卷积核组成, 其中短时自适应卷积模块的参数存储在两个张量中, 分别代表频率特征的实部和虚部。通过基于输入生理信号参数微调和优化, 卷积核能够精确响应输入信号的多维频率特征。为了进一步增强模型的场景适应能力, 在每个卷积核中引入了频率加权因子, 以实现频率特定特征的差异化关注。全局传递卷积核仅需存储一个初始卷积核, 其余卷积核均基于该初始卷积核生成, 有效抑制了动态卷积模型参数的冗余增长。实验结果表明, STDyConv与传统的卷积神经网络(CNN)架构具有良好的兼容性, 并且与传统方法相比, 吞咽动作识别的平均准确率提高了20%。

关键词

人体运动识别, 柔性压电传感器, 时间序列分类, 吞咽动作, 动态卷积

Studies on Swallowing Action Recognition Using Short-Time Dynamic Convolution

Zisen Zhou¹, Jian Yang^{2*}

¹School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

²School of Mechanical and Automation Engineering, Wuyi University, Jiangmen Guangdong

Received: March 24, 2026; accepted: April 22, 2026; published: April 30, 2026

*通讯作者。

文章引用: 周子森, 杨建. 基于短时动态卷积的吞咽动作识别研究[J]. 计算机科学与应用, 2026, 16(4): 428-438.
DOI: 10.12677/csa.2026.164142

Abstract

Dysphagia or swallowing dysfunction seriously affects human health, especially for the elderly. At present, the existing swallowing recognition models have problems such as insufficient accuracy and redundant model parameters. To address the aforementioned issues, this study proposes using a self-developed flexible piezoelectric polylactic acid sensor as the data acquisition terminal and employing a Short-Time Dynamic Convolution (STDyConv) method to analyze physiological signals collected during swallowing. This method consists of a short-time frequency adaptive convolution module and a globally propagating convolution kernel. The parameters of the short-time adaptive convolution module are stored in two tensors, representing the real and imaginary parts of the frequency features, respectively. Through parameter fine-tuning and optimization based on the input physiological signals, the convolution kernel can accurately respond to the multidimensional frequency features of the input signal. To further enhance the model's scene adaptability, a frequency weighting factor is introduced into each convolution kernel to achieve differentiated attention to frequency-specific features. The globally propagating convolution kernel only needs to store one initial convolution kernel, and all other convolution kernels are generated based on this initial kernel, effectively suppressing the redundant growth of parameters in the dynamic convolution model. Experimental results show that STDyConv has good compatibility with traditional convolutional neural network (CNN) architectures, and compared with traditional methods, the average accuracy of swallowing action recognition is improved by 20%.

Keywords

Human Action Recognition, Flexible Piezoelectric Sensor, Time Series Classification, Swallow Action, Dynamic Convolution

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

吞咽是由神经肌肉系统精确调控的复杂生理协同过程。正常吞咽依赖于舌体蠕动产生的压力梯度与咽肌收缩的节律性配合,共同驱动食团由口腔向咽部安全转运[1][2],从而构成了防止误吸的核心生理屏障[2]-[4]。一旦该机制失调(即吞咽障碍),食物颗粒误入喉腔极易诱发致命的吸入性肺炎[5],并导致营养不良与生活质量显著下降[6]。

目前,临床诊断虽以视频透视吞咽功能检查(Video Fluoroscopic Swallowing Study, VFSS) [7]与纤维内窥镜吞咽功能评估(Fiberoptic Endoscopic Evaluation of Swallowing, FEES) [8]被公认为评估吞咽功能的“金标准”。但其高昂成本、放射性风险及侵入性不适限制了其在大规模筛查与居家监测中的普及。

近年来,柔性可穿戴传感器凭借无创、实时及高顺应性等优势,为吞咽生理信号的客观量化提供了新路径。文献[9]开发了一种基于超轻铝氮化物(AIN)的传感器,其结构基于一层聚酰亚胺薄膜,并通过一层聚二甲硅氧烷-聚乙二醇醚聚合物(PDMS-PEIE)粘合剂固定在皮肤上。当固定在颈部时,该传感器的可贴合性和灵敏度能够提供高性能且不会干扰正常的喉部功能。文献[10]开发了一种离子聚合金属复合材料(IPMC)压电传感器,用于检测吞咽动作,并采用支持向量机(SVM)来区分吞咽动作和非吞咽动作。

类似地, 文献[11]设计了一种项链式的创新可穿戴设备, 在脖子上放置两个垂直的压电传感器, 并利用长短期记忆网络(Long Short-Term Memory, LSTM) [12]来检测和量化吞咽。

然而, 现有文献多聚焦于动作的“发生性”检测或阶段划分, 缺乏对复杂康复训练动作规范性的深度解析[13][14]。此外, 针对老年群体等复杂应用场景, 现有深度学习模型普遍面临特征表征精度瓶颈与计算资源冗余的双重挑战。针对上述痛点, 本研究提出利用高灵敏度、强贴合性的柔性压电传感器实现颈部微弱生理信号的精准捕获, 并构建了涵盖多种复杂康复模式的原始数据集。在此基础上, 本文提出一种频率信息引导的动态卷积架构, 通过信号频域特性驱动卷积核的自适应生成。该模型作为即插即用的模块集成于常规 CNN 中, 在强化复杂特征提取能力的同时, 通过参数共享传递机制显著降低了计算开销。在自建数据集及公开人体活动识别任务中的实验结果表明, 该方法在保持轻量化的同时, 具有显著的性能优势与工程应用潜力。

