用于准确语音词识别的前馈脉冲神经网络自适应轴突延迟

Pengfei Sun[1], *Ehsan Eqlimi*[1], *Yansong Chua*[2],*Paul Devos*[1],*Dick Botteldoore* n^1

[1]WAVES研究组，根特大学，比利时[2]中国南湖电子与信息技术研究院，中国

arXiv:2302.08607v1 [cs.NE] 16 Feb 2023

摘要

脉冲神经网络（SNN）是构建准确、高效自动语音识别系统的一条前景广阔的研究方向。近期在音频到脉冲编码和训练算法方面的进展，使脉冲神经网络得以应用于实际任务。受生物学启发的脉冲神经网络通过稀疏异步事件进行通信。因此，脉冲时序对脉冲神经网络的性能至关重要。在这方面，大多数研究聚焦于训练突触权重，只有少数研究考虑了事件传输中的延迟，即轴突延迟。在本研究中，我们考虑一种可学习的轴突延迟，并将其限制在最大值以内；该延迟可根据各网络层中的轴突延迟分布进行调整。结果表明，所提出的方法在SHD数据集（92.45%）和NTIDIGITS数据集（95.09%）上取得了已报告的最佳分类结果。本研究展现了训练轴突延迟在具有复杂时间结构的任务中的潜力。

索引词— 自适应轴突延迟、脉冲神经网络、语音处理、听觉建模。

1 引言

准确、高效的语音识别模型是实现低功耗移动设备和智能家电中自动语音识别（ASR）应用的关键 [1]。具备先进深度人工神经网络（ANN） [2], 、强大特征提取算法（如梅尔频率倒谱系数（MFCC）、对数梅尔）以及海量数据的自动语音识别系统已取得巨大成功，在某些任务中达到了人类水平的表现。然而，性能提升也带来了更高的能耗，且当前的特征提取方法在生物学上不够真实 [3] ，并以同步方式处理音频信号。这促使我们探索更高效的解决方案。

近年来，结合脉冲神经网络（SNN）的生物学逼真听觉编码前端 [4, 5, 6]日益受到关注。尤其是，近期研究 [7, 8, 9] 采用替代梯度或迁移学习 [10, 11, 12] 来规避不可微脉冲函数这一问题，为在大型数据集上应用相对深层的脉冲神经网络（SNN）铺平了道路。然而，将脉冲神经网络（SNN）应用于自动语音识别（ASR）仍处于早期阶段，而较新的研究已在事件驱动音频处理算法 [13, 14, 15, 16, 17, 18]方面取得进展。

然而，大多数脉冲神经网络研究都着重于训练突触权重 [19]和膜时间常数 [20, 21]。一些研究还提出训练突触延迟，以学习正确的脉冲时序 [8, 22]。这些方法可视为将信息编码在脉冲时间中，而针对不同传输延迟影响的研究则很少 [8]。Sun 等人[23] 采用了可学习的轴突延迟，在时间复杂度较高的学习任务中效果显著。然而，该方法采用的轴突延迟受固定的最大延迟限制，且无法自适应调整。

本文提出一种训练调度器，可在训练过程中为每一层独立调整轴突延迟上限，从而提升自动语音识别任务的性能。我们采用两阶段训练策略。第一阶段，网络经过若干训练轮次的预训练，以获得初始延迟分布。第二阶段，引入一种用于确定逐层延迟上限的机制，通过滑动窗口大小和上限比例这两个可调参数校准延迟上限，并重新分配轴突延迟。

