文章编号: 1007-4252(2025)05-0364-15 DOI: 10.20027/j.gncq.2025.0046 CSTR: 32180.14.j.gncq.2025.0046

面向视觉目标检测的脉冲神经网络综述:从生物机制到前沿应用

作赛飞,张 渊,谢 迪,俞 海*,朱 江
(海康威视研究院,杭州 310052)

摘要: 脉冲神经网络(spiking neural network, SNN)作为第三代神经网络模型,凭借其生物启发的脉冲动力学机制、事件驱动的异步计算特性以及低功耗优势,在目标检测领域展现出广阔的应用前景。系统综述面向视觉目标检测的 SNN 方法,涵盖其生物机制基础、神经元模型、神经编码策略、数据集分类及主流算法方案。在神经元模型方面,分析积分-发放(integrate-and-fire, IF)、泄漏积分-发放(leaky integrate-and-fire, LIF)、Izhikevich 和 Hodgkin-Huxley 等模型在生物合理性与计算效率之间的权衡;在编码机制上,对泊松编码、强度-延迟编码等输入编码方法以及速率、时间、群体解码策略进行阐述;在数据集层面,系统总结静态、神经形态转换、神经形态捕获和模拟生成等 4 类数据集的特点与局限性。在 SNN 算法方面,详细分析基于人工神经网络(artificial neural network, ANN)-to-SNN 转换的方法与基于代理梯度的直接训练方法,并对比其在精度、能效、延迟和硬件适配性等方面的表现。对 SNN 在训练-硬件协同优化、新架构设计、多模态扩展及工具链生态等方向的发展趋势进行展望,为低功耗、高能效脉冲视觉系统的研究与应用提供参考。

关键词: 脉冲神经网络;目标检测;ANN-to-SNN 转换;代理梯度训练;低功耗
中图分类号: TP183 **文献标志码:** A

Comprehensive survey of spiking neural networks for visual object detection: from biological mechanisms to state-of-the-art applications

WU Saifei, ZHANG Yuan, XIE Di, YU Hai*, ZHU Jiang
(Hikvision Research Institute, Hangzhou 310052, China)

Abstract: Spiking neural network (SNN), recognized as the third generation of neural network models, exhibit substantial potential in the domain of object detection. This potential arises from their biologically inspired spiking mechanisms, event-driven asynchronous computational characteristics, and low-power consumption benefits. This paper presents a systematic review of SNN methodologies for visual object detection, encompassing their biological foundations, neuron models, neural encoding techniques, dataset

收稿日期: 2025−09−26; **修订日期:** 2025−10−14
基金项目: 新一代人工智能国家科技重大专项(No. 2020AAA0109004).
作者简介: 作赛飞(1995—),男,硕士,主要研究方向为计算机视觉(E-mail: wusaifei@hikvision.com).
通信作者: 俞 海(1981—),男,硕士,副研究员,主要研究方向为类脑计算(E-mail: yuhai@hikvision.com).

classifications, and prevailing algorithmic frameworks. Concerning neuron models, the balance between biological plausibility and computational efficiency is analyzed across models such as IF (integrate-and-fire), LIF (leaky integrate-and-fire), Izhikevich, and Hodgkin-Huxley. Regarding encoding mechanisms, input encoding techniques like Poisson encoding and intensity-latency encoding, along with decoding strategies including rate, time, and population decoding, are thoroughly discussed. At the dataset level, the attributes and limitations of four dataset categories —static, neuromorphic conversion, neuromorphic acquisition, and simulation-generated—are systematically summarized. For SNN algorithmic frameworks, methods based on ANN (artificial neuralnetwork)-to-SNN conversion and direct training using surrogate gradients are explored in detail, with comparisons of their performance in terms of accuracy, energy efficiency, latency, and hardware compatibility. This paper delineates future trends in SNN development, including training-hardware co-optimization, innovative architecture design, multimodal expansion, and toolchain ecosystem advancement, offering insights into the research and application of low-power, high-efficiency spiking vision systems.