2. 短时动态卷积

本研究提出了一种短时动态卷积(Short-Time Dynamic Convolution, STDyConv), 该卷积由频率调制模块与参数传递机制协同构成。在频率调制模块中, 卷积参数解构为两个特征张量, 分别映射频率信号的实部与虚部特征; 通过获取输入生理信号的频率信息, 驱动卷积核权重实现对输入生理信号特征的差异化聚焦进行精确的参数微调。此外, 为增强模型的场景泛化性, 研究引入了动态权重因子以实现输入生理信号的精准响应。在参数量优化方面, 参数传递机制仅需单一基础卷积核, 其余卷积参数均由此衍生, 有效遏制了动态网络中参数冗余的增长。整体流程如图 1 所示, 其中包括两个部分, 基于频率信息的动态卷积生成部分和全局参数传递部分。

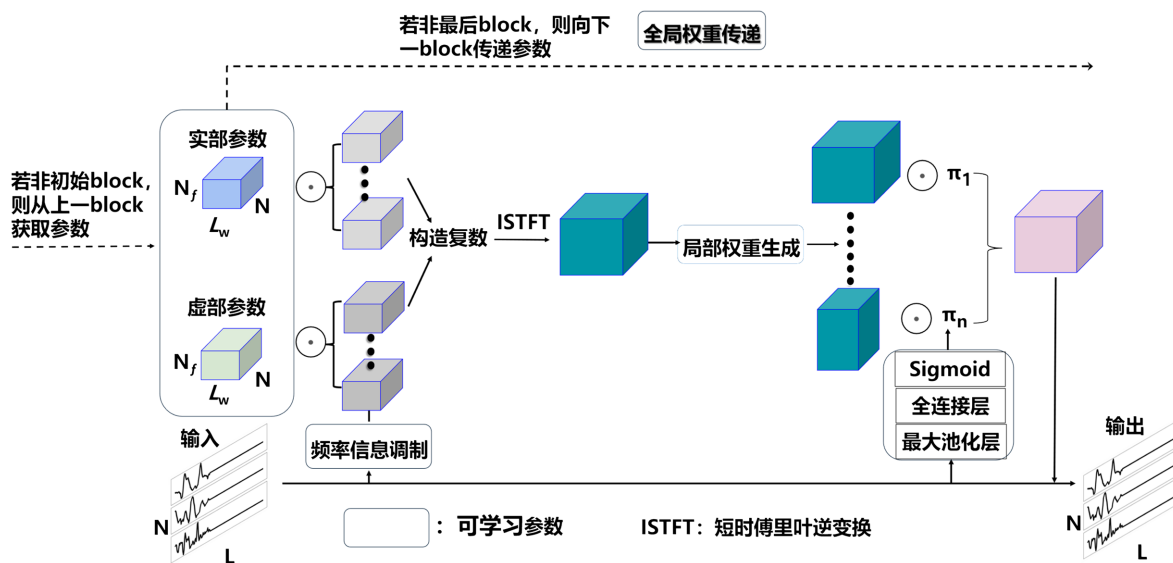


Figure 1. STDyConv model structure
图 1. 短时动态卷积 STDyConv 模型结构

2.1. 频率信息调制

动态卷积(Dynamic Convolution)神经网络[15]的核心是通过动态调整模型结构或参数来适应不同的样本。近年来, 凭借其输入自适应计算特性在计算机视觉任务中表现出显著优势。例如, CondConv [16]和 ODConv [17]并行学习多组权重, 并通过注意力机制进行融合。利用 sigmoid 函数完成多核线性混合。

ODConv 通过引入三重注意力机制(通道、过滤和空间维度的联合加权), 以实现更精细的权重调节。该策略虽然增强了模型的表达能力, 但多组参数的引入导致计算开销明显增加。

动态卷积通过将标准一维卷积中的卷积核权重替换为一组卷积核长度相同的 n 个“专家”卷积核 $\{W_1, W_2, \dots, W_n\}$ 来提高卷积的适应性。每个权重的输出通过一组基于输入动态生成的注意力系数 $\{\pi_1, \pi_2, \dots, \pi_n\}$ 。在大部分的动态卷积设置中, 这组注意力系数是通过全局池化层后再经过一个全连接层处理所得到的。综上所述, 动态卷积的计算过程可以表示为:

$$Y = \sum_{i=1}^n \pi_i W_i * X \quad (1)$$

在时间序列分析任务中, 频域特征始终是捕获信号动态特性的关键。受此启发, 本研究提出了一种基于频率信息的卷积核权值优化策略, 如图 2 所示。在 STDyConv 模型中, 可学习的卷积核参数被解耦并存储于两个张量中, 分别映射卷积核在频域上的实部与虚部分量, 具体来说, STDyConv 的基础可学习参数张量为 $P_a, P_b \in \mathbb{R}^{L_w \times N_f \times N}$, 该模块通过调制张量(Modulation Tensor)对基础参数张量进行动态调整, 而调制张量的生成则高度依赖于输入信号的频率分布。

具体的调制流程如下。首先, 对一输入信号 $X \in \mathbb{R}^{C_{in} \times L}$ 应用离散傅里叶变换(DFT)以提取频域特征, 随后, 基于获取的频率信息构建一组自适应特征字典, 即:

$$\{f_1, f_2, \dots, f_k\} = \arg \operatorname{topk}_{i>0}^k (|\operatorname{DFT}(X)|) \quad (2)$$