2 方法描述

我们首先在第 2.1 节介绍脉冲神经元模型，然后在第 2.2 节介绍带有自适应延迟上限的轴突延迟模块。

2.1 脉冲响应模型

脉冲神经元模型得名于其采用脉冲进行通信这一特殊计算范式。由于没有输入事件时无需进行输入积分，因此这种范式具有很高的能效。在我们的工作中，我们采用脉冲响应模型（SRM），其膜电位的描述如下

$$u_i^l(t) = \sum_j (W_{ij}^{l-1}(\varepsilon * s_j^{l-1})(t) + (\nu * s_i^l)(t)) \qquad (1)$$

pengfei.sun@ugent.be, caiyansong@cnaeit.com 本研究获弗兰德斯研究基金会资助（资助编号 G0A0220N），并得到佛兰德政府“Onderzoeksprogramma Artificiele Intelligentie (AI) Vlaanderen”项目支持。本研究还获得中国国家重点研发计划资助（项目编号：2021ZD0200300）。

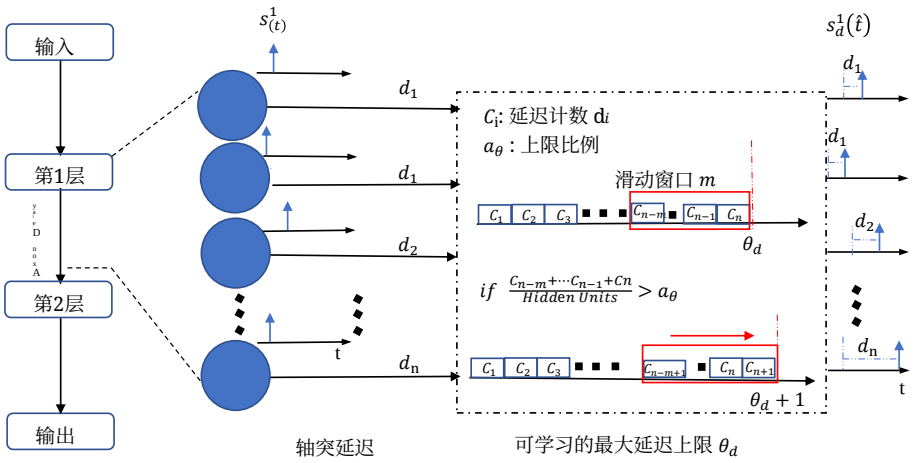

图1。自适应延迟上限的确定方式以及轴突延迟的调整示意图。生成的脉冲 $s^l(t)$ 将在轴突延迟模块中按时间偏移 d_i，随后作为脉冲序列 $s^l_d(\hat{t})$ 输出。自适应调度器会相应地调整延迟上限。不同神经元的延迟值可能相同，例如顶部两个神经元具有相同的延迟值 d_1。该层可以是传统的卷积层、全连接层或循环层。

其中，W l−1ij 表示第 l 层中从神经元 j 到神经元 i 的突触权重，-1 而 u_i^l 表示第 l 层神经元 i 的膜电位，s_j^{l-1} 则表示来自前一神经元 j 的输入脉冲模式。在本实验中，我们通过将输入脉冲 $s^{l-1}(t)$ 与响应核 ε进行卷积，使用响应信号 $a(t) = (\varepsilon * s^{l-1})(t)$ 描述神经元的响应，其中 $\varepsilon(t) = t\tau s \exp(1 - t\tau s)\Theta(t)$。这里，$\Theta(t)$ 表示 Heaviside阶跃函数。同样，不应期信号可表示为 $(\nu * s^l)(t)$，其中 $\nu(t) = -2\theta_u t\tau r \exp(1 - t\tau r)\Theta(t)$。这里，参数 τ_s 和 $\bar{\tau}_r$ 分别为相应核的时间常数。每当 u_i^l 超过预定义阈值θ_u时，便会产生一个输出。该脉冲生成过程可表示为

$$s_i^l(t) = \Theta(u_i^l(t) - \theta_u) \tag{2}$$

2.2 轴突延迟模块和自适应延迟上限

图 1 展示了轴突延迟模块以及用于设置延迟上限的自适应训练调度器。轴突延迟是脉冲传输的一部分，我们对其进行建模，使其能够与优化算法联合学习。

$$s_d^l(\hat{t}) = \delta(t - d^l) * s^l(t) \tag{3}$$

在第l层中，d^l 表示延迟集合 $\{d_1, d2, .., d_n\}$，且满足约束条件 $\{d_1 < d2, .., < d_n \leq \theta_d\}$。同时，$s_d^l(\hat{t})$表示延迟模块在移位后的时间 $\hat{t}$输出的脉冲序列。从优化的角度来看，限制延迟值有助于学习。在这里，我们计算滑动窗口内神经元数量占总数的比例