Key words: spiking neural network; object detection; ANN-to-SNN conversion; surrogate gradient training; low power consumption

0　引言

脉冲神经网络(spiking neural network, SNN)[1-9]作为人工智能与计算神经科学领域的新型计算建模范式,通过高度模拟生物大脑中神经元的动态发放机制,在结构与功能层面实现了对生物神经系统的更紧密拟合。与传统的人工神经网络(artificial neural network, ANN)[10-12]相比,SNN 在处理具有异步性与时空动态性的事件流数据时表现出显著优势:传统 ANN 方法难以有效融合事件发生的时间与空间信息,而 SNN 通过模拟生物神经元的脉冲发放机制,以离散脉冲序列对信息进行编码与传递,能够更自然地整合时空特征。

在 SNN 中,神经元的膜电位随时间动态更新,仅当其超过特定阈值时才发放二值脉冲,这一机制使得时空信息被嵌入脉冲的时序与空间分布中。由于采用了事件驱动的稀疏计算模式,SNN 仅在接收到输入脉冲时触发神经元活动,显著降低了计算与通信成本。这一特性使其特别适合在 Loihi[13]、TrueNorth[14]等神经形态硬件上实现高效部署,并被广泛应用于动态视觉传感器(dynamic vision sensor, DVS)、边缘智能等对实时性与能效要求较高的情境中。

目标检测作为计算机视觉中的核心任务,旨在对图像中的多个目标进行识别与精确定位,在自动驾驶、安防监控、医学图像分析等领域具有重要应用价值。当前的主流检测框架(如 YOLO 系列、R-CNN 系列等)大多基于 ANN 构建,虽具备较高的性能,但常伴随过高的计算复杂度与能耗开销。作为第三代神经网络模型,SNN 提供了一种结构更接近生物神经系统、能效更高的替代方案:利用脉冲信号进行稀疏、异步的事件通信。该方案不仅降低了数据传输与存储需求,也避免了冗余计算,为开发高效、低功耗的目标检测系统提供了新的技术方向。

1　神经元动力学

本节系统阐述 SNN 中采用的神经元模型,涵盖生物神经元的行为机制、多样化的计算建模方法以及神经元的动态重置过程,还将探讨将输入数据编码为脉冲序列的神经编码策略,以及基于脉冲输出进行信息解码以支持检测分类等任务的技术途径。对 SNN 中各类神经元模型的深入理解,为分析其计算特性与网络动态行为奠定重要基础。

1.1　生物神经元

本小节主要阐述生物神经元对 SNN 的启发作用,为 SNN 的设计与理论理解提供基本蓝图。生物神经元通过电脉冲,即动作电位或峰电位(spike)进行信息传递,持续时间通常约为 10 ms[15]。如图 1 所示,生物神经元主要由 3 个部分构成:树突、轴突和

体细胞(即细胞体)[16]。树突负责接收来自其他神经元轴突的信号,这些信号经树突传导至体细胞后会改变其膜电位。一旦膜电位超过特定阈值,体细胞立即产生动作电位,该动作电位随后沿轴突传播至其他神经元。在轴突末端,动作电位的到达会促使突触囊泡向突触间隙(即突触前与突触后神经元之间的空隙)释放神经递质[17]。突触通过神经递质实现轴突与树突间的通信。这些化学物质与突触后神经元树突上的受体结合,通过调控离子通道的开、闭状态,影响新峰电位产生的概率[18]。

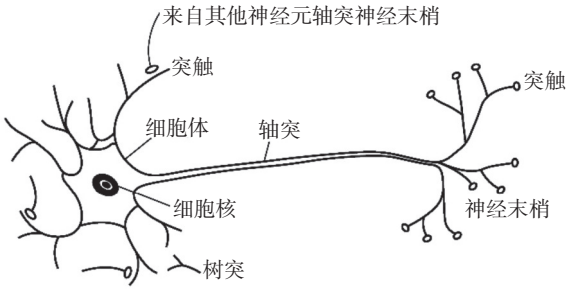

图 1 生物神经元模型

Fig. 1 Model of biological neuron

从电位变化的动态过程来看,生物神经元在产生动作电位期间,其膜电位会经历以下典型阶段:

(1)静息状态:膜电位维持在约−70 mV;

(2)去极化:钠离子通道开放,钠离子内流导致膜电位快速上升;

(3)复极化:钾离子通道开放,钾离子外流使膜电位逐渐回落;

(4)超极化:膜电位暂时低于静息电位水平,从而抑制新的动作电位的立即产生。

神经元在动作电位发放后进入不应期,该时期可分为两个阶段:

(1)绝对不应期:对应于去极化与复极化阶段,在此期间无论刺激强度如何,神经元均无法再次激发动作电位,这一机制可有效防止神经元过度频繁放电;

(2)相对不应期:紧随绝对不应期之后,对应于超极化阶段,此时膜电位低于静息电位水平,神经元虽可被激发,但所需刺激强度高于正常水平,激发难度显著提高。

突触强度(或称权重)受神经递质释放量、受体可用性及信号传导特性等因素的调控。兴奋性突触和抑制性突触分别会提高和降低突触后神经元的膜电位。在无外部输入的情况下,膜电位随时间呈指数衰减,并逐渐恢复至静息电位水平。