其中, $\operatorname{DFT}(\cdot)$ 表示对一维序列执行离散傅里叶变换; $|\cdot|$ 为取模运算, 对应频率中的幅值谱; $\arg \operatorname{topk}_{i>0}^k(\cdot)$ 表示提取幅值序列中降序排列的前 k 个点的位置索引, 且 $k = N_f$ 为超参数, 这使得在后续生成调制张量时对齐基础可学习参数张量。对于多通道时间序列, 每个 f_i 是一个元素个数等于通道数的列向量。通过该方法, 我们能够捕获贡献度最高的几个频率成分, 这些成分通常映射了时域中的核心特征。在这里, 本研究剔除了零频率分量, 因为这通常是代表直流偏移量等无关项, 不包含吞咽等生理动作的动态特征 [18]。接下来, 我们对这组特征字典应用 LSTM, 生成调制张量:

$$P' = \operatorname{LSTM}(\{f_1, f_2, \dots, f_n\}) \quad (3)$$

$$\{P_{a_1}, P_{b_1}, \dots, P_{a_m}, P_{b_m}\} = \operatorname{split}(\operatorname{reshape}(P')) \quad (4)$$

其中 $P' \in \mathbb{R}^{L_N \times N}$, $L_N = L_w \times k \times m \times 2$, m 是超参数, 表示卷积滤波器的数量; $\operatorname{reshape}(\cdot)$ 表示对张量改变形状, 将 P_m 变为 $(m \times 2 \times L_w, k, N)$ 的形状; $\operatorname{split}(\cdot)$ 表示将张量进行拆分为一系列调制张量序列 $\{P_{a_1}, P_{b_1}, \dots, P_{a_m}, P_{b_m}\}$, 且 $P_{a_i}, P_{b_i} \in \mathbb{R}^{L_w \times k \times N}$, $i=1, 2, \dots, m$, 与 P_a, P_b 形状一致。将 P_a, P_b 分别与调制张量序列执行 Hadamard 积并将其重构为复数频域表征。最后, 通过短时傅里叶变换逆变换(ISTFT)还原得到时域卷积核 $\{F_1, F_2, \dots, F_m\}$, $F_i \in \mathbb{R}^{L_k \times N}$, $i=1, 2, \dots, m$ 。在默认配置下, ISTFT 设置为中心对称映射, 则:

经由上述流程, 我们利用输入信号的频域信息, 实现了对 m 个卷积滤波器动态调制, 为了便于后续的卷积运算, 我们将这些独立的滤波器进行维度拼接, 从而构建出完整的动态权重张量。综上所述, 整个调制过程可表示为:

$$W_1 = \operatorname{FM}(X) \quad (5)$$

其中, $\operatorname{FM}(\cdot)$ 表示频率调制操作, $W_1 \in \mathbb{R}^{L_k \times N \times m}$ 表示整个调制模块通过输入信号得到的全部卷积滤波器拼接得到的张量。

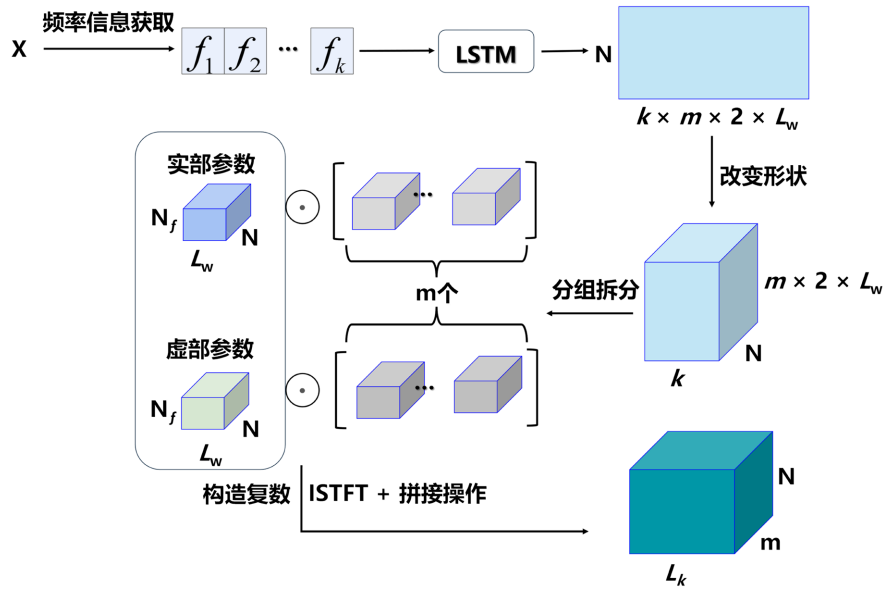


Figure 2. Frequency information modulation
图 2. 频率信息调制的模型结构

2.2. 参数传递机制

在前述章节中, 我们利用信号频率信息动态生成了 m 个滤波器。常规做法通常通过增加基础参数张量的数量并重复调制过程来获取多个“专家”卷积核, 但这会导致模型参数量与计算复杂度显著激增。为平衡表征能力与计算成本, 本研究提出了一种参数传递机制, 该机制通过对基础卷积核执行线性变换, 即卷积操作, 来衍生新内核, 并分为“局部”与“全局”两个维度, 如图 3 所示。