，并将其用于自适应训练调度器中，以优化训练。

考虑滑动窗口 m，当窗口内延迟神经元所占比例超过预设的上限比例 α_θ 时，滑动窗口向右移动 1 格，延迟上限 θ_d 也随之增加 1。所提出的自适应训练调度器的伪代码见算法 1。

训练期间（在第二个 while 循环中），延迟将按如下方式截断

$$d = max(0, min(d, \theta_d)) \tag{4}$$

其中，θ_d 和延迟 d 将根据我们的调度器进行自适应调整。

3 实验设置

3.1 数据集

我们在 SHD [25]和 NTIDIGITS [26]数据集上评估所提出的方法。这些数据集是由人工耳蜗模型[25, 26] 对原始音频信号进行处理后得到的脉冲模式，具有较高的时间复杂度[27]。SHD数据集包含 12 位说话人的 10,420 条不同长度的话语（0.24s 至 1.17s）。其中共有 20 个数字类别，涵盖英语和德语中的 '0' 至 '9'。我们采用了 [28]，中相同的数据预处理方法，并使用与官方 [25] 相同的数据集训练/验证/测试集划分。

NTIDIGITS 是另一项基于事件的语音识别任务。原始 TIDIGITS 由 64 通道 CochleaAMS1bsensor 处理，并记录其输出。总共有 11 个口语数字（从“0”到“9”的数字，

算法 1 自适应训练调度器的伪代码 输入： θ_d：延迟上限，d：延迟值， α_θ：上限比例，m：滑动窗口大小，C_i：延迟值为 d_i 的神经元数量，n：最大延迟值的索引，T steps：训练步数，HiddenUnits：此层中的神经元数量 初始化：$\{\theta_d\}$， {d}预训练模型：x 个训练轮次 初始化：n $\leftarrow$C.index(max(d))，counter $\leftarrow$ 0滑动窗口内神经元的比例：$\alpha = (Cn_{-m+\ldots+}Cn_{-1}+Cn)$ HiddenUnits当 $\alpha > \alpha_\theta$ 时执行 n $\leftarrow$n $+1\theta_d \leftarrow \theta_d + 1$ counter $\leftarrow$ 0当 counter $<$ T steps 时执行 训练 d 根据公式（4）裁剪 d counter $\leftarrow$ counter $+1$

输出： 最大延迟上限 θ_d，延迟d。

英语和“oh”）。我们采用与 [26] 相同的训练集和测试集划分。

3.2 实现细节

在实验中，我们使用 Adam 优化器，以恒定的学习率 0.1 和 128 的小批量大小联合更新突触权重和轴突延迟。SHD 和 NTIDIGITS 的初始延迟上限分别设为 64 和 128。预训练轮次x 设为 40，两个数据集的仿真时间步长均为 1毫秒 。表 1 列出了实验中使用的参数。

我们使用 SLAYER-Pytorch 框架[8]训练脉冲神经网络。该 GPU 加速软件是近期推出的公开软件，已证明在训练脉冲神经网络方面十分有效。对于这两项任务，我们使用具有两个隐藏全连接层的脉冲神经网络；SHD 使用 128 个隐藏单元，NTIDIGITS 使用 256 个神经元。模型参数总数约为 0.11M（SHD数据集）和 0.08M（NTIDIGITS数据集）。

