PTS-SNN：一种用于高效语音情感识别的提示调优时间移位脉冲神经网络

Xun Su、Huamin Wang、Qi Zhang

arXiv:2602.08240v1 [cs.AI] 9 Feb 2026

摘要——**语音情感识别（SER）广泛应用于人机交互，但传统模型的高计算成本阻碍了其在资源受限的边缘设备上的部署。脉冲神经网络（SNN）凭借其事件驱动特性，提供了一种能效更高的替代方案；然而，将其与连续的自监督学习（SSL）表征相结合，面临着分布失配这一根本性挑战：动态范围较大的嵌入会降低基于阈值的神经元的信息编码能力。为解决这一问题，我们提出提示调优脉冲神经网络（PTS-SNN），这是一种参数高效的神经形态适配框架，可使冻结的自监督学习骨干网络与脉冲动态相匹配。具体而言，我们引入时间移位脉冲编码器，通过无参数的通道移位捕捉局部时间依赖，从而建立稳定的特征基底。为弥合领域差距，我们设计了一种上下文感知膜电位校准策略。该机制利用脉冲稀疏线性注意力模块，将全局语义上下文聚合至可学习的软提示中，以动态调节参数化漏积分发放（PLIF）神经元的偏置电压。这种调节能有效地将异质输入分布置于响应性发放范围内，从而缓解功能性静默或饱和。在五个多语言数据集（如 IEMOCAP、CASIA、EMODB）上的大量实验表明，PTS-SNN 在 IEMOCAP 上的准确率达到 73.34%，与具有竞争力的人工神经网络（ANN）相当，而可训练参数仅需 1.19M，每个样本的推理能耗仅为 0.35 mJ。**

索引词—**脉冲神经网络、语音情感识别、提示微调、能效、神经形态适应。**

I. 引言

SPEECH情感识别（SER）是人机交互（HCI）的基础，使具身智能体能够感知情感状态并理解人类意图[1]。尽管特征提取已从手工设计描述符 [2]发展到自监督学习（SSL）表征 [3],，最先进系统通常仍采用密集的人工神经网络（ANN）进行下游推理。

然而，这些架构会带来高昂的计算成本和内存开销 [4]，，限制了它们在资源受限的边缘设备上的部署 [5]。在此背景下，脉冲神经网络（SNNs）提供了一种能效更高的替代方案[6]。通过利用生物膜电位动力学以及稀疏的事件驱动计算，SNN 能够自然地建模时间演化，同时显著降低功耗 [7]。

尽管将连续表征学习与离散的事件驱动动力学相结合具有理论价值，但

Xun Su、Huamin Wang 和 Qi Zhang 就职于中国重庆 400715 的西南大学人工智能学院，同时也隶属于中国重庆 400715 的重庆类脑计算与智能芯片重点实验室（电子邮箱：swuring@email.swu.edu.cn；hmwang@swu.edu.cn；myhonor151@email.swu.edu.cn）

两者信息编码机制的固有异质性带来了显著的适配挑战 [6]。实证观察表明，将脉冲神经网络直接应用于自监督特征会导致性能下降 [8]。其根本障碍在于上游连续表征与下游脉冲神经元之间存在特征分布偏移 [7], [9]。具体而言，预训练嵌入向量通常具有较大的动态范围和稠密的流形分布，而漏积分发放（LIF） [10]神经元则充当对输入强度敏感的基于阈值的滤波器 [11]。若不进行针对性校准，高方差特征流会使神经元陷入两种极端状态：输入微弱导致功能性静默，或信号过强引发发放饱和 [12], , 如图1所示。这两种极端状态都会破坏有效特征整合所需的稀疏时间编码机制 [13]。

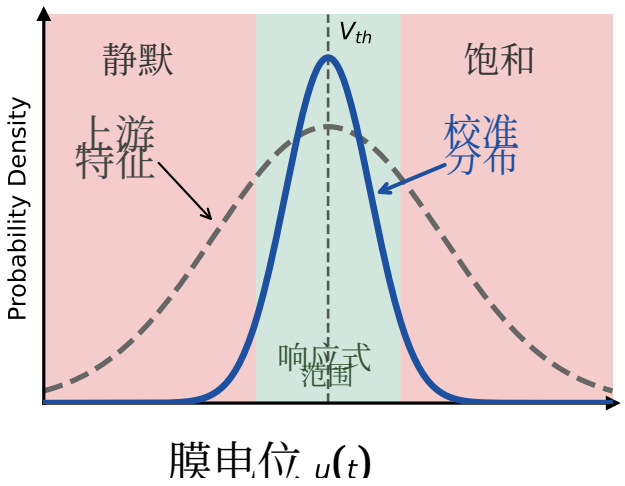


图 1：**分布失配示意图**。原始上游特征（灰色虚线）具有较大的动态范围，显著延伸至静默区或饱和区（红色区域），导致信息丢失。相比之下，所提出的 PTS-SNN 将这些特征重新对齐为校准分布（蓝色实线），使其集中于发放阈值 V_{th} 附近的响应范围（绿色区域）内，从而最大限度地提升脉冲神经元的时间编码容量。

为解决连续表征与离散动态之间的特征分布失配 [9], , 我们提出了 PTS-SNN，这是一种神经形态适配框架。我们的方法不修改庞大的预训练骨干网络 [14], , 而是通过两阶段校准过程，使冻结的连续特征与离散脉冲动态相对齐。首先，我们引入时间移位脉冲编码器，利用无参数的通道移位来捕获局部时间依赖关系 [15]。这能稳定原始特征流形中的高频波动

，从而提供稳健的输入基础。尤为重要的是，为解决阈值敏感性问题，我们设计了一种基于脉冲稀疏线性注意力的上下文感知膜电位校准策略 [16]。该机制将全局语义上下文聚合到可学习的软提示中 [17], [18], ，再由软提示生成动态偏置电压，主动调节参数化漏积分发放（PLIF）神经元的基线电位 [19]。这种受生物启发的调节能够有效地将异质输入分布置于响应性发放范围内，从而降低功能性静默或饱和的风险。本文的贡献可总结如下：

(1) 我们提出了 PTS-SNN，这是一种参数高效的类神经形态适配框架，旨在解决连续 SSL 表示与离散脉冲动态之间的特征分布失配。该架构将时间移位脉冲编码器与脉冲稀疏线性注意力相结合，在冻结上游骨干网络的同时，高效构建稳健的特征空间。

(2) 我们提出了一种受生物稳态调节启发的上下文感知膜电位校准策略。与传统语义标记不同，我们可学习的软提示充当主动调节器，用于生成动态偏置电压。该机制使异构特征分布与 PLIF 神经元的敏感响应范围相匹配，从而有效避免神经元停止放电或进入饱和状态。

(3) 我们在五个多语言数据集（IEMOCAP、CASIA、EMODB、EMOVO 和 URDU）上验证了该模型。实验结果表明，PTS-SNN 的准确率具有竞争力（在 IEMOCAP 上达到 73.34%），可媲美最先进的人工神经网络；与此同时，它仅需 1.19 M 个参数，且每个样本的推理能耗为 0.35 mJ，验证了其在能耗受限的边缘智能领域的应用潜力。

本文结构如下。第二节回顾了自监督学习和脉冲神经网络的相关研究。第三节阐述了分布失配问题，并详细介绍所提出的 PTS-SNN 框架，重点关注时间移位编码器和上下文感知的脉冲校准。第四节评估模型在五个多语言基准数据集上的性能和能效，第五节则总结全文。