1.2 人工神经元模型

理解 SNN 中采用的各类神经元模型对于把握其通用性与功能多样性具有重要意义。不同神经元模型在生物合理性与计算效率方面存在差异,这些差异直接影响神经网络的设计与实现。

1.2.1 积分-发放(integrate-and-fire, IF)神经元模型

在常见模型中,IF 神经元模型因计算效率高、结构简洁且易于实现等特点而被广泛采用。该模型通过整合输入脉冲使膜电位随时间累积;当电位超过预设阈值时,神经元发放脉冲(动作电位),随后膜电位被重置至静息电位(通常为零)。IF 神经元的动力学行为可由如下方程描述:

$$\tau_{\mathrm{m}}\frac{\mathrm{d}v}{\mathrm{d}t}=RI(t) \tag{1}$$

式中,τ_{m}为膜时间常数,R 为膜电阻,$I(t)$为输入电流。

IF 模型尽管捕捉了脉冲发放的核心动态,但忽略了诸如不应期和复杂突触动力学等关键生物机制[19]。

1.2.2 泄漏积分-发放(leaky integrate-and-fire, LIF)神经元模型

LIF 神经元模型是对基础 IF 神经元模型的重要扩展,通过引入膜漏电流机制,该模型可更准确地模拟在无外界输入时生物神经元的膜电位随时间指数衰减至静息电位的过程。该模型由 Lapicque 于 1907 年基于电容器电路和青蛙坐骨神经实验提出[20],较好地平衡了生物合理性与计算效率现,已成为 SNN 研究中应用最广泛的计算模型之一。当膜电位因输入电流累积而超过发放阈值时,神经元产生脉冲,并在发放后重置电位,且通常辅以短暂的不应期模拟以提升生物真实性[21]。

LIF 神经元模型的膜电位 $v(t)$ 动态行为由以下常微分方程描述:

$$\tau_{\mathrm{m}}\frac{\mathrm{d}v}{\mathrm{d}t}=-v(t)+RI(t) \tag{2}$$

漏电流项$-v(t)$ 的引入,模拟了在缺乏外部输入情况下膜电位随时间呈指数衰减至静息电位水平的被动过程,反映了生物神经元因膜漏电导而逐渐损失电荷的生理特性。该衰减机制有效避免了膜电位的无界累积,从而确保神经元对输入刺激表

现出动态和短暂的响应特性, 这一行为对于模拟真实神经系统中的信息处理过程至关重要[19]。

1.2.3　脉冲响应模型(spike-response model, SRM)

SRM 是对 LIF 模型的广义扩展, 通过引入核函数(kernel)描述神经元的动态响应行为, 在数学表达和计算实现方面具有更高的灵活性。SRM 最初于文献 [16] 中被提出, 其采用两个核核函数 η 和 κ 分别刻画神经元的内源性动态与外部输入响应: η 负责模拟历史脉冲对当前膜电位的影响, 涵盖诸如不应期(即脉冲发放后神经元兴奋性暂时降低的时期)和后电位(脉冲引发膜电位的短暂波动)等现象; κ 则用于描述神经元对外部输入电流 $I_{\text{ext}}(t)$ 的响应特性。

SRM 框架中的膜电位取决于该神经元的脉冲发放历史及所接收的外部输入。静息状态下, 膜电位维持在静息电位水平; 当施加外部电流时, 膜电位偏离静息电位水平, 其演化过程受核函数 κ 所控制。这一响应行为可通过 κ 核函数与输入电流的卷积运算进行数学表征。因此, SRM 中的膜电位由两部分共同决定: 一是历史脉冲通过 η 核产生的影响, 二是外部输入经由 κ 核卷积所贡献的响应。

膜电位 $v(t)$ 的完整数学表达式可写为:

$$v(t) = v_{\text{rest}} + \sum_{t_i \in \mathcal{F}(t)} \eta(t - t_i) + \int_0^{\infty} \kappa(s) I_{\text{ext}}(t - s)\mathrm{d}s \tag{3}$$

式中, v_{rest} 为静息膜电位, $\mathcal{F}(t)$ 为该神经元在时间 t 之前的所有脉冲发放时刻的集合, $\eta(t - t_i)$ 为脉冲响应核函数, $\kappa(s)$ 为输入响应核函数。积分项表示输入电流与核函数 κ 的卷积, 用于捕捉输入引起的电位变化。该方程将膜电位分解为静息基准、历史脉冲效应和外部输入响应, 这 3 项共同构成了 SRM 对生物神经元动态的高度抽象与数学描述。

1.2.4　Izhikevich 神经元模型

Izhikevich 神经元模型是计算神经科学中一类兼具生物合理性与计算效率的动力学模型, 由 Izhikevich [22] 提出。该模型通过两个耦合的一阶常微分方程来描述神经元膜电位 v 和恢复变量 u 的演化行为, 其数学表达式如下:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 0.04v^2 + 5v + 140 - u + I \tag{4}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t} = a(bv - u) \tag{5}$$