2.2.1. “局部”可传递机制

“局部”传递机制侧重于模块内部的参数演化, 将卷积权值重新审视为一维时间序列信号, 将经典的信号处理手段引入到了权重生成领域。可以将原始卷积核视为 m 个具有 N 维输入维度 L_k 个时间步的时间序列信号, 并对其应用深度卷积运算[19]:

$$W_{same} = DWConv_{padding}(W_1) \quad (6)$$

$$W_{short} = DWConv(W_1) \quad (7)$$

其中, $W_1 \in \mathbb{R}^{L_k \times N \times m}$ 表示初始的卷积核; $DWConv_{padding}(\cdot)$ 表示进行两侧零填充的深度卷积操作, 这将不会改变输出卷积核的形状, 即 $shape(W_{same}) = shape(W_1)$, $shape(\cdot)$ 表示求得张量的形状; $DWConv(\cdot)$ 表示不进行两侧零填充的深度卷积操作, 这将输出卷积核长度更短的卷积核, 从而获得一个不同尺度的卷积核。本研究将用于传递的深度卷积核长度统一设置为 3, 即 $K_{DWConv} = 3$, 则 $shape(W_{short}) = (m, N, L_k - 2)$ 。容易得到, 只要传递卷积的核尺寸远小于原始参数规模, 该机制便能以极低的参数代价获得丰富的多尺度“专家”卷积核。在 STDyConv 架构中, 这些衍生出的多尺度卷积核通过注意力系数进行动态加权。传统动态卷积通常采用加权求和后再执行推理的策略[16]。针对本研究涉及的多尺度特征对齐问题, 我们提出了一种基于零填充的对齐加权方案。对于较短尺度的卷积核, 通过在其末端进行零填充, 使其在维度上与长尺度核对齐, 确保多尺度专家核能够在动态加权框架下协同工作:

$$W' = \{W_1, W_{same}, \dots, align(W_{short})\} \quad (8)$$

$$Y = \sum_{i=1}^n (\pi_i W_i') * X \tag{9}$$

其中 $\text{align}(\cdot)$ 表示通过对序列末端进行零填充的对齐操作。

综上所述, 利用深度卷积的轻量化特性, 本研究实现了“局部”参数传递机制。该设计在丰富卷积核多样性与多尺度表征能力的同时, 有效遏制了参数冗余, 并通过对齐机制实现了多尺度动态特性的高效融合。

2.2.2. “全局”传递机制

“局部”参数传递机制可进一步拓宽至模型“全局”, 在 STDyConv 架构中, 卷积核参数以频域形式存储, 即表现为基础可学习参数张量。受此启发, 本研究提出了跨模块全局传递方案: 当参数由当前模块向下一模块传递时, 通过执行深度可分离卷积操作[19], 实现参数张量的自适应演化与生成。其数学表达式为:

$$\{P_a^{\text{next}}, P_b^{\text{next}}\} = \left\{ \text{DSCConv}_{\text{padding}}(P_a), \text{DSCConv}_{\text{padding}}(P_b) \right\} \tag{10}$$

其中, $P_a^{\text{next}}, P_b^{\text{next}}$ 表示下一模块的频率参数张量, $\text{DSCConv}_{\text{padding}}(\cdot)$ 表示执行两侧进行零填充的深度可分离卷积操作, 为保持计算的一致性, 深度可分离卷积的卷积核长度同样为 3, 即 $K_{\text{DSCConv}} = 3$ 。通过该机制, 模型仅需维护极少量的初始参数, 即可通过链式映射衍生出各层所需的卷积权重。