II. 相关工作

A. 语音情感识别中的自监督学习

早期的语音情感识别主要依赖梅尔频率倒谱系数和扩展大五声学参数集等手工设计的声学特征 [2]。尽管计算效率较高，这些静态描述符难以捕捉复杂情感表达中固有的层级时间依赖关系 [1]。近年来，自监督学习已成为特征提取的主流范式 [20]。Wav2Vec 2.0 [21] 等预训练框架以及领域专用的 emotion2vec[22] 利用大规模无标注数据学习稳健的上下文表征，与传统基线相比展现出更强的泛化能力 [23]。

然而，将这些模型适配到下游任务通常需要进行全量微调，或附加密集型 ANN解码器 [24]。这些策略会造成参数冗余度高、推理延迟长的问题，从而阻碍其部署于资源受限的边缘设备 [25]。更关键的是，人工神经网络固有的密集型浮点矩阵乘法，与神经形态计算的异步稀疏范式根本不兼容 [26]–[28]。因此，如何制定轻量级适应策略，使冻结的自监督学习特征与脉冲动态相契合，仍是一项关键的研究挑战 [29], [30]。

B. 用于音频处理的脉冲神经网络

脉冲神经网络采用事件驱动范式以尽可能降低计算能耗，因此天然适用于音频处理 [31], [32]。近期的进展已将这些架构扩展至关键词检测 [33]和大词汇量连续语音识别 [4],[34]等复杂任务。在语音情感识别领域，研究者开发了混合架构，通常采用卷积神经网络（CNN）作为前端，为脉冲后端提取特征 [35], [36]。这些方法利用脉冲时序动力学捕捉情感韵律 [37] ，同时保持较低的计算开销。

然而，现有基于脉冲神经网络的情感识别模型仍受限于其输入表征，主要依赖梅尔频率倒谱系数或频谱图等低层声学特征 [38], [39]。尽管混合卷积神经网络-脉冲神经网络架构在挖掘局部模式方面表现有效，在标准六分类基准上取得了约 65.3% 的准确率 [40],，与传统人工神经网络相当，但其本质上受限于浅层声学描述符固有的语义上限 [1]。因此，这些模型无法捕捉先进人工神经网络系统通过自监督学习提取的深层、高级情感线索 [22],[23]。因此，核心挑战并非网络的拟合能力，而是如何有效融合深度连续表征与离散的神经形态动力学 [7]。主要障碍在于预训练流形与脉冲阈值之间的分布异质性，因此需要有针对性的适应策略，才能实现高性能、能效的识别 [41]。

C. 提示学习与适应技术

提示学习最初为自然语言处理（NLP） [42]，而开发，如今已成为参数高效微调（PEFT） [43]的主流策略。通过优化少量可训练向量，同时冻结预训练骨干网络 [44],，这一范式能够以极低的计算开销有效地将基础模型适配到下游任务 [14]。近期研究已成功将这一方法扩展至计算机视觉 [45]和音频处理 [46],，验证了其在缓解跨模态分布偏移方面的有效性 [46], [47]。

尽管取得了这些成果，提示微调在神经形态计算中的应用仍然有限。尽管近期研究致力于增强脉冲神经网络的表征能力[48], [49],现有的优化策略主要聚焦于替代梯度学习 [50] 或 ANN到SNN转换

[51]。虽然混合架构试图弥合这一差距，但往往受限于浅层声学输入，导致语义深度不足 [52], [53]。因此，针对预训练连续表征的高效适应协议仍然缺失 [54]。根本障碍在于机制上的不兼容：标准提示是连续加性向量 [43]，，而脉冲神经网络的发放动态则是离散且基于阈值的 [55]。将连续提示直接注入二值脉冲序列会破坏时间编码所需的精确时序 [56],，从而导致信息丢失。

为解决这一问题，我们提出了一种专为脉冲神经网络设计的提示机制生物物理重构。不同于人工神经网络中使用的语义令牌，我们将提示建模为类似于生物控制机制的稳态调节信号 [57]。通过将可学习向量映射为动态偏置电压，该方法主动校准基线膜电位[17]。这一策略无需重构底层网络，即可精确对齐连续上游特征与离散神经元响应范围 [54]，，从而为特征分布不匹配问题提供一种参数高效的解决方案。

III. 方法

A. 问题 *ϕορмαλεрц*化

任务定义。令 $\mathcal{D} = \{(\mathbf{z}_i, y_i)\}_{i=1}^{N}$ 表示标注数据集，其中 $\mathbf{z}_i \in \mathcal{Z}$ 表示原始语音波形， $y_i \in \mathcal{Y} = \{1, \ldots, K\}$ 表示对应的类别情感标签。我们采用预训练自监督学习模型（emotion2vec）作为特征提取器，其映射函数记为f_{ssl}：$\mathcal{Z} \to \mathbb{R}^{T \times D_{in}}$。为保持参数效率，骨干网络参数 Θ_{ssl} 被冻结。因此，对于每个样本$\mathbf{z}$，我们得到一个确定性的连续表示序列$\mathbf{X} \in \mathbb{R}^{T \times D_{in}}$，其中 T 为序列长度，D_{in} 为通道维度。

神经元动力学。下游分类器使用PLIF 神经元构建。与具有固定超参数的标准LIF模型不同，PLIF 神经元引入可学习膜时间常数，以自适应地建模时间依赖关系。令 $u_j^{(l)}[t]$ 和 $s_j^{(l)}[t]$ 分别表示第 l 层中第 j个神经元在时间步 t 的膜电位和脉冲输出。离散时间动力学由以下差分方程控制：

$$u_j^{(l)}[t] = \tau_j^{(l)} \cdot u_j^{(l)}[t-1] + i_j^{(l)}[t] - V_{th} \cdot s_j^{(l)}[t-1] \quad (1)$$

其中 $i_j^{(l)}[t]$ 为突触输入电流， V_{th} 为发放阈值， $\tau_j^{(l)} \in$（0，1）表示可学习衰减因子。脉冲生成遵循Heaviside阶跃函数$\Theta(\cdot)$：

$$s_j^{(l)}[t] = \Theta(u_j^{(l)}[t] - V_{th}) \quad \text{s.t.} \quad s_j^{(l)}[t] \in \{0, 1\} \quad (2)$$

分布失配。 $\mathbf{X}$ 的连续流形与公式 (2) 中离散的切换动态之间存在根本性的兼容性障碍。预训练特征通常具有较大的动态范围，导致分布相对于固定阈值 V_{th} 发生偏移。我们将突触电流 $\mathbf{I}$ 的这种情况归纳为两种病态状态：

$$\begin{cases} \mathbb{E}_t[\mathbf{I}[t]] \ll V_{th} \implies \sum_{t=1}^{T} s[t] \to 0 \\ \mathbb{E}_t[\mathbf{I}[t]] \gg V_{th} \implies \sum_{t=1}^{T} s[t] \to T \end{cases} \quad (3)$$

这两种极端情况都会使脉冲序列 $H(\mathbf{S})$的熵最小化，从而降低脉冲神经网络的时间编码容量。

学习目标。为解决分布失配，我们提出一种可学习的神经形态适配器 $\mathcal{G}_\theta$ ： $\mathbb{R}^{T \times D_{in}} \to \mathbb{R}^{T \times D_{out}}$。其目标是将连续输入 X 映射到与PLIF分类器 $\mathcal{F}_\phi$的感受野对齐的校准潜在空间。预测的类别概率分布$\hat{}y$ 通过对输出神经元在 T 个时间步上的平均发放率进行解码得到：

$$\hat{y} = \text{Softmax}\left(\frac{1}{T} \sum_{t=1}^{T} \mathcal{F}\phi(\mathcal{G}\theta(X)[t]) \right) \quad (4)$$