表1。不同数据集的详细参数设置

数据集	τ_s	τ_r	初始 θ_d	θ_u	T_{steps}
SHD	1	1	64	10	150
NTIDIGITS	5	5	128	10	150

4 结果

4.1 不同上限比例和滑动窗口大小的消融研究

窗口大小和上限比例是获得合理延迟上限的重要参数。为了评估它们的影响，我们

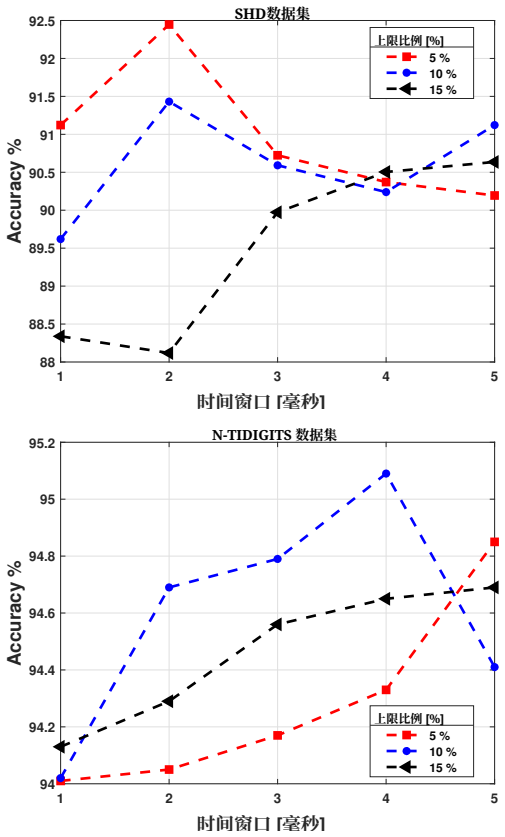

图 2。基于（上）SHD 和（下）N-TIDIGITS，对不同滑动窗口大小m和上限比例 α_θ 进行消融研究。x 轴表示滑动窗口大小，y 轴表示准确率。实验中使用了 3 种不同的上限比例（5%、10% 和 15%）以及 5 种不同的窗口大小（1、2、3、4 和 5）。

设计了消融研究来比较其影响，其中包含5种不同的窗口大小和3种不同的比例参数。这些参数对SHD和NTIDIGITS数据集的影响如图2所示。对于SHD数据集，我们观察到，较小的比例始终能够取得较好的结果；通过将最大的两个延迟神经元数量控制在神经元总数的5%以内，可以获得最佳结果，这意味着窗口大小设为2、上限比例设为5%。而对于NTIDIGITS数据集，较大的窗口大小和比例更加有利；除窗口大小为5且上限比例为10%的情况外，准确率会随着窗口大小增大而持续提升。

4.2 SHD

表 2 给出了前馈脉冲神经网络、具有自适应发放阈值的循环脉冲神经网络（RSNN）、异构循环脉冲神经网络、具有时间注意力的循环脉冲神经网络以及我们所提出的方法的性能结果。如表 2 所示，我们所提出的方法达到了 92.45% 的准确率，这是目前针对该任务报告的最佳性能

表 2。在网络规模和准确率方面与当前最先进方法的比较。

数据集	方法	参数	准确率
SHD	前馈脉冲神经网络 [25]	0.11 MB	48.6±0.9%
	循环脉冲神经网络 [25]	1.79 MB	83.2±1.3%
	带自适应的循环脉冲神经网络 [28]	0.14 MB	84。4%
	异构循环脉冲神经网络 [20]	0.11 MB	82.7±0.8%
	带时间注意力的脉冲神经网络 [29]	0.12 MB	91.08%
	本工作（m=2, α_θ = 5%）	**0.11 MB**	**92.45%**
NTIDIGITS	门控循环单元循环神经网络 [26]†	0.11 MB	90。90%
	分相长短期记忆网络 [26]†	0.61 MB	91.25%
	ST-RSBP [30]	0.35 MB	93.63±0.27%
	采用轴突延迟的脉冲神经网络 [23]	0.08 MB	94.45%
	本工作 (m=4, α_θ = 10%) 0.08 MB		**95.09%**