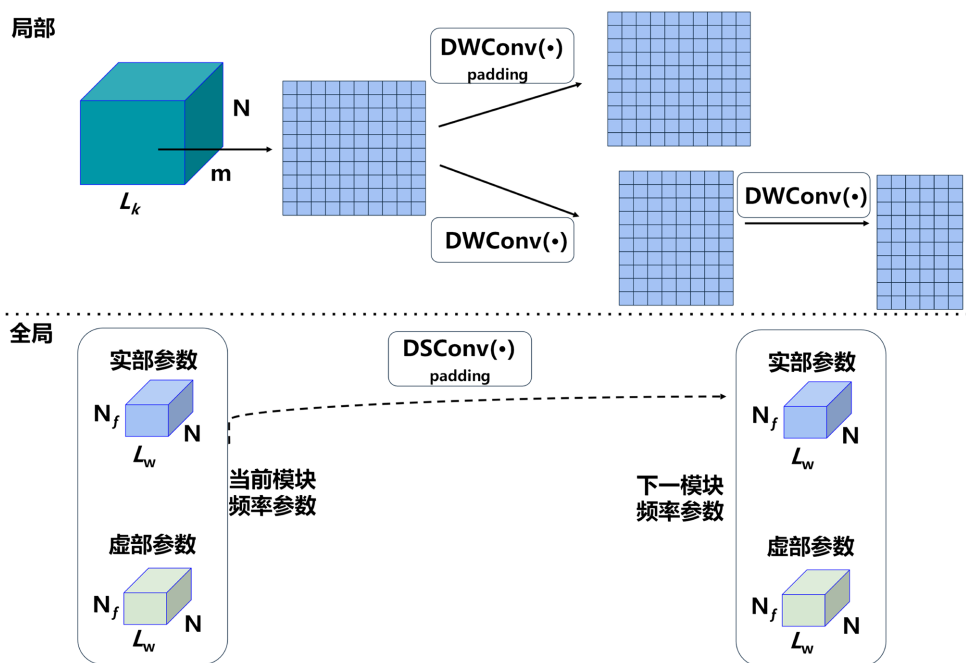


Figure 3. Principle of the transferable parameter mechanism

图 3. 可传递参数机制原理

3. 实验与测试

3.1. 吞咽动作采集

本研究采用自主研发的柔性压电传感器[20]进行人体吞咽行为的数据获取。该传感器以离子液体掺

杂的聚左旋乳酸(PLLA)为核心材料,旨在精准捕捉人体体表的压力特征信号。通过引入 1-丁基-3-甲基咪唑双(三氟甲基磺酰)亚胺([BMIm]TFSI)离子液体(IL), PLLA 薄膜的压电效应得到显著增强。实验表明,当 IL 含量为 1 wt%、经 140°C退火 2 小时及 4 倍拉伸倍率处理后, PLLA/IL 薄膜的输出电压与冲击灵敏度分别提升至原先的 3 倍和 105.1%。此前,该传感器已成功应用于人体咳嗽动作的识别,充分验证了其在人体行为感知领域的有效性。在此研究基础上,我们将该传感器的应用范围进一步延伸至人体吞咽行为监测,显著拓宽了其在人体活动监测领域的应用潜能。本次实验的采集装置维持前期设定:传感器贴附于受试者的喉结处,通过 NI 采集卡实现信号传输,详细的硬件架构如图 4 所示。

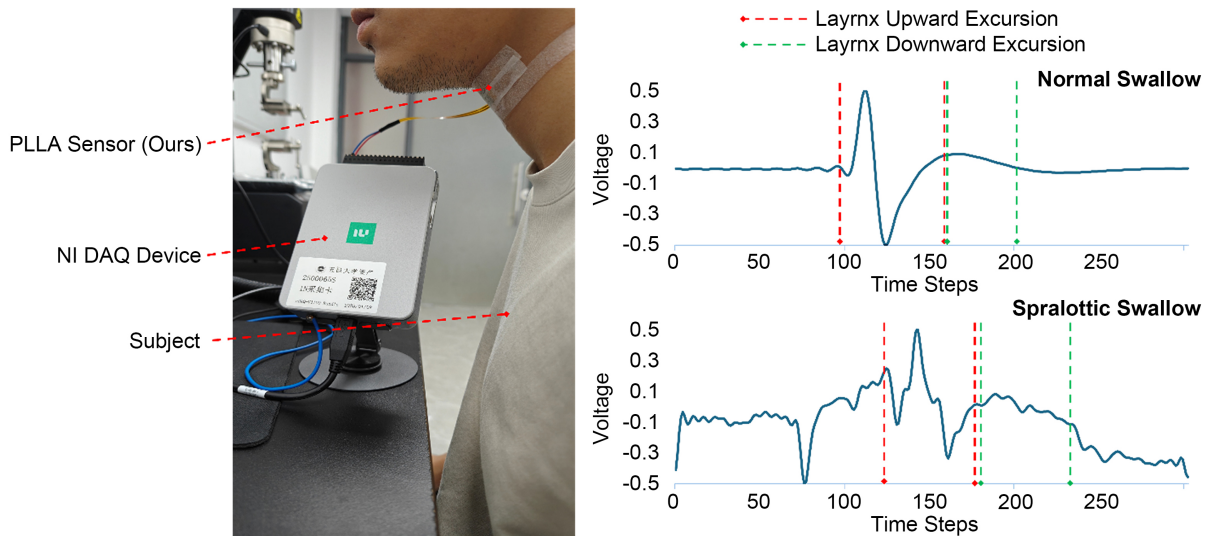


Figure 4. Acquisition and setting of the human swallowing action
图 4. 人体吞咽动作采集设置

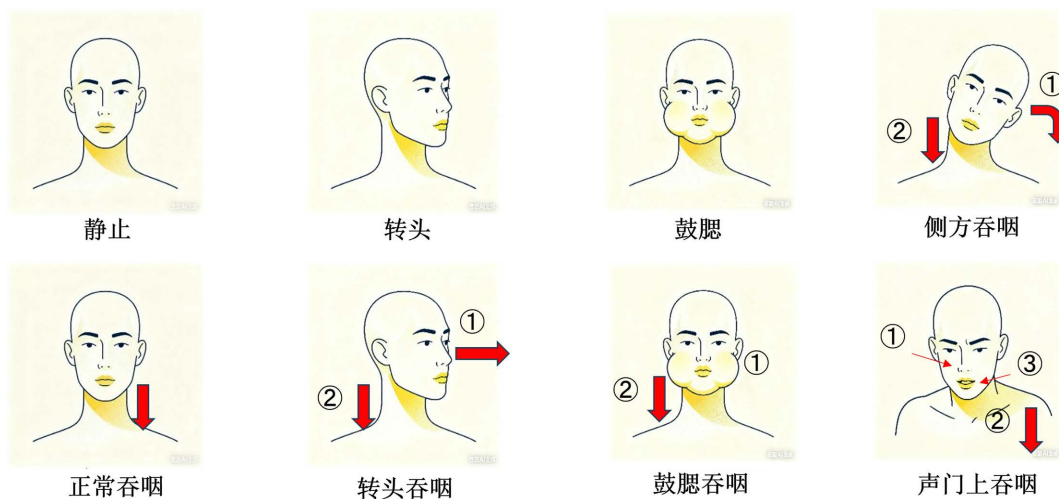


Figure 5. Schematic diagram of the classification of the deglutition action design
图 5. 吞咽动作设计分类示意图

为丰富吞咽动作识别任务的数据维度,本研究精心设计了八类代表性动作:静坐、转头、低语、正常吞咽、鼓腮吞咽、转头吞咽、侧方吞咽及声门上吞咽,如图 5。这些动作的设计综合考量了生理特征、