我们的目标是联合优化适配器参数 θ和SNN权重 ϕ ，以最小化$\hat{}y$ 与真实标签 y之间的差异。针对类别不平衡和标签噪声的具体损失函数设计详见第III.F节。

B. 总体架构

如图2所示，我们提出了PTS-SNN，这是一种参数高效框架，旨在协调连续的SSL表示与离散的脉冲动态。通过冻结预训练的emotion2vec骨干网络，该架构仅优化一个轻量级适配器，并通过三个级联阶段将静态特征转换为动态脉冲序列。

首先，时间移位脉冲编码器在残差脉冲块中采用无参数的通道移位操作。该模块能够捕获局部时间依赖关系，并平滑特征流形中的高频波动，从而建立稳定的输入基。

随后，脉冲稀疏线性注意力（SSLA）模块将可学习的软提示与特征序列相结合。SSLA通过线性交互将全局语义上下文聚合到提示向量中，同时利用神经元阈值过滤背景冗余。

最后，富含上下文信息的提示会驱动稳态偏置生成器。受生物调节机制启发，该组件会生成动态电压偏置，以移动基线电位，主动将异质输入分布调整至 PLIF 后端的敏感响应范围，从而实现有效的脉冲解码。

C. 时间移位脉冲编码器

从冻结骨干网络提取的原始特征序列通常表现出帧级独立性和高幅值离群点。这些特征缺乏对短时韵律的显式建模，并可能导致下游神经元的发放动态不稳定。为缓解这些问题，我们构建了时间移位脉冲编码器。该模块旨在

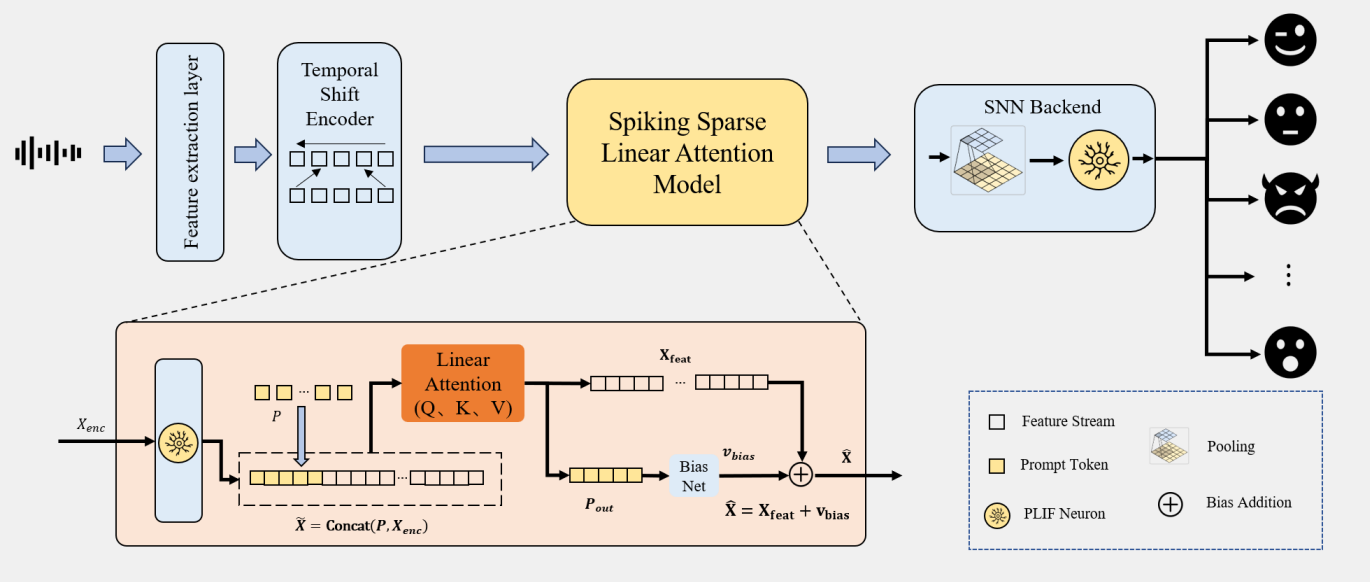


图 2：**所提出的 PTS-SNN 框架概览**。该架构对冻结骨干网络提取的特征进行处理，依次经过三个级联阶段：通过无参数移位捕捉局部依赖关系的时间移位脉冲编码器、用于聚合全局语义上下文的 SSLA 模块，以及用于情绪分类的 SNN 后端。下方展示了 SSLA 的工作机制：可学习的软提示与稀疏脉冲图交互，生成动态电压偏置，以校准膜电位。

通过软饱和预处理和高效时序移位这两个关键阶段，平滑输入流形并提取紧凑的局部上下文。

软饱和预处理。为防止数值不稳定，我们首先限制输入特征 $X \in \mathbb{R}^{T\times D}$ 的动态范围。我们设计了一种软饱和函数，将层归一化与缩放后的双曲正切变换相结合：

$$X_{pre} = \tanh(\xi \cdot \mathrm{LayerNorm}(X)) \tag{5}$$

我们通过经验将缩放因子 ξ 设为 0.1。与硬裁剪不同，这种可微操作能够平滑地将极端离群值压缩到线性区域。它确保输入电流处于适合神经形态计算的稳定范围，同时保留反向传播所需的梯度流。

脉冲时序移位块。预处理后的序列 X_{pre} 进入一个由两个级联残差脉冲块组成的编码网络。该网络逐步降低特征维度（768 → 512 → 256），以过滤冗余信息。在每个块中，我们以无参数的时间移位操作取代标准卷积核。通过沿时间轴移动一部分特征通道，该机制无需浮点运算（FLOPs）即可促进相邻帧之间的信息交换，如图 3 所示。第 l 个块的计算流程形式化如下：

$$H^{(l)} = \mathrm{PLIF}(\mathrm{BN}(W^{(l)}(\mathrm{Shift}(H^{(l-1)})) + b^{(l)})) \tag{6}$$

其中，$W^{(l)}$ 和 $b^{(l)}$ 分别表示线性投影的权重和偏置。Shift 操作显式地捕捉局部时间依赖关系，为 PLIF 神经元提供具有上下文感知能力的输入电流。PLIF 神经元的可学习时间常数进一步优化了这些

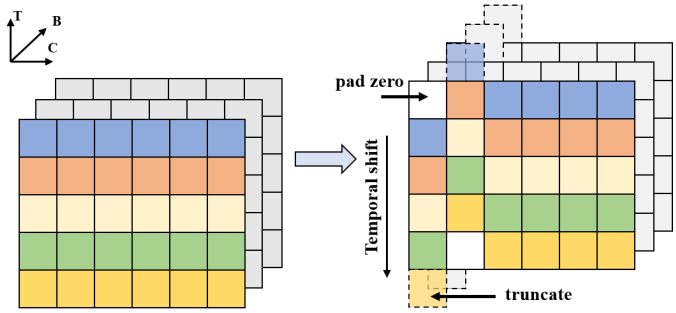


图 3：**时间移位操作**。**左：**原始输入特征图，其维度包括时间（T）、批次（B）和通道（C）。**右：**沿时间轴移动特定通道后的变换特征图。顶部的空位以零填充，底部溢出的部分则被截断，以保持序列长度，从而在不增加额外参数的情况下促进局部信息交换。