† 非脉冲神经网络实现。

。更重要的是，可以观察到循环脉冲神经网络的性能优于前馈脉冲神经网络，这表明该任务本质上具有动态性。然而，我们的结果表明，前馈脉冲神经网络无需循环连接，仅通过采用自适应轴突延迟模块，便能够以更少的参数实现灵活的短期记忆和更高的准确率。

4.3 NTIDIGITS

将人工神经网络（循环神经网络和分相长短期记忆网络）应用于数字识别任务时，其准确率分别达到 90.90% 和 91.25%（见表 2）。然而，这些网络无法充分发挥基于稀疏事件的信息的优势，并且必须依赖事件合成算法，实际上失去了处理嵌入脉冲时序中的信息时所获得的优势。利用我们的自适应轴突延迟模块，可实现 95.09% 的最佳性能；与 Zhang 等人[30]直接使用脉冲序列级特征相比，我们的模型仅使用 23% 的参数，性能便提升了 1.4%。

4.4 延迟值的影响

如表3所示，延迟上限对模型性能至关重要。对于这两项音频分类任务，即使不限制延迟值的范围，模型性能仍具有竞争力，这证明了轴突延迟模块的有效性。不过，限制延迟范围仍然是必要的；我们的实验表明，合适的延迟上限能够提升模型的分类能力。以SHD数据集为例，我们的自适应训练调度器能够确定最优延迟分布，使模型达到最佳性能。采用其他延迟上限组合时，准确率会下降，这可能表明每种网络结构都有其最优延迟上限，但手动寻找这些延迟上限并不容易。我们的方法提供了一种自适应训练调度器，用于搜索这些最优参数。值得注意的是，NTIDIGITS数据集也符合这一现象。

5 结论

本文将可学习的轴突延迟模块集成到脉冲神经元模型中，随后引入自适应训练调度器，以调整每个网络层。与采用静态延迟上限的先前工作相比，我们提出的方法在不增加额外参数的情况下，显著提升了分类能力。此外，我们的自适应调度器可以轻松集成到现有延迟模块中，并自适应地确定网络的最优延迟分布。我们在 SHD（92.45%）和 NTIDIGITS（95.09%）数据集上，以最少的参数取得了最佳性能。这些结果表明，采用自适应延迟上限的神经元轴突延迟可用于构建轻量、灵活的短期记忆模块，从而实现准确且高效的口语词识别系统。我们推测，轴突延迟机制无需增加可训练参数数量，即可引入一种短期记忆。对于自动语音识别中的某些数据集，如果 1）信息以短序列组织，无需长期记忆；且 2）数据量有限，因而容易过拟合，那么轴突延迟机制与前馈的小型脉冲神经网络相结合，可能效果最佳。我们的实验似乎印证了上述观点，并进一步证实了将脉冲时序作为自动语音识别问题解决方案之一的巨大潜力。此外，采用基于脉冲的损失 [31]可以加快决策，从而进一步减轻额外延迟的影响。

表3。不同延迟上限方法的消融研究，以及延迟上限 θ_d 在轴突延迟模块中的作用。θ_{d_i} 表示i^{th} 层的延迟。“手动”指使用静态延迟上限，而“自适应”指使用我们提出的自适应调度器。

数据集	方法	$(\theta_{d_1}, \theta_{d_2})$	参数	准确率
SHD	手动	(0, 0)	108,820	67.05%
	手动	(64, 64)	109,076	86。84%
	手动	(128, 128)	109,076	87。24%
	自适应	（107，175）	109,076	**92**.%
	手动	(+∞,+∞)	109,076	84.99%
NTIDIGITS	手动	(0，0)	85,259	78.86%
	手动	(128, 128)	85,771	94。19%
	自适应	(215, 215)	85,771	**95.09**%
	手动	(+∞,+∞)	85,771	93.83%