信号获取的优化以及临床实际应用场景[13][14],旨在全面覆盖吞咽过程中的多样化特征。实验共招募十名受试者(八男二女),每组动作重复 10 次,单次采样时长为 10 秒。NI 采集卡的采样频率设定为 1000 Hz,辅以 0~20 Hz 的带通滤波器。最后,根据模型输入需求,我们将原始数据(每组 10,000 个采样点)压缩至 1000 个点进行后续处理。

3.2. 实验设置

为了衡量泛化性,实验中引入 UEA [21]时间序列档案库的 Handwriting [22]公开数据集,该数据依托智能穿戴设备获取,详细记录了受试者在书写 26 个英文字母过程中的手部运动轨迹。数据集由 150 组训练样本与 850 组测试样本组成,涵盖了三轴加速度与三轴角速度共六个物理维度的动态读数,能够全面表征书写过程中的空间运动特征。为了规避实验结果的偶然性,每个模型在每个数据集上均进行 100 个周期的训练,重复三次,取平均值,并且始终使用 Adam 优化器作为优化器。我们的测试环境是 RTX 2060TI 和英特尔(R)酷睿(TM) i7-10875H 处理器(频率 2.30 GHz)。

为评估模型的表现,我们采用模型推理准确率和 Micro-F1 分数[23]作为评价指标。Micro-F1 分数适用于多类别的 F1 分数计算,通过计算所有类别总和的精确率(Precision)与召回率(Recall)来得到:

$$F1_{\text{micro}} = 2 \cdot \frac{\text{Precision}_{\text{all}} \cdot \text{Recall}_{\text{all}}}{\text{Precision}_{\text{all}} + \text{Recall}_{\text{all}}} \quad (11)$$

4. 结果与讨论

STDyConv 可作为标准卷积算子的有效替代方案。为客观评估其性能,本研究采用全卷积网络模型 [24] (FCN)和 ResNet [25]作为基准架构,将 STDyConv 集成其中来比较各项指标,并且参考了 TimesNet [26]和 PatchTST [27]作为基准参考模型进行实验。FCN、ResNet 和 PatchTST 的实现均参考了开源时序分析库 tsai [28]。STDyConv 的参数如表 1 所示。

Table 1. STDyConv parameter settings
表 1. STDyConv 参数设置

参数	值
多尺度卷积的组数	4
每组多尺度卷积的个数	1
短时傅里叶变换的帧数	7
短时傅里叶变换的频率样本数	4

准确率实验结果如表 2 所示,实验结果表明,引入 STDyConv 模块后,模型在评价指标上均实现了增长,有力证明了该模块在特征增强方面的优越性。同时,实验发现基于 Transformer 的方法并不适用于人体动作时序分类任务,在各项指标上均表现不佳。

接下来,本研究探讨了 STDyConv 模块的集成深度对模型推理性能的影响。实验以 FCN 为基准架构,并在自建吞咽动作数据集上展开评估。由于基础 FCN 包含三个卷积层,我们采用逐层替换策略将 STDyConv 引入模型,旨在考察性能增益与集成数量之间的演变关系。如表 3 所示,随着 STDyConv 集成密度的增加,模型分类准确率呈现稳步上升趋势。这一结果表明,STDyConv 能有效增强模型的特征映射能力,且多层次特征生成的协同效应能显著优化最终的识别效果。

Table 2. Experimental results of integrating STDyConv
表 2. 集成 STDyConv 的实验结果

模型	推理准确率(%)		Micro-F1 分数	
	吞咽动作	Handwriting	吞咽动作	Handwriting
FCN	51.5	41.5	0.52	0.42
FCN + STDyConv	61.7	52.7	0.62	0.53
ResNet	50.5	35.1	0.50	0.35
ResNet + STDyConv	58.9	46.8	0.57	0.53
TimesNet	60.9	32.4	0.61	0.33
PatchTST	50.3	19.5	0.51	0.20

此外, 本研究针对 STDyConv 核心的参数传递机制进行了严谨的消融实验验证。该机制的设计初衷是通过共享基础参数张量来驱动动态卷积核的生成, 以降低模型的存储开销。在验证实验中, 我们构建了一个等效的对比模型, 即弃用参数传递机制, 转而采用多个相互独立的可学习参数张量来执行卷积核的生成过程。实验数据显示, 采用参数传递机制的模型与采用独立参数张量的模型在分类准确率上表现持平。这一结论有力地证实了参数传递机制的科学性: 它在未造成特征表达能力降级的前提下, 通过数学等价的参数复用逻辑, 极大地提升了模型的参数能效比, 验证了其作为轻量化算子的核心优势。

Table 3. Effect of different number of STDyConv modules on model accuracy
表 3. 不同个数 STDyConv 模块对模型准确率的影响