局部韵律线索。最终，该编码器输出稳健的低维特征序列 X_{conv}，为后续的全局注意力机制奠定稳定基础。

D. 脉冲稀疏线性注意力

尽管时间移位脉冲编码器能够有效提取局部韵律特征，但要理解复杂的情绪状态，还需要显式建模长程全局依赖。为高效实现这一目标，我们设计了 SSLA 模块，其详细内部结构如图2下半部分所示。该组件将基于提示的序列增强与线性化注意力机制相结合，在聚合全局上下文的同时尽可能降低计算复杂度。

我们首先引入一组可学习提示向量$\mathbf{P} \in \mathbb{R}^{L_p\times D}$，其中 L_p 表示

提示词令牌的数量，这些令牌用作上下文查询。这些提示沿编码器输出$\mathbf{X}_{enc} \in \mathbb{R}^{T\times D}$ 的时间维度拼接，构成扩展后的联合序列$\tilde{\mathbf{X}}$：

$$\tilde{\mathbf{X}} = \text{Concat}(\mathbf{P}, \mathbf{X}_{enc}) \in \mathbb{R}^{(L_p+T)\times D} \tag{7}$$

此操作确保提示在物理上嵌入特征空间，使其能够直接参与后续的特征交互与上下文聚合。

$\tilde{}$为过滤背景冗余，联合序列$\mathbf{X}$ 由单步PLIF神经元层处理。该层充当基于阈值的滤波器，将连续特征转换为稀疏二值脉冲矩阵$\mathbf{S}$。因此，仅保留强度超过发放阈值的显著特征，同时将低幅噪声严格抑制为零：

$$\mathbf{S} = \text{PLIF}(\tilde{\mathbf{X}}) \in 0, 1^{(L_p+T)\times D} \tag{8}$$

这种二值化显著降低了数据密度，从而有效减少下游矩阵运算的计算成本。

随后，通过无偏置线性层将稀疏矩阵$\mathbf{S}$ 投影为查询（$\mathbf{Q}$）、键（$\mathbf{K}$）和值（$\mathbf{V}$）张量。为规避标准Soft max注意力固有的二次复杂度$O(L^2)$，我们利用矩阵乘法的结合律采用线性化公式，并优先计算全局上下文项$\mathbf{K}^T\mathbf{V}$。注意力输出正式定义为：

$$\text{SSLA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{Q} \cdot (\sigma(\mathbf{K})^\top \cdot \mathbf{V}) \tag{9}$$

其中，$\sigma(\cdot)$ 表示按列归一化函数。首先计算 $D \times D$ 上下文矩阵 $\mathbf{K}^T\mathbf{V}$ ，将计算复杂度降低至 $O((L_p+T)D^2)$，使其随序列长度线性增长。通过这一机制，$\mathbf{Q}$中的提示向量能够有效聚合全局上下文中的情感相关语义，从而为后续校准阶段生成紧凑表示。

E. 上下文感知膜电位校准

在 SSLA 阶段之后，输出提示序列$\mathbf{P}_{out}$ 封装了聚合后的情感上下文。然而，这些潜在表示无法直接纠正脉冲后端中的分布失配。为弥合这一差距，我们引入稳态偏置生成器，将抽象上下文转化为物理电压调节信号。该机制会动态调整特征基线，使输入分布与 PLIF 神经元的敏感发放范围对齐。

校准过程始于提取紧凑的上下文描述符 $\mathbf{c} \in \mathbb{R}^D$。通过从联合输出序列中分离出更新后的提示词令牌，我们沿时间轴执行全局平均池化：

$$\mathbf{c} = \frac{1}{L_p}\sum_{i=1}^{L_p} \mathbf{P}_{out}[i] \tag{10}$$

这一操作有效地将序列级语义信息压缩为统一的全局状态向量，从而全面表征情感基调。

随后，将描述符 $\mathbf{c}$ 输入瓶颈式多层感知机（MLP），以生成动态偏置向量$\mathbf{v}_{bias} \in \mathbb{R}^D$。为尽量减少参数开销，同时确保非线性拟合能力，生成器采用“先压缩、再扩展”的架构。映射函数表示为：

$$\mathbf{v}_{bias} = \kappa \cdot \tanh\left(\mathbf{W}_{up} \cdot \text{ReLU}\left(\mathbf{W}_{down}\mathbf{c} + \boldsymbol{\beta}_1\right) + \boldsymbol{\beta}_2\right) \tag{11}$$

其中，$\mathbf{W}_{down} \in \mathbb{R}^{D_h\times D}$ 和$\mathbf{W}_{up} \in \mathbb{R}^{D\times D_h}$ 分别表示降维权重和扩展权重，其中 $D_h < D$。经过因子 v_{bias}缩放的 tanh 激活函数可生成双向偏置，使模型能够进行兴奋性或抑制性调节。

为确保优化稳定性，我们严格约束生成器的初始状态。我们对输出层参数采用零初始化策略：

$$\mathbf{W}_{up} \leftarrow \mathbf{0}, \quad \beta_2 \leftarrow \mathbf{0} \tag{12}$$

因此，在训练开始时，生成的偏置 $\mathbf{b}$初始化为零向量。这确保模型以恒等映射的形式启动，避免随机初始化造成的分布冲击，并促进其平滑过渡到自适应状态。

最后，生成的偏置 $\mathbf{v}_{bias}$ 会进行广播，并注入从 SSLA ˆ输出得到的原始特征流$\mathbf{X}^{feat}$ 中。校准后的输入$\mathbf{X}$ 定义如下：

$$\hat{\mathbf{X}}[t] = \mathbf{X}_{feat}[t] + \mathbf{v}_{bias} \tag{13}$$

从生物物理学视角来看，该操作将依赖于上下文的背景电流 I_{bias} =$\mathbf{v}_{bias}$ 引入 PLIF 动力学中。这种主动调节能够自适应地将输入膜电位的中心调整至发放阈值 V_{th}附近，有效防止跨域分布偏移引起的功能性静默或饱和。

F. 优化目标

为实现端到端训练，同时应对语音情感识别数据集中固有的类别不平衡和标签噪声，我们采用混合损失函数。该目标函数将焦点损失（用于挖掘困难样本）与标签平滑交叉熵（用于防止对噪声标注过拟合）进行线性组合。令 $\mathbf{o} \in \mathbb{R}^K$ 表示 SNN 后端的输出向量（即时间步 T上的平均发放率），其中 K 为情感类别的数量。对该输出应用 Softmax 函数即可得到预测概率分布$\mathbf{p}$：

$$p_k = \frac{\exp(o_k)}{\sum_{j=1}^{K}\exp(o_j)}, \quad k \in \{1, \ldots, K\} \tag{14}$$

总优化目标 $\mathcal{L}$ 表示为焦点损失项（$\mathcal{L}_{FL}$）和标签平滑项（$\mathcal{L}_{LS}$）的加权和：

$$\mathcal{L} = \lambda_1\mathcal{L}_{FL} + \lambda_2\mathcal{L}_{LS} \tag{15}$$

其中， λ_1 和 λ_2 是用于平衡两个损失分量的超参数。

焦点损失项：为减轻类别不平衡的影响，并使模型关注难以分类的样本，我们使用真实硬标签$y \in \{1, \ldots, K\}$计算焦点损失：

$$\mathcal{L}_{FL} = -\alpha(1 - p_y)^{\gamma} \log(p_y) \tag{16}$$

标签平滑项：为增强泛化能力并处理标签歧义，我们构造平滑目标分布$\mathbf{y}^{ls}$。对于目标类别 y，平滑标签向量定义如下：

$$y_k^{ls} = \begin{cases} 1 - \epsilon + \frac{\epsilon}{K}, & \text{if } k = y \\ \frac{\epsilon}{K}, & \text{otherwise} \end{cases} \tag{17}$$