式 (4) 和 (5) 中, I 代表突触输入电流或其他外部刺激, a、b 为调控系统动力学的核心变量。

当膜电位 $v \geqslant 30$ mV时, 发放与重置机制为:

$$v \leftarrow c \tag{6}$$

$$u \leftarrow u + d \tag{7}$$

其中, 恢复变量 u 用于模拟钾离子通道激活和钠离子通道失活等过程的综合效应, 参数 c、d 为调控系统动力学的核心变量。

在该模型中: 参数 a 控制 u 的时间尺度, 影响神经元发放后 v 恢复至静息电位水平的速度; 参数 b 调节 u 对 v 变化的敏感性, 从而影响神经元兴奋性与发放模式之间的耦合强度; 参数 c 决定着动作电位峰值后 v 的重置水平, 直接影响神经元再次达到发放阈值的难易程度; 参数 d 指定每次发放后 u 的复位增量, 进而影响后续的发放行为。通过合理配置这些参数, Izhikevich 模型能够重现持续发放、簇状放电和谐振振荡等多种神经元的典型放电模式, 因此在大规模 SNN 仿真中得到了广泛应用。

1.2.5　霍奇金-赫胥黎模型(Hodgkin-Huxley model)

霍奇金-赫胥黎模型通过刻画动作电位产生过程中离子通道的动态机制, 实现了对神经元电生理行为的定量描述[23]。该模型由 Hodgkin 与 Huxley 于 1952 年基于巨型乌贼轴突的实验数据建立, 其核心机制是通过门电压来控制钠离子(Na^+)与钾离子(K^+)通道的电导变化。

该神经元模型通过以下微分方程来控制膜电位 v 的演化:

$$C_{\text{m}}\frac{\mathrm{d}v}{\mathrm{d}t} = -g_{\text{Na}}m^3h(v - E_{\text{Na}}) - g_{\text{K}}n^4(v - E_{\text{K}}) - g_{\text{L}}(v - E_{\text{L}}) + I_{\text{ext}} \tag{8}$$

式中, C_{m} 为膜电容; g_{Na}、g_{K}、g_{L} 分别为最大钠电导、钾电导和漏电导; E_{Na}、E_{K}、E_{L} 为相应离子的平衡电位; m、h、n 为门控变量, 分别用于描述钠通道激活、钠通道失活和钾通道激活的动态过程, 其演化服从一阶动力学方程。该模型通过上述机制精确再现了动作电位的产生与传播过程, 成为计算神经科学中生物物理建模的重要基础。

1.2.6　膜电位重置机制

膜电位重置机制虽非一类独立的神经元模型, 但对理解神经元发放后的动态行为具有重要意义。该机制决定神经元在产生动作电位后是彻底复位至基准电位(硬重置), 还是保留部分超阈值残余电位(软重置), 从而直接影响信息保持能力与后续响应特性。这类机制常见于 LIF、IF 等简化脉冲神经

元模型中，而 Izhikevich 与 Hodgkin-Huxley 等生物模型则通过连续微分方程自然地描述膜电位演化，因此不依赖显式离散重置操作[22, 24-25]。

硬重置型脉冲神经元[26-27]：在硬重置机制中，当膜电位超过发放阈值时，其值被强制重置为零。该方法具备结构简洁与计算高效的特点，但会丢弃超阈值部分的所有电位信息，可能导致信号积分过程中的信息损失，进而影响神经元对持续输入响应的准确性。

软重置型脉冲神经元[24, 28-29]：软重置机制在发放后使膜电位减去阈值电压，从而保留超出阈值的残余电位，因此该机制也被称为“减法重置”[28-29]或“残余膜电位重置”（residual membrane potential, RMP）[24]。这一方法有助于更精确地编码输入信号历史并提升推理任务中的性能表现。

1.2.7 小结

在 SNN 的研究中，若干经典神经元模型在计算效率与生物合理性之间实现了较好的平衡。IF 神经元模型及其泄漏变体 LIF 神经元模型因兼具较高的仿真效率与一定的生物可信度而被广泛采用。LIF 神经元模型通过引入泄漏项，在不显著增加计算负担的前提下提升了膜电位动力学的生物学合理性。SRM 神经元模型在 LIF 神经元模型的基础上进一步扩展，借助核函数刻画历史脉冲与输入响应，增强了模型的表达灵活性。Izhikevich 模型能够高效复现多种生物神经元发放模式，而 Hodgkin-Huxley 模型则以较高的计算成本精确模拟离子通道动力学。