STDyConv 集成的个数	推理准确率(%)
1	53.0
2	55.6
3	61.7

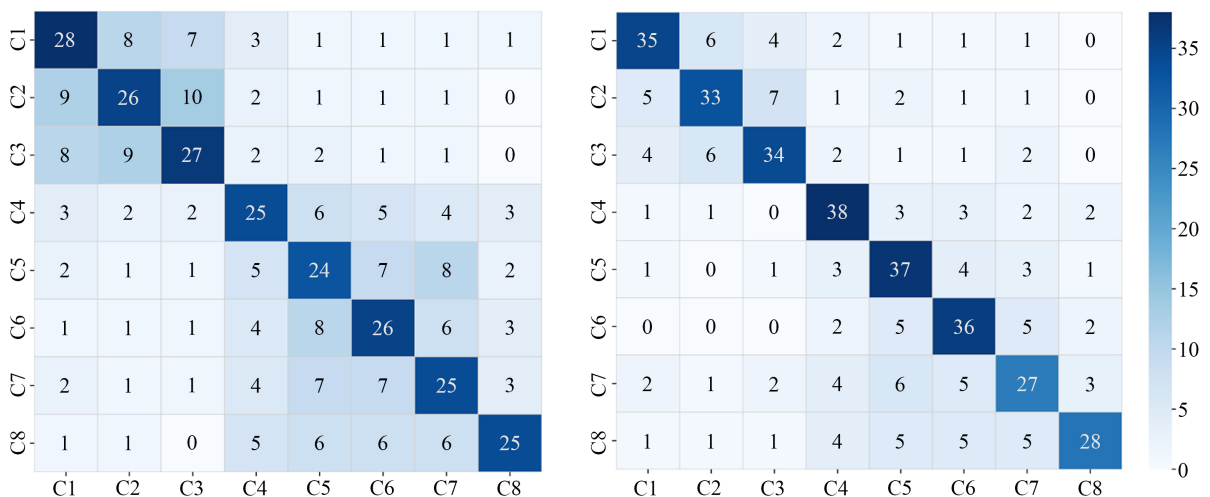


Figure 6. Comparison of confusion matrices for swallowing action classification
图 6. 吞咽动作分类的混淆矩阵对比

为了深层次评估模型对不同吞咽及其伴随动作的细微表征识别能力, 本实验引入混淆矩阵对分类结果进行了可视化剖析。如图 6 所示, 性能的显著增益高度集中于几类吞咽动作之上, 并能够与用于对比的类似非吞咽动作进行区分。在基准模型中, 由于上述动作引发的颈部生理信号在时域包络上存在严重的包络重叠与能量分布交叉, 模型往往难以在单一尺度下构建清晰的线性判别边界, 导致类间误判率居高不下。相比之下, 本文提出的改进模型凭借频率信息引导的动态卷积机制, 实现了对瞬态频率波动的自适应捕捉。这种机制允许模型在处理复杂吞咽生理信号时, 自动调节卷积核权重以聚焦于高维纹理特征, 从而将其与运动轨迹相似但频谱特征迥异的非吞咽动作进行精准解耦。

5. 结语

本研究针对人体吞咽行为识别任务中复杂信号特征提取难、深度模型参数冗余等问题, 构建了从高性能柔性传感硬件到创新时频域深度学习算法的端到端识别系统。首先, 在硬件层面, 本研究使用了自制的 PLLA 柔性压电传感器。传感器在颈部喉结处的成功应用, 证明了其捕获微弱体表生理压力信号的卓越能力, 为后续算法处理提供了高质量的原始数据支撑。其次, 在算法层面, 本研究提出了短时动态卷积模块(STDyConv)。该模块突破了传统卷积参数固定的局限, 利用信号的频域信息实现权重的动态调制, 并通过引入“可传递卷积核机制”解决了动态网络中专家核增多导致的参数激增问题。通过对不同尺度卷积核进行零填充对齐, 有效地融合了多尺度的动态表征, 实现了模型轻量化与判别能力的平衡。最后, 通过在自建吞咽数据集及 UCR 公共基准数据集上的广泛实验, 验证了 STDyConv 具有极强的即插即用特性。实验表明, 将其集成至 FCN 或 ResNet 等架构后, 模型的准确率与 Micro-F1 分数均获得 20% 显著增益, 且性能随集成深度的增加表现出明显的正向累积效应。

尽管本研究提出的参数传递机制与动态卷积架构在轻量化性能上表现卓越, 但在临床复杂环境下的鲁棒性仍存在潜在局限。参数共享逻辑虽然降低了冗余, 但在处理极端异构的病理吞咽波形时, 可能受限于基底张量的表征空间, 产生“特征表达饱和”风险。其次, 频率信息引导机制对高频噪声较为敏感, 若预处理阶段未能完全剔除环境伪影, 动态核的自适应调节可能引发伪特征增强, 影响跨个体场景下的泛化精度。此外, 为追求边缘端极低延迟而简化的通道交互, 在面对长时程、连续监测任务时, 是否存在上下文语义缺失导致的累积误差, 仍需在更大规模的临床随访中进一步验证。

综上所述, 本研究的研究成果不仅为高精度、非侵入式的吞咽功能监测提供了可靠的技术手段, 也为生理信号处理领域的动态特征建模提供了新的理论视角。未来的工作将聚焦于算法在低功耗嵌入式平台上的实时部署, 并探索多模态生理信号融合在更广泛临床场景中的应用潜力。

参考文献

- [1] 孙平秀. 科普小课堂: “安全吞咽, 守护健康” [C]//2025 年“《健康大湾区》-科普引领健康”论坛暨第 6 期健康科普作品征集活动. 2025: 1153-1155.
- [2] Hennessy, M. and Goldenberg, D. (2016) Surgical Anatomy and Physiology of Swallowing. *Operative Techniques in Otolaryngology—Head and Neck Surgery*, **27**, 60-66. <https://doi.org/10.1016/j.otot.2016.04.002>
- [3] Shaker, R., Belafsky, P.C., Postma, G.N., et al. (2013) Principles of Deglutition: A Multidisciplinary Text for Swallowing and Its Disorders. Vol. 19, Springer.
- [4] Ono, T., Hori, K., Masuda, Y. and Hayashi, T. (2009) Recent Advances in Sensing Oropharyngeal Swallowing Function in Japan. *Sensors*, **10**, 176-202. <https://doi.org/10.3390/s100100176>
- [5] 祝文月. 神经损伤引发吞咽困难与误吸的预防攻略[J]. 科技视界, 2025, 15(31): 18-19.
- [6] 肖金霞. 吞咽障碍不可忽视, 营养管理与功能训练正当时[J]. 健康必读, 2025(28): 52-53.
- [7] Boaden, E., Nightingale, J., Bradbury, C., Hives, L. and Georgiou, R. (2020) Clinical Practice Guidelines for Videofluoroscopic Swallowing Studies: A Systematic Review. *Radiography*, **26**, 154-162.