其中， ϵ 是平滑参数。标签平滑损失通过计算平滑目标与预测概率之间的交叉熵得到：

$$\mathcal{L}_{LS} = -\sum_{k=1}^{K} y_k^{ls} \log(p_k) \tag{18}$$

根据经验调优，我们将聚焦参数设为 $\alpha = 0.8$ 和 $\gamma = 2.5$，平滑因子设为 $\epsilon = 0.12$。损失分量采用 $\lambda_1 = 0.7$ 和 $\lambda_2 = 0.3$加权，以优先挖掘困难样本，同时保持正则化。

IV. 实验

本节对所提出的PTS-SNN 框架在五个多语言基准数据集上的表现进行全面的实证评估。我们从分类准确率、参数效率和神经形态能耗等方面分析模型的有效性，并与现有方法进行详细比较。

A. 实验设置

我们在五个多语言基准数据集上评估所提出的框架：IEMOCAP、CASIA、EMODB、EMOVO 和 URDU。表 I 总结了语言、说话人统计信息和情绪类别的详细统计数据。遵循 emotion2vec 中的方案，我们对 IEMOCAP 数据集采用留一会话法五折交叉验证策略。对于 CASIA，我们采用随机留一说话人法四折交叉验证。对于其余数据集（EMODB、EMOVO 和 URDU），则采用随机留一法十折交叉验证方案，以确保评估的严谨性。

为评估分类性能，我们采用加权准确率（WA）和非加权准确率（UA）。WA 衡量所有样本的总体准确率，反映整体性能。UA 计算各类别准确率的算术平均值，可为类别不平衡的数据集提供公平的评估。

所有模型均使用 PyTorch 框架（Python 3.10.14）在一台配备单块英伟达 GeForce RTX 4090 图形处理器（24GB 显存）的工作站上实现。优化采用 AdamW，初始学习率为 2×10^{-4}，权重衰减为 0.05。学习率通过余弦退火热重启调度器动态调整，其初始重启周期（T_0）设为 10 个训练轮次，乘法因子（T_{mult}）设为 2，并逐渐衰减至最小值 1×10^{-6}。训练共进行 200 个训练轮次，批次大小为 32。对于 SNN 后端，采用陡峭度因子为 $\alpha_{sg} = 5$ 的替代梯度法，以实现反向传播。

表 I：本研究使用的多语言语音情绪识别数据集概述。

数据集	语言	说话人	话语	情感
IEMOCAP	英语	10	5,531	4
CASIA	中文	4	1,200	4
EMODB	德语	10	535	7
EMOVO	意大利语	6	588	7
URDU	Urdu	38	400	4

B. 主要结果

神经形态适应的有效性。我们通过比较 PTS-SNN 与 emotion2vec 骨干网络及直接脉冲基线，评估该适应策略。如表 IV 所示，原始脉冲神经网络将连续特征直接投影到脉冲神经元上，加权准确率（WA）为 41.78%。性能下降表明预训练特征与脉冲域之间存在分布偏移。相比之下，PTS-SNN 达到 73.34%，超过了原始 emotion2vec 骨干网络（71.79%）。这一结果表明，所提出的基于提示的校准能够使连续表征与脉冲后端对齐，且不会造成性能下降。

与最先进方法的比较。表 II 展示了 PTS-SNN 与近期语音情感识别架构在 IEMOCAP 数据集上的比较。基线方法包括基于注意力的模型（Co-attention、ENT）和基于移位的网络（ShiftFormer）。对于 MSTR，我们报告了基于官方实现复现所得的结果。在分类准确率方面，PTS-SNN（73.34%）优于 MSTR（70.6%）和 ShiftFormer（72.1%），与 ENT（72.43%）和 DST（71.8%）表现相当。此外，如表中所示，与这些基于人工神经网络的基线方法相比，PTS-SNN 的参数量更少，推理能耗更低。下一节将对计算效率进行详细分析。

跨语言泛化。为验证所提方法的普适性，我们将评估扩展至四个非英语数据集：CASIA、EMODB、EMOVO和URDU。结果汇总于表 III。PTS-SNN在所有测试语言上均持续优于emotion2vec基线。在德语 EMODB数据集上，准确率从84.34%提升至87.49%。在 CASIA（+3.%）和URDU（+2.75%）上也观察到显著提升，

表 II：在 IEMOCAP 数据集上与最先进方法的比较。 †：使用官方代码实现复现的结果。

模型	核心机制	WA (%)	UA (%)	参数量 (M)	能量（毫焦）
协同注意力 [58]	交叉注意力	69.80	71.05	175.30	103.40
MSTR† [59]	多尺度 Transformer	70.60	71.60	30.03	51.06
DST [60]	双流融合	71.80	73.60	22.78	35.34
ShiftFormer [61]	稀疏移位	72.10	72.70	9.50	16.20
ENT [62]	高效Transformer	72.43	73.88	8.55	6.35
PTS-SNN（我们的方法）	**脉冲提示**	**73.34**	**73.72**	**1.19**	**0.35**

表明学习得到的提示捕捉的是与语言无关的情感韵律，而非语言内容。即便在资源有限的意大利语EMOVO数据集上，我们的方法仍保持性能优势，进一步证明了其在多样声学环境中的稳健性。

表 III：CASIA、EMODB、EMOVO 和 URDU 数据集上的跨语言性能比较（WA %）。

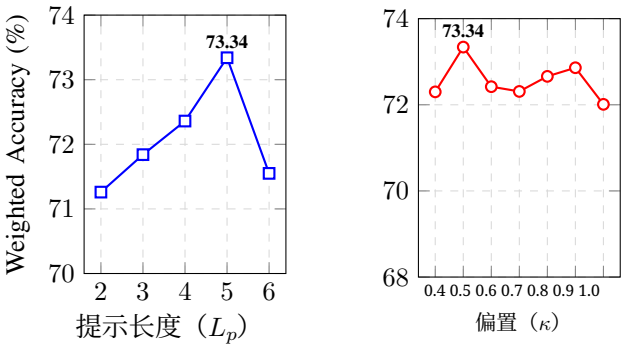
数据集	语言	emotion2vec	PTS-SNN（本文方法）
CASIA	中文	69.20	**72.50**
EMODB	德语	84.34	**87.49**
EMOVO	意大利语	61.21	**61.94**
URDU	Urdu	81.50	**84.25**

表 IV：在 IEMOCAP 数据集上不同适应策略的性能比较。

模型	适应策略	WA (%)
emotion2vec [22]	标准人工神经网络微调	71.79
原始脉冲神经网络	直接脉冲转换	41.78
PTS-SNN（本文提出）	神经形态提示	**73.34**

表V：IEMO-CAP 数据集上关键组件的消融研究。“完整模型”表示所提出的 PTS-SNN。

配置	WA (%)	UA (%)
完整模型 (PTS-SNN)	**73.34**	**73.72**
不使用提示微调	71.76	71.90
不含注意力模块	68.35	69.10
仅时间移位	61.27	62.05