尽管后几种模型能够更好地模拟生物物理机制，但 IF 和 LIF 神经元模型因在效率与真实性之间的良好平衡，仍在实践中占主导地位。

1.3 神经编码

神经元的编码机制是将静态或动态输入信息转化为动作电位（脉冲）序列，并进一步解码这些序列以完成信息处理与传递的关键过程。在本小节中，我们系统阐述输入编码与输出解码的基本原理及其具体实现方式。

1.3.1 输入编码

输入编码（input encoding）负责将外部刺激或数据转换为神经元可处理的脉冲序列。根据刺激特性与编码机制的不同，主要有静态编码与动态编码这两类编码方式。

（1）泊松编码（Poisson encoding）：将输入帧中每个像素的强度值转换为服从泊松过程的脉冲序列，发射率与像素强度成正比。高亮度像素对应于较高的发放频率，低亮度像素则对应于发放频率较低的情况。通过引入发放时间的随机性来模拟生物神经元响应中的随机特性，以增强模型对噪声的鲁棒性。该编码方法常见于处理图像或光强信息的场境中[30-31]。

（2）强度-延迟编码（intensity-to-latency encoding）：每个像素仅产生一个脉冲，其发放时刻与像素强度成反比，即亮度越高脉冲发生越早。这种编码方法显著降低了脉冲数量和计算开销，适用于高时间分辨率场景和低功耗神经形态系统[32-33]。

（3）增量调制（delta modulation）：作为一种动态编码方法，其根据输入信号强度的变化触发脉冲生成，仅当像素强度发生显著改变时才产生脉冲。该方法能够高效捕捉时序动态特征，提供稀疏的事件型表示，广泛应用于事件相机与 DVS[34]。

1.3.2 输出解码

输出解码（output decoding）将 SNN 产生的脉冲序列重新转换为可用于分类或决策任务的明确信号，主要包括以下 3 类方法。

（1）速率解码（rate decoding）：基于神经元在单位时间内的脉冲发放频率进行类别判断，通常选择发放频率最高的神经元所代表的类别作为结果输出。这种方法通过对多个脉冲进行统计聚合，显著提高了抗噪声能力和系统稳定性[16]。

（2）时间解码（temporal decoding）：依据首个达到发放阈值的神经元的脉冲时间进行决策，认为响应更早的神经元对当前输入具有更高的置信度。时间解码具备高速度和低能耗的特点，适用于对实时性要求较高的硬件系统[35]。

（3）群体解码（population decoding）：利用一组神经元共同表达某一类别的信息，可结合速率或时间解码策略进行集成判断，如使用群体向量或投票机制。群体解码通过神经冗余显著提高了系统的容错性和表达能力[36]。

1.3.3 小结

神经编码机制的设计直接影响 SNN 的信息表示能力、计算效率及生物合理性。不同的输入编码方法各有优势：泊松编码通过概率性发放模拟生

物神经元的不确定性; 强度-延迟编码以时间信息代替频率信息, 提高了处理效率; 增量调制专注于动态信息捕捉, 适用于对事件流数据的处理。输出解码方面, 速率解码稳定性高, 时间解码响应迅速, 群体解码则提升了系统的鲁棒性。

这些编码与解码机制不仅为 SNN 研究提供了理论依据, 也在神经形态计算、事件相机、脑机接口等领域具有广阔的应用前景[37]。未来的研究可侧重于编码-解码机制的协同优化及其在复杂认知任务与边缘设备中的应用拓展。

2　脉冲神经网络数据集

当前所采用的数据集可系统性地划分为 4 类: 静态数据集(static datasets)、神经形态转换数据集(neuromorphic converted datasets)、神经形态捕获数据集(neuromorphically captured datasets)以及模拟生成数据集(simulated datasets)。本节综述每一类别中的代表性数据集, 探讨其与 SNN 应用之间的关联性、数据生成或采集方法, 以及各类数据集的优势与局限性。

2.1　静态数据集

静态数据集源自传统计算机视觉与机器学习社区, 最初以静态图像或固定帧率视频的形式存在。为适应 SNN 的处理特性, 通常借助编码算法(如泊松编码或强度-延迟编码)将这类数据集的原始像素信息转换为脉冲序列。典型的数据集包括 MNIST[38]、CIFAR-10[39] 与 ImageNet[40], 它们为 SNN 提供了与 ANN 进行性能对比的基础平台, 尤其在图像分类任务中广泛用于模型验证与基准测试。生成脉冲数据的一般方法, 是通过强度-发放频率映射将每一图像帧转化为时空事件流。