- <https://doi.org/10.1016/j.radi.2019.10.011>
- [8] Prikladnicki, A., Santana, M.G. and Cardoso, M.C. (2022) Protocols and Assessment Procedures in Fiberoptic Endoscopic Evaluation of Swallowing: An Updated Systematic Review. *Brazilian Journal of Otorhinolaryngology*, **88**, 445-470. <https://doi.org/10.1016/j.bjorl.2021.03.002>
- [9] Natta, L., Guido, F., Algieri, L., Mastronardi, V.M., Rizzi, F., Scarpa, E., *et al.* (2021) Conformable ALN Piezoelectric Sensors as a Non-Invasive Approach for Swallowing Disorder Assessment. *ACS Sensors*, **6**, 1761-1769. <https://doi.org/10.1021/acssensors.0c02339>
- [10] Lee, J., Chee, P., Lim, E. and Tan, C. (2021) Artificial Intelligence-Assisted Throat Sensor Using Ionic Polymer-Metal Composite (IPMC) Material. *Polymers*, **13**, Article No. 3041. <https://doi.org/10.3390/polym13183041>
- [11] Nguyen, D.T., Cohen, E., Pourhomayoun, M. and Alshurafa, N. (2017) Swallownet: Recurrent Neural Network Detects and Characterizes Eating Patterns. 2017 *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kona, 13-17 March 2017, 401-406. <https://doi.org/10.1109/percomw.2017.7917596>
- [12] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [13] 曹静. 科学护理, 让脑卒中患者吞咽无忧[J]. 医食参考, 2025(22): 55.
- [14] 韩思园, 王峥, 秦雯, 等. 画说老年患者吞咽功能康复锻炼[J]. 康复, 2025(33): 70-71.
- [15] Verelst, T. and Tuytelaars, T. (2020) Dynamic Convolutions: Exploiting Spatial Sparsity for Faster Inference. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 2320-2329. <https://doi.org/10.1109/cvpr42600.2020.00239>
- [16] Yang, B., Bender, G., Le, Q.V., *et al.* (2019) CondConv: Conditionally Parameterized Convolutions for Efficient Inference. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, 8-14 December 2019, 1307-1318.
- [17] Li, C., Zhou, A. and Yao, A. (2022) Omni-Dimensional Dynamic Convolution. *ICLR 2022*, 25-29 April 2022. <https://openreview.net/forum?id=DmpCfq6Mg39>
- [18] 刘璐瑶, 张森, 肖文栋. 基于小波分析和自相关计算的非接触式生理信号检测[J]. 工程科学学报, 2021, 43(9): 1206-1214.
- [19] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 1800-1807. <https://doi.org/10.1109/cvpr.2017.195>
- [20] He, Z., Liu, Y., Babichuk, I.S., Zhou, Z., Liu, Z., Yang, X., *et al.* (2024) Flexible Piezoelectric Sensors Based on Ionic Liquid-Doped Poly(l-lactic Acid) for Human Coughing Recognition. *Advanced Materials Technologies*, **9**, Article ID: 2400386. <https://doi.org/10.1002/admt.202400386>
- [21] Bagnall, A., Dau, H.A., Lines, J., *et al.* (2018) The UEA Multivariate Time Series Classification Archive, 2018. <https://doi.org/10.48550/arXiv.1811.00075>
- [22] Shokoohi-Yekta, M., Hu, B., Jin, H., Wang, J. and Keogh, E. (2017) Generalizing DTW to the Multi-Dimensional Case Requires an Adaptive Approach. *Data Mining and Knowledge Discovery*, **31**, 1-31. <https://doi.org/10.1007/s10618-016-0455-0>
- [23] Takahashi, K., Yamamoto, K., Kuchiba, A. and Koyama, T. (2022) Confidence Interval for Micro-Averaged F1 and Macro-Averaged F1 Scores. *Applied Intelligence*, **52**, 4961-4972. <https://doi.org/10.1007/s10489-021-02635-5>
- [24] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P. (2019) Deep Learning for Time Series Classification: A Review. *Data Mining and Knowledge Discovery*, **33**, 917-963. <https://doi.org/10.1007/s10618-019-00619-1>
- [25] Wang, Z., Yan, W. and Oates, T. (2017) Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. 2017 *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, 14-19 May 2017, 1578-1585. <https://doi.org/10.1109/ijcnn.2017.7966039>
- [26] Wu, H., Hu, T., Liu, Y., *et al.* (2022) TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. <https://doi.org/10.48550/arXiv.2210.02186>
- [27] Nie, Y., Nguyen, N.H., Sinthong, P., *et al.* (2022) A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. *ICLR 2023*, Kigali, 1-5 May 2023. <https://openreview.net/forum?id=Jbdc0vTOcol>
- [28] Oguiza, I. (2023) Tsai—A State-of-the-Art Deep Learning Library for Time Series and Sequential Data. <https://github.com/timeseriesAI/tsai>