6 参考文献

[1] Roshan Gopalakrishnan、Yansong Chua、Pengfei Sun、Ashish Jith Sree- jith Kumar 和 Arindam Basu，“Hfnet：一种针对神经形态硬件协同设计、具有突触交叉阵列的卷积神经网络架构”，神经科学前沿，第14卷，第907页，2020年。[2] Alex Graves、Abdel-rahman Mohamed 和 Geoffrey Hinton，“采用深度循环神经网络进行语音识别”，载于*2013*年 *IEEE* 国际声学、语音与信号处理会议。IEEE，2013年，第6645–6649页。[3] Deepak Baby、Arthur Van Den Broucke 和 Sarah Verhulst，“用于实时应用的人类耳蜗力学与滤波器调谐卷积神经网络模型”，*Nature Machine Intelligence*，第3卷，第2期，第134–143页，2021年。[4] Jibin Wu、Yansong Chua 和 Haizhou Li，“一种基于脉冲神经网络、符合生物学原理的语音识别框架”，载于*2018*年国际神经网络联合会议（*IJCNN*），2018年，第1–8页。[5] Zihan Pan、Yansong Chua、Jibin Wu、Malu Zhang、Haizhou Li 和 Eliathamby Ambikairajah，“一种高效且基于感知的脉冲神经网络听觉神经编码与解码算法”，神经科学前沿，第13卷，第1420页，2020年。[6] Jibin Wu、Yansong Chua、Malu Zhang、Haizhou Li 和 Kay Chen Tan，“一种用于鲁棒声音分类的脉冲神经网络框架”，神经科学前沿，第12卷，第836页，2018年。[7] Emre O Neftci、Hesham Mostafa 和 Friedemann Zenke，“脉冲神经网络中的替代梯度学习”，*IEEE Signal Processing Magazine*，第36卷，第61–63页，2019年。[8] Sumit Bam Shrestha 和 Garrick Orchard，“SLAYER：脉冲层时间误差重新分配”，载于神经信息处理系统进展第*31*卷，第1412–1421页。Curran Associates, Inc.，2018年。[9] Malu Zhang、Jiadong Wang、Jibin Wu、Ammar Belatreche、Burin Amornpaisannon、Zhixuan Zhang、Venkata Pavan Kumar Miriyala、Hong Qu、Yansong Chua、Trevor E Carlson 等，“用于深度脉冲神经网络反向传播的修正线性突触后电位函数”，*IEEE Transactions on Neural Networks and Learning Systems*，第33卷，第5期，第1947–1958页，2021年。[10] Jibin Wu、Yansong Chua、Malu Zhang、Guoqi Li、Haizhou Li 和 Kay Chen Tan，“一种用于深度脉冲神经网络有效训练与快速推理的串联学习规则”，*IEEE Transactions on Neural Networks and Learning Systems*，2021年。[11] Jibin Wu、Chenglin Xu、Xiao Han、Daquan Zhou、Malu Zhang、Haizhou Li 和 Kay Chen Tan，“面向深度脉冲神经网络模式识别的渐进式串联学习”，*IEEE Transactions on Pattern Analysis and Machine Intelligence*，第44卷，第11期，第7824–7840页，2021年。[12] Qu Yang、Jibin Wu、Malu Zhang、Yansong Chua、Xinchao Wang 和 Haizhou Li，“采用局部串联学习训练脉冲神经网络”，*arXiv*预印本 *arXiv:2210.04532*，2022年。[13] Jibin Wu、Zihan Pan、Malu Zhang、Rohan Kumar Das、Yansong Chua 和 Haizhou Li，“鲁棒声音识别：一种神经形态方法”，载于 *INTERSPEECH*，2019年，第3667–3668页。[14] Jibin Wu、Emre Yılmaz、Malu Zhang、Haizhou Li 和 Kay Chen 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，第199页，2020年。[15] Rong Xiao、Rui Yan、Huajin Tang 和 Kay Chen Tan，“一种用于声音识别的脉冲神经网络模型”，载于认知系统与信号处理，Fuchun Sun、Huaping Liu 和 Dewen Hu 编，新加坡，2017年，第584–594页，Springer Singapore。[16] Peter Blouw 和 Chris Eliasmith，“采用神经形态计算系统的事件驱动信号处理”，载于*ICASSP 2020—2020*年 *IEEE* 国际声学、语音与信号处理会议（*ICASSP*），2020年，第8534–8538页。