（a）提示长度的影响　　(b) 偏置参数的影响

图4：参数敏感性分析。**(a)** 不同提示长度下的准确率 L_p。**(b)** 偏置参数变化对准确率的影响 κ 。

C. 消融实验

为验证各组件的贡献并分析超参数敏感性，我们在 IEMOCAP 数据集上开展了一系列实验。组件消融的定量结果汇总于表V。我们首先通过逐一移除模块来评估所提出模块的影响。移除注意力模块后，性能从 73.34% 降至 68.35%，表明全局上下文聚合对于区分复杂情感状态至关重要。同样，将可学习提示替换为标准微调后，准确率降至 71.76%，这证实了所提出的提示策略比直接更新权重更能有效地适配冻结骨干网络。仅保留时间移位机制时，准确率进一步降至 61.27%，这表明尽管时间动态是基础，但若缺少高级特征调制，仍不足以取得良好效果。

我们进一步研究了模型对提示长度（L_p）以及漏积分发放神经元中的偏置参数（b）的敏感性。

如图 4(a) 所示，随着 L_p 增加，加权准确率起初有所提升，并在 $L_p = 5$ 时达到峰值 73.34%。然而，将长度扩展到 6 会导致性能下降（71.55%），这表明过多的提示词令牌可能会引入冗余。关于偏置参数，图 4(b) 展示了在 $[0.4, 1.0]$ 范围内的性能变化。准确率在 $b = 0.5$ 时达到最佳。偏离这一最优点的值会导致性能下降；具体而言，较低的偏置值可能无法充分激活神经元，而较高的值则容易引入噪声，从而损害脉冲表示的信噪比。

D. 能量消耗

参数效率。我们首先从可训练参数的角度评估模型复杂度。如图5(a)所示，PTS-SNN包含119万个参数。参数量较少得益于基于适配器的设计：预训练骨干网络保持冻结，仅更新稀疏脉冲模块。相比之下，全量微调基线

例如 MSTR 和协同注意力分别需要 30.03 M 和 175.30 M 个参数。即使与轻量级 ShiftFormer（9.50 M）相比，我们的方法也将参数量减少了约 87%，便于部署在内存受限的边缘设备上。

能量消耗。为量化推理成本，我们采用基于 45 nm CMOS 技术的标准神经形态协议。能量估算区分了人工神经网络（ANN）中的乘累加（MAC）操作（4.6 pJ）与脉冲神经网络（SNN）中的累加（AC）操作（0.9 pJ） [63]。三种不同架构范式的计算方法定义如下：

$$EC_{ANN} = MACs^{ANN} \times 4.6 \tag{19}$$

$$EC_{pure} = Synapsed^{SNN}_{activated} \times 0.9 \tag{20}$$

$$\begin{aligned} EC_{non-pure} = Synapsed^{SNN}_{activated} \times 0.9 \\ + MACs^{ANN} \times 4.6 \end{aligned} \tag{21}$$

基于这些公式，能量比较结果如图5(b)所示。纯人工神经网络架构的 MSTR 模型由于密集矩阵计算，每个样本消耗 51.06 mJ。类似地，ShiftFormer 需要 16.20 mJ。相比之下，本文提出的 PTS-SNN 仅消耗 0.35 mJ。这一显著降低归因于脉冲动态的稀疏性，验证了该方法适用于功率受限的应用场景。

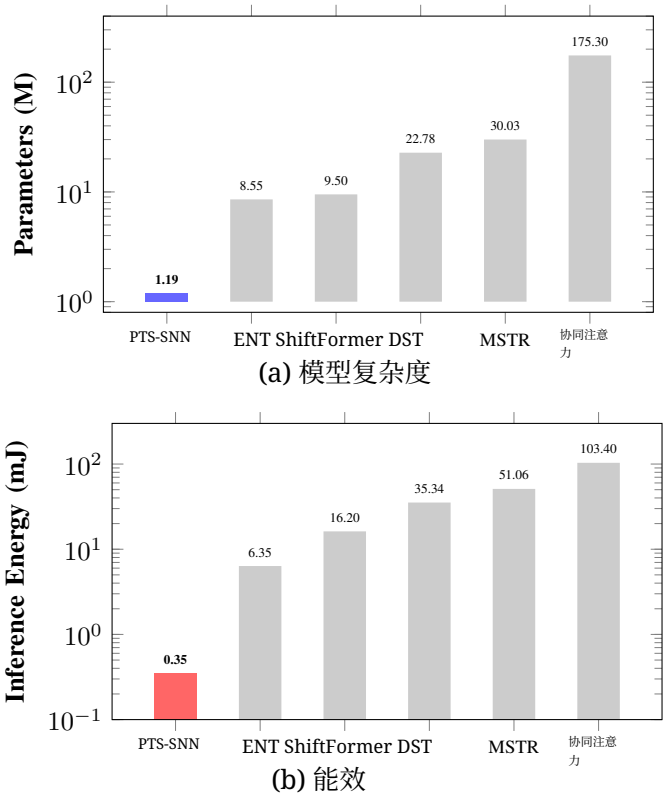


图5：与最先进基线方法的计算效率比较。**(a)** 可训练参数数量（对数刻度）。**(b)** 每个样本的推理能耗（对数刻度）。所提出的**PTS-SNN** 在这两项指标上均实现了数量级的提升。

五、结论

本文提出了 PTS-SNN，这是一种参数高效的神经形态适配框架，旨在使冻结的自监督学习（SSL）骨干网络与离散脉冲动态相匹配，以实现能效语音情感识别。为有效解决通常会削弱神经元编码能力的分布异质性，该架构整合了时间移位脉冲编码器，通过无参数通道移位捕捉局部时间依赖关系，并结合上下文感知膜电位校准策略。该机制利用脉冲稀疏线性注意力，将全局语义上下文聚合至可学习的软提示中，生成动态偏置电压，以调节参数化漏积分发放（PLIF）神经元，使输入分布有效地集中于敏感发放范围内，从而缓解功能性静默或饱和。在五个多语言数据集（IEMOCAP、CASIA、EMODB、EMOVO 和 URDU）上的实证评估表明，PTS-SNN 在 IEMOCAP 上取得了 73.34% 的加权准确率，与最先进的人工神经网络相当，同时每个样本仅需 119 万个参数和 0.35 mJ 的推理能耗。然而，当前框架仅依赖声学信号，忽略了人类交流中面部表情和文本内容所蕴含的丰富互补信息。未来工作将着重把这种神经形态提示机制扩展到多模态脉冲架构，旨在融合视听表征，从而在复杂的真实环境中实现更稳健的情感理解。