此类数据集的主要优势在于规模庞大、标注完善且易于获取, 有利于经典算法的迁移与比较研究。然而, 其显著局限在于缺乏真实场景中的时序动态与事件稀疏性, 难以充分体现 SNN 在异步、事件驱动环境下的计算优势与生物合理性。

2.2　神经形态转换数据集

神经形态转换数据集是通过算法手段从传统动态视觉数据(如高帧率视频)转换而成的, 以模拟神经形态传感器的事件流输出。常见的代表包括 N-MNIST、N-CALTECH101[41] 等数据集。这类数据集为研究 SNN 在事件模态下的泛化能力与特征编码机制提供了重要桥梁, 其生成方式通常基于亮度变化检测与阈值判断, 将连续视频帧转化为异步事件序列。

这类数据集的优势在于部分保留了视觉动态信息, 比纯静态数据更接近实际事件相机的输出, 同时兼容主流评估流程, 但在转换过程中可能引入运动模糊、时间精度损失等人工伪影, 无法完全还原真实事件数据的噪声特性与时空统计结构。

2.3　神经形态捕获数据集

神经形态捕获数据集由专用神经形态传感器(如 DVS 和事件相机)直接采集得到, 以异步事件流格式记录视觉信息。典型的数据集包括 DVS128 Gesture[42]、Poker-DVS[43]、Gen1 Automotive[44] 以及 Scenes-DVS[45] 等。这类数据集最能体现 SNN 处理生物启发时空数据的优势, 被广泛用于对象追踪、动作识别、实时 SLAM 等任务。数据集通过事件相机捕获亮度变化而产生, 输出(x, y, 时间戳, 极性)格式的事件序列。

这类数据集的突出优势包括高时间分辨率、低时间延迟、高动态范围、低功耗以及数据的时空稀疏性。然而, 这类数据集也面临标注成本高、环境噪声显著(如背景活动事件)以及现有数据规模有限、多样性不足等挑战。

2.4　模拟生成数据集

模拟生成数据集通过合成的方式, 在仿真环境(或借助生成式算法)构建事件流数据, 常用于特定或极端场景下的算法验证。典型的数据集包括 ESIM[46]、DVS Gazebo[47]、PIX2NVS[48] 以及 Video-to-Events[49] 等。该类数据集适用于鲁棒性测试、噪声敏感度分析和极端光影或运动条件下的性能评估, 其生成方法依赖计算机图形渲染、物理模型和事件生成算法, 可精确调控事件密度、噪声水平和场景参数。

这类数据集在数据生成可控性、标注准确度及可扩充性方面存在明显优势, 适合于系统性验证与敏感度分析; 不足之处在于仿真与真实数据间可能存在域差异, 从而影响模型在真实环境中的泛化表现。

3 主流脉冲神经网络方法

本节主要概述面向视觉目标检测的脉冲神经网络方法，可分为两种：基于转换的 SNN（ANN-to-SNN）[2-5] 方法和基于直接训练的梯度替代方法（gradient surrogate）[7-9]。

3.1 基于转换的 SNN 方法

基于转换的 SNN 是一种将训练好的传统 ANN 转换为 SNN 的技术路径，其核心思想是利用激活值与发放频率的等效原理，将 ANN 中 ReLU 等激活函数的连续输出映射为 SNN 中神经元的脉冲发放频率。SNN 与 ANN 之间的模型转换机制，有效实现了预训练模型的能力迁移与高效复用，显著提升了 SNN 与现有主流深度学习框架的兼容性，同时大幅降低了模型训练与推理过程中的计算和存储资源开销。该方法不仅支持迁移学习与微调策略的顺利实施，还进一步增强了神经网络模型的生物合理性与可解释性。此外，SNN 固有的稀疏计算特性与事件驱动处理机制使其能够与神经形态硬件实现高度协同，从而显著推动了神经形态计算系统在扩展性与能效方面的提升。

基于 ANN 激活值与 SNN 脉冲发放频率近似等效的基本假设，研究者提出了多种 ANN-to-SNN 转换方法[2-5]，旨在充分利用深度神经网络已取得的性能优势。这类方法将连续激活函数输出的实值响应映射为离散脉冲神经元的发放行为，从而构建性能较高的脉冲神经网络模型（图 2）。