[17] Emre Yılmaz、Ozg¨ur Bora Gevrek、Jibin Wu、Yuxiang Chen、Xuanbo Meng 和 Haizhou Li，“用于关键词检测的深度卷积脉冲神经网络”，载于*INTERSPEECH*会议论文集，2020年，第2557–2561页。[18] Malu Zhang、Jibin Wu、Yansong Chua、Xiaoling Luo、Zihan Pan、Dan Liu 和 Haizhou Li，“Mpd-al：一种用于脉冲神经元的高效膜电位驱动聚合标签学习算法”，载于*AAAI*人工智能大会论文集，2019年。[19] Sumit Bam Shrestha 和 Qing Song，“基于脉冲事件权重更新的鲁棒脉冲序列学习”，《神经网络》，第96卷，第33–46页，2017年。[20] Nicolas Perez-Nieves、Vincent CH Leung、Pier Luigi Dragotti 和 Dan FM Goodman，“神经元异质性促进鲁棒学习”，《自然-通讯》，第12卷，第1期，第1–9页，2021年。[21] Wei Fang、Zhaofei Yu、Yanqi Chen、Timoth´ee Masquelier、Tiejun Huang 和 Yonghong 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，第2661–2671页。[22] Malu Zhang、Jibin Wu、Ammar Belatreche、Zihan Pan、Xiurui Xie、Yansong Chua、Guoqi Li、Hong Qu 和 Haizhou Li，“具有突触延迟权重可塑性的脉冲神经网络监督学习”，《神经计算》，第409卷，第103–118页，2020年。[23] Pengfei Sun、Longwei Zhu 和 Dick Botteldooren，“轴突延迟作为前馈深度脉冲神经网络的短期记忆”，*IEEE*声学、语音与信号处理国际会议论文集，2022年。

[24] Xuetong Niu、Li Wang 和 Xulei Yang，“信用卡欺诈检测比较研究：监督式与无监督式”，*arXiv* 预印本 *arXiv:1904.10604*，2019年。

[25]Benjamin Cramer、Yannik Stradmann、Johannes Schemmel 和 Friedemann Zenke，“用于系统评估脉冲神经网络的海德堡脉冲数据集”，*IEEE Transactions on Neural Networks andLearningSystems*，2020年。

[26] Jithendar Anumula、Daniel Neil、Tobi Delbruck 和 Shih-Chii Liu，“用于序列分类的预训练循环神经网络的特征表示”，神经科学前沿，第12卷，第23页，2018年。

[27] Laxmi R Iyer、Yansong Chua 和 Haizhou Li，“神经形态 MNIST 是否名副其实？在时域中分析神经形态数据集的判别能力”，神经科学前沿，第15卷，第297页，2021年。[28] Bojian Yin、Federico Corradi 和 Sander M Boht´e，“利用多时间尺度脉冲循环神经网络实现高效计算”，载于*2020*年国际神经形态系统会议，2020年，第1–8页。[29] Man Yao、Huanhuan Gao、Guangshe Zhao、Dingheng Wang、Yihan Lin、Zhaoxu Yang 和 Guoqi Li，“用于事件流分类的时间维度注意力脉冲神经网络”，载于*IEEE/CVF*国际计算机视觉会议论文集，2021年，第10221–10230页。[30] Wenrui Zhang 和 Peng Li，“用于训练深度循环脉冲神经网络的脉冲序列级反向传播”，神经信息处理系统进展，第32卷，2019年。[31] Sumit Bam Shrestha、Longwei Zhu 和 Pengfei Sun，“Spikemax：用于分类的基于脉冲的损失方法”，载于*2022*年国际神经网络联合会议（*IJCNN*），2022年，第1–7页。