参考文献

[1] Y. Gao、L. Wang、J. Liu、J. Dang 和 S. Okada，“用于跨语料库语音情感识别的对抗域泛化变换器”，*IEEE*情感计算汇刊，第 15 卷，708，2023 年，第 2 期，页码 – 697。[2] F. Eyben、K. R. Scherer、B. W. Schuller、J. Sundberg、E. André、C. Busso、L. Y. Devillers、J. Epps、P. Laukka、S. S. Narayanan 等，“用于语音研究与情感计算的日内瓦极简声学参数集（GeMAPS）”，*IEEE*情感计算汇刊，第 7 卷，第 2 期，页码 190–202，2015 年。[3] S. Chen、C. Wang、Z. Chen、Y. Wu、S. Liu、Z. Chen、J. Li、N. Kanda、T. Yoshioka、X. Xiao 等，“WavLM：用于全栈语音处理的大规模自监督预训练”，*IEEE*信号处理选题期刊，第 16 卷，第 6 期，页码 1505–1518，2022 年。[4] J. Wu、E. Yılmaz、M. Zhang、H. Li 和 K. C. Tan，“用于大词汇量自动语音识别的深度脉冲神经网络”，《神经科学前沿》，第 14 卷，文章编号 199，2020 年。[5] C.-K. Lu、C.-W. Lu 和 G. B. Lin，“面向陪伴机器人的轻量级多模态情感识别：融合面部与语音特征的深度学习框架”，*IEEE*情感计算汇刊，2025 年。[6] Z. Zhou、Y. Zhu、C. He、Y. Wang、S. YAN、Y. Tian 和 L. Yuan，“Spikformer：脉冲神经网络与变换器的结合”，载于第十一届国际学习表征会议，2023 年。[7] M. Yao、J. Hu、Z. Zhou、L. Yuan、Y. Tian、B. Xu 和 G. Li，“脉冲驱动的变换器”，《神经信息处理系统进展》，第 36 卷，页码 64 043–64 058，2023 年。[8] M. Bal 和 A. Sengupta，“SpikingBERT：通过隐式微分将 BERT 蒸馏为脉冲语言模型”，载于*AAAI*人工智能会议论文集，第 38 卷，第 10 期，2024 年，页码 10 998–11 006。[9] D. Hong 和 Y. Wang，“用于 ANN 到 SNN 蒸馏中精确中间层匹配的自注意力时空校准”，载于论-

文集，AAAI人工智能会议，第 39 卷，第 16 期，2025 年，页码 17 240–17 248。[10] W. Gerstner 和 W. M. Kistler，《脉冲神经元模型：单个神经元、群体与可塑性》。剑桥大学出版社，2002 年。[11] S. Wu、G. Wang、Y. Song、Y. Zhao、Y. Zhou、Q. Meng 和 Y. Liao，“用于丰富脉冲信息增强目标检测的双向动态阈值脉冲神经网络”，《神经科学前沿》，第 19 卷，文章编号 1661916，2025 年。[12] Y. Li、Y. Guo、S. Zhang、S. Deng、Y. Hai 和 S. Gu，“可微脉冲：重新思考训练脉冲神经网络的梯度下降”，载于《神经信息处理系统进展》，A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，2021 年。[13] K. Wang、J. Zhang、Y. Ren、M. Yao、D. Shang、B. Xu 和 G. Li，“Spikevoice：通过高效脉冲神经网络实现高质量文本到语音”，载于计算语言学协会第 62 届年会论文集（第 1 卷：长篇论文），2024 年，页码 7927–7940。[14] Z. Fu、H. Yang、A. M.-C. So、W. Lam、L. Bing 和 N. Collier，“论参数高效微调的有效性”，载于 AAAI人工智能会议论文集，第 37 卷，第 11 期，2023 年，页码 12 799–12 807。[15] K. Yu、T. Zhang、Q. Xu、G. Pan 和 H. Wang，“TS-SNN：脉冲神经网络的时间移位模块”，载于第四十二届国际机器学习会议，2025 年。 [在线]。可访问：https://openreview.net/forum?id=eFgtUFYe6v [16] X. Song、A. Song、R. Xiao 和 Y. Sun，“具有线性复杂度的单步脉冲变换器”，载于第三十三届国际人工智能联合会议论文集，丛书：IJCAI ’ 2024 [ ] 10 24963 2024 。在线。可访问：https://doi.org/ . /ijcai. / [17] 348 N. Shan、Y. Lin、C. Xie、K. Wu、W. Ma、C. Mu、X. Yan 和 A. Zhang，“由膜电位驱动的脉冲神经网络自适应阈值可塑性：结合反向去极化速率与比例膜电位动力学的仿生机制”，IEEE计算智能新兴主题汇刊，2025 年。[18] K. Takala、W. Thamviset 和 S. Wongthanavasu，“Biasnn：用于图像分类的脉冲神经网络生物启发式注意力机制”，《科学报告》，第 15 卷，第 1 期，文章编号 38753，2025 年。[19] W. Fang、Z. Yu、Y. Chen、T. Masquelier、T. Huang 和 Y. Tian，“引入可学习膜时间常数以增强脉冲神经网络的学习能力”，载于IEEE/CVF国际计算机视觉会议论文集，2021 年，页码 2661–2671。[20] Z. Gong、P. Shi、K. Donbekci、L. Ai、R. Chen、D. Sasu、Z. Wu 和 J. Hirschberg，“以少求多：低资源语音情感识别的自监督方法”，载于 Interspeech 2025，2025 年，页码 151–155。[21] A. Baevski、Y. Zhou、A. Mohamed 和 M. Auli，“Wav2Vec 2.0：用于语音表征自监督学习的框架”，《神经信息处理系统进展》，第 33 卷，页码 12 449–12 460，2020 年。[22] Z. Ma、Z. Zheng、J. Ye、J. Li、Z. Gao、S. Zhang 和 X. Chen，“emotion2vec：语音情感表征的自监督预训练”，载于计算语言学协会论文集：ACL 2024 论文集，2024 年，页码 15 747–15 760。[23] S. Latif、R. Rana、S. Khalifa、R. Jurdak、J. Qadir 和 B. Schuller，“语音情感识别的深度表征学习综述”，IEEE情感计算汇刊，第 14 卷，第 2 期，页码 1634–1654，2023 年。[24] F. D’Asaro、J. J. Márquez Villacís 和 G. Rizzo，“大型语音模型在意大利语语音情感识别中的迁移学习”，载于2024 年IEEE第18届信息与通信技术应用国际会议（AICT），2024 年，页码 1–6。[25] Y. Zhao、K. Wang、Q. Dong 和 T. Ko，“端到端语音翻译的参数高效迁移学习”，载于2024年计算语言学、语言资源与评测联合国际会议论文集（LREC-COLING 2024），2024 年，页码 12 592–12 598。[26] D. Yu、C. Lv、X. Du、L. Jiang、Q. Yin、W. Tong、X. Zheng 和 S. Deng，“利用脉冲神经网络实现高性价比的端侧序列推荐”，载于第三十四届国际人工智能联合会议论文集，2025 年，页码 3579–3587。[27] H. Xu、X. Qiu、Y. Xu、M. E. Elbtity、P. Zhou、Y. Tian、R.-J. Zhu、J. Zhang、S. Gu、Y. Pan 等，“神经形态脉冲大语言模型”，《国家科学评论》，页码 nwaf551，2025 年。[28] M. Yao、O. Richter、G. Zhao、N. Qiao、Y. Xing、D. Wang、T. Hu、W. Fang、T. Demirci、M. De Marchi 等，“基于脉冲的动态计算与异步感知计算神经形态芯片”，《自然·通信》