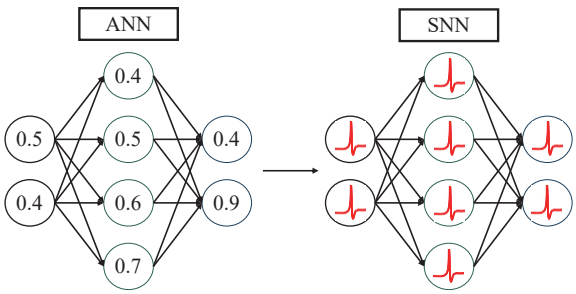

图 2 激活神经元与脉冲神经元的映射关系

Fig. 2 Mapping relationship between activation neurons and spiking neurons

3.1.1 基于 ReLU 激活函数的转换策略

Cao 等[2] 率先提出了一种转换策略，将使用 ReLU 激活函数且不含偏置项的 CNN 映射至 IF 神经元的 SNN 结构中。ReLU 激活函数的数学表达式为：

$$y(x)=\max(0,\sum W_i x_i) \tag{9}$$

式中，W_i 为突触权重，x_i 表示输入激活。IF 神经元的定义如下：

$$s(t)=\Theta[v(t-1)+\sum W_i s_i(t-1)-\theta] \tag{10}$$

式中，s 为输出，v 为膜电位，θ 为放电阈值，Θ 为 Heaviside 阶跃函数，其表达式如下：

$$\Theta(x)=\begin{cases}1, & \text{if } x\geqslant 0\\ 0, & \text{else}\end{cases} \tag{11}$$

研究表明，ReLU 激活函数在功能上等价于 IF 神经元模型。该模型可视为 LIF 神经元在特定参数设置下的简化形式：

$$y(x)\approx\frac{\sum_{i=1}^{T}s(t)}{T} \tag{12}$$

式中，T 为总时间步数。

3.1.2 通道归一化技术

继 Cao 等[2] 的开创性工作之后，Kim 等[3] 针对 SNN 在目标检测任务中存在的显著精度损失及脉冲发放频率与人工神经网络激活值分布不匹配的问题，提出了通道归一化技术。

该技术通过权重归一化和阈值调整两个核心步骤，显著提升了 SNN 的精度与能效。在权重归一化阶段，对卷积层每个输出通道的权重张量进行缩放，使其具有单位方差。具体实现时，计算各通道权重的标准差，并将权重除以该标准差，从而消除权重尺度的不一致性。这一操作不仅能够稳定训练过程，防止梯度爆炸或消失，还确保了脉冲神经元在推理时合理的发放脉冲频率。在阈值调整阶段，基于归一化后的权重，采用数据驱动的方式动态设置每个通道的脉冲发放阈值。通过统计训练数据中各通道的激活分布，将阈值设定为激活分布的特定百分位数（如 99%），从而在确保大多数输入能够有效触发脉冲的同时避免过度激活。

因为解决了不合理的阈值所导致的脉冲发放不稳定或能效低下问题，这种调整增强了网络的鲁棒性并显著降低了能耗。在 PASCAL VOC 数据集上的实验结果表明，该方法得到的全类平均正确率（mean Average Precision, mAP）性能仅比原 ANN 低 2%，却实现了数个数量级的能耗降低。然而，Spiking-YOLO 在神经形态硬件部署中仍需 2 500 个时序步才能满足实时检测需求，因此时间步数较

多是其在实际应用中的瓶颈。

3.1.3　脉冲校准与池化层优化方法

Li 等[4] 针对 SNN 在目标检测和分割任务中转换精度不足的问题, 从理论层面系统分析了转换误差的来源, 并提出了脉冲校准与池化层优化方法, 显著提升了 SNN 的转换精度与推理速度。

该方法将转换误差归纳为裁剪误差、未激活神经元脉冲(spike of inactive neurons, SIN)误差和最大池化误差, 并采用有针对性的措施逐项进行解决。在误差分析方面, 研究发现神经元的发放频率与其接收的突触电流随时间变化的峰值相关, 而非最终响应值。这一发现揭示了以往被忽视的错误脉冲触发机制, 为误差修正提供了理论依据。在误差修正阶段, 提出了脉冲校准(SpiCalib)方法, 通过监测神经元的平均脉冲间隔来识别 SIN 误差, 并对错误触发的脉冲进行动态校正。同时, 通过改进 LIPooling 方法, 实现了最大池化层的无损转换, 构建出完整的 SNN 转换框架。实验结果表明, 所提方法在分类、目标检测和分割任务方面均实现了最优异的性能, 且在 128 个时间步内即可在 PASCAL VOC 数据集上达到与 ANN 相当的检测与分割精度, 显著优于其他转换方法。

该工作首次实现了极低时间步数下 SNN 对复杂视觉任务的高性能适配, 为 SNN 的实际部署提供了有效的解决方案。

3.1.4　快速 SNN

快速 SNN(Fast-SNN)[5] 旨在实现竞争性的性能(与 ANN 相当)和低推理延迟, 以保持 SNN 的能效优势, 其核心思想是通过减小量化误差与累积误差来实现 ANN 到 SNN 的高效转换。