，第15卷，第1期，第4464页，2024年。

[29] A. Zhang、J. Shi、J. Wu、Y. Zhou 和 W. Yu，“具有自驱动自适应阈值可塑性的低延迟稀疏计算脉冲神经网络”，*IEEE*神经网络与学习系统汇刊，2023。[30] T. Feng 和 S. Narayanan，“Peft-ser：关于使用参数高效迁移学习方法和预训练语音模型进行语音情感识别”，载于*2023*年第*11*届国际情感计算与智能交互会议（*ACII*）。IEEE，2023，第1–8页。[31] S. Deng、Y. Wu、K. Du 和 S. Gu，“脉冲令牌混合器：一种适用于事件驱动的脉冲神经网络前置结构”，神经信息处理系统进展，第37卷，128 846，2024，128 825页–。[32] N. Abuhajar、Z. Wang、M. Baltes、Y. Yue、L. Xu、A. Karanth、C. D. Smith 和 J. Liu，“用于语音增强的三阶段混合脉冲神经网络微调”，《神经科学前沿》，第19卷，文章编号1567347，2025。[33] L. Qiutian，“用于关键词检测的轻量级计算模型”，载于*2024*年第*21*届国际小波活性媒体技术与信息处理计算机会议（*ICCWAMTIP*）。IEEE，2024，第01–05页。[34] A. Bittar 和 P. N. Garner，“通过替代梯度脉冲神经网络探索语音感知过程中的神经振荡”，《神经科学前沿》，第18卷，文章编号1449181，2024。[35] C. A. Buscicchio、P. Górecki 和 L. Caponetti，“使用脉冲神经网络进行语音情感识别”，载于智能系统方法论国际研讨会。Springer，2006，第38–46页。[36] C. Revathy 和 R. Sureshbabu，“利用猫群优化脉冲神经网络对语音信号进行噪声去除和情感识别的有效技术”，《波动与噪声快报》，第21卷，第02期，文章编号2250019，2022。[37] E. Liebenthal、D. A. Silbersweig 和 E. Stern，“情感的语言、语调与韵律：口语情感感知的神经基质与动态”，《神经科学前沿》，第10卷，文章编号506，2016。[38] S. P. Mishra、P. Warule 和 S. Deb，“利用频谱图像和卷积神经网络进行语音情感识别”，载于*2023*年*IEEE*第*20*届印度理事会国际会议（*INDICON*）。IEEE，2023，第161–166页。[39] S. R. Gudivaka、P. Polipogu、R. Kodali、V. R. Dhulipalla、V. S. K. Tatavarty、J. Penumudi 和 P. R. Gogulamudi，“成人与儿童的语音情感识别：传统特征与原始波形模型的全面综述”，国际语音技术期刊，第29卷，第1期，文章编号21，2026。[40] C. Du、F. Liu、B. Kang 和 T. Hou，“基于脉冲神经网络与卷积神经网络的语音情感识别”，人工智能工程应用，第147卷，110314，2025。[41] X. Wang、D. Zhu 和 J. Li，“一种具有负脉冲的快速精准人工神经网络到脉冲神经网络转换算法”，载于第三十四届国际人工智能联合会议论文集，2025，第6460–6468页。[42] T. Brown、B. Mann、N. Ryder、M. Subbiah、J. D. Kaplan、P. Dhariwal、A. Neelakantan、P. Shyam、G. Sastry、A. Askell 等，“语言模型是少样本学习器”，神经信息处理系统进展，第33卷，第1877–1901页，2020。[43] P. Liu、W. Yuan、J. Fu、Z. Jiang、H. Hayashi 和 G. Neubig，“预训练、提示与预测：自然语言处理中提示方法的系统综述”，*ACM*计算调查，第55卷，第9期，第1–35页，2023。[44] B. Lester、R. Al-Rfou 和 N. Constant，“规模对参数高效提示微调的作用”，载于*2021*年自然语言处理实证方法会议论文集。计算语言学协会，2021。[45] M. U. Khattak、H. Rasheed、M. Maaz、S. Khan 和 F. S. Khan，“Maple：多模态提示学习”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2023，第19 113–19 122页。[46] J. Liang、X. Liu、W. Wang、M. D. Plumbley、H. Phan 和 E. Benetos，“声学提示调优：赋予大语言模型听觉能力”，*IEEE*音频、语音与语言处理汇刊，2025。[47] H. Xu、L. Liu、T. Liu、S. Zhi、S. Sun 和 M.-M. Cheng，“面向无源跨域少样本学习的逐步分布对齐风格提示调优”，*IEEE*模式分析与机器智能汇刊，2025。[48] Z. Li、T. Gao、Y. An、T. Chen、J. Zhang、Y. Wen、M. Liu 和 Q. Zhang，“受大脑启发的能效脉冲神经网络用于目标检测

”，载于计算机视觉与模式识别会议论文集，2025，第3552–3562页。[49] W. Doh、D. Lee、M. Chang、J. Mok 和 S. Yoon，“通过多样化脉冲模式增强脉冲神经网络的表征能力”，2025。[50] J. K. Eshraghian、M. Ward、E. O. Neftci、X. Wang、G. Lenz、G. Dwivedi、M. Bennamoun、D. S. Jeong 和 W. D. Lu，“借鉴深度学习经验训练脉冲神经网络”，*IEEE*会刊，第111卷，第9期，第1016–1054页，2023。[51] Z. Huang、W. Fang、T. Bu、P. Xue、Z. Hao、W. Liu、Y. Tang、Z. Yu 和 T. Huang，“用于免训练人工神经网络到脉冲神经网络转换的差分编码”，载于第四十二届国际机器学习会议，2025。[52] A. Islam、M. Foysal 和 M. I. Ahmed，“使用CNN-BiLSTM混合模型从语音音频信号中识别情感”，载于*2024*年第*3*届电气与电子工程进展国际会议（*ICAEEE*）。IEEE，2024，第1–6页。[53] S. Mishra、N. Bhatnagar、P. P 和 S. T. R，“使用深度CNN与BiLSTM混合模型进行语音情感识别与分类”，《多媒体工具与应用》，第83卷，第13期，第37 603–37 620页，2024。[54] M. Li、X. Yang 和 C. Deng，“通过异构知识蒸馏应对脉冲神经网络中的长尾数据挑战”，载于第三十四届国际人工智能联合会议论文集，2025，第1404–1412页。[55] J. Gygax 和 F. Zenke，“阐明脉冲神经网络中替代梯度学习的理论基础”，《神经计算》，第37卷，第5期，第886–925页，2025。[56] Y. Kim、Y. Li、H. Park、Y. Venkatesha、A. Hambitzer 和 P. Panda，“探索脉冲神经网络中的时间信息动态”，载于*AAAI*人工智能会议论文集，第37卷，第7期，2023，第8308–8316页。[57] V. Geadah、S. Horoi、G. Kerg、G. Wolf 和 G. Lajoie，“具有优化单神经元适应能力的神经网络揭示了生物学上合理的正则化”，《*PLOS* 计算生物学》，第20卷，第12期，文章编号e1012567，2024。[58] H. Zou、Y. Si、C. Chen、D. Rajan 和 E. S. Chng，“基于协同注意力的多层次声学信息语音情感识别”，载于*ICASSP 2022——2022*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2022，第7367–7371页。[59] Z. Li、X. Xing、Y. Fang、W. Zhang、H. Fan 和 X. Xu，“用于语音情感识别的多尺度时间变换器”，*arXiv*预印本 *arXiv:2410.00390*，2024。[60] W. Chen、X. Xing、X. Xu、J. Pang 和 L. Du，“DST：用于情感识别的可变形语音变换器”，载于*ICASSP 2023——2023*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2023，第1–5页。[61] S. Shen、F. Liu 和 A. Zhou，“交融还是错位？利用预训练表征和时间移位进行语音情感识别”，载于*ICASSP 2023——2023*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2023，第1–5页。[62] S. Shen、Y. Gao、F. Liu、H. Wang 和 A. Zhou，“用于细粒度语音情感识别的情感神经转导器”，载于*ICASSP 2024——2024*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*）。IEEE，2024，第10 111–10 115页。[63] M. Horowitz，“1.1 计算能耗问题（以及我们可以采取的措施）”，载于*2014*年*IEEE*国际固态电路会议技术论文集（*ISSCC*）。IEEE，2014，第10–14页。