在量化误差最小化方面, 文献 [5] 揭示了 SNN 中的时间量化与 ANN 中的空间量化之间的等效映射关系。基于该映射, 通过量化神经网络的监督训练可有效降低量化误差, 具体包括寻找最优的裁剪范围以及每层权值和激活值的分布, 从而促进从 ANN 到 SNN 的高精度转换; 此外, 还推导了推理延迟的上界, 为量化从 ANN 到 SNN 的无损转换提供了理论保证。与此同时, 进一步提出了减少累积误差的方法。研究发现, 从 ANN 到 SNN 的转换过程中, 逐层的顺序误差是累积误差的主要来源。为此, 引入了带符号的中频神经元机制, 以抑制错误触发脉冲的影响, 降低单层顺序误差; 同时提出了分层微调策略, 以降低层间累积误差。实验结果表明, Fast-SNN 在包括图像分类、目标检测和语义分割在内的多个视觉任务上均表现出先进性能与低延迟。

该方法在保持高性能的同时显著降低了推理延迟, 为深度 ANN 模型向 SNN 模型的转换提供了有效的解决方案。

3.1.5　局限性与解决方案

基于转换的方法通过预训练 ANN 的参数来初始化 SNN, 可有效解决脉冲神经元不可微性带来的训练难题, 使研究者能够充分借鉴 ANN 领域丰富的优化策略与架构设计, 同时发挥 SNN 在推理阶段的能效优势[2]。然而, ANN 与 SNN 在表示形式及动力学机制上存在显著差异: ANN 处理连续的激活值, 而 SNN 则依赖于离散脉冲和时序动态。这导致该方法存在本质上的局限, 转换后的 SNN 往往无法充分利用脉冲时序中所蕴含的信息。

为缓解上述问题, 研究者们提出了多种策略, 包括使用软 LIF 激活函数来近似模拟脉冲行为、实施权重归一化以稳定膜电位以及引入软复位与阈值优化机制等[3-5]。尽管这些方法显著提升了转换性能与模型鲁棒性, 但间接学习仍难以完全捕捉 SNN 固有的时间计算特性。因此, 当前研究一方面致力于将时间学习机制嵌入 ANN 的训练阶段, 另一方面持续优化转换流程以更好地兼容脉冲动力学, 从而推动 SNN 在更复杂认知任务与实际场景中的应用。

3.2　基于直接训练的梯度替代方法

3.2.1　基本原理与 BPTT 训练策略

基于直接训练的梯度替代方法主要利用可微替代函数来克服脉冲发放函数在时间域上的不可微性, 从而解决误差反向传播过程中的梯度中断问题。通过引入连续可微的代理函数, 实现了对脉冲生成过程的平滑近似, 使标准优化算法(如随机梯度下降法或 Adam)能够应用于 SNN 的端到端训练中, 显著简化了训练流程并提高了可扩展性。

具体而言, 为应对脉冲激活函数的不连续性, 研究者通常采用替代梯度策略, 即使用一个连续可微函数的导数来近似地模拟脉冲非线性机制的导数。在该框架下, SNN 可被建模为一种具有二值脉冲输出的递归神经网络(recurrent neural network,

RNN)，并借助时间反向传播算法(backpropagation through time, BPTT)进行训练[1]。与 RNN 中梯度的链式传播机制类似，BPTT 将 SNN 在时间维度上展开，并将损失函数产生的梯度时间逐步反向传播至网络各层及突触权重。突触权重的更新可表示为：

$$\frac{\partial L}{\partial W}=\sum_{t=1}^{T}\frac{\partial L}{\partial s(t)}\frac{\partial s(t)}{\partial \vartheta(t)}\frac{\partial \vartheta(t)}{\partial W} \tag{13}$$

式中，L 为损失函数。

为避免不可微脉冲机制问题，替代梯度 $\frac{\partial s(t)}{\partial \vartheta(t)}$ 通过可微函数对真实的脉冲发放机制进行平滑逼近，从而在数学上确保梯度的可计算性与训练的稳定性。例如，使用一种线性函数作为代理梯度，用于对脉冲发放函数(Heaviside 阶跃函数)的导数的近似，其数学表达式如下：

$$y(x)=\begin{cases}x-\theta+m, & \text{if } \theta\leqslant x\leq\theta+m\\ -x+\theta+m, & \text{else if } \theta-m\leqslant x<\theta\\ 0, & \text{else}\end{cases} \tag{14}$$

其中，θ 为放电阈值，m 为常数。

3.2.2 SNN 的梯度传播挑战与改进方法

深度 SNN 的监督学习面临梯度消失与爆炸的共性挑战。在误差反向传播过程中，梯度可能因层数或时间步的增加而呈指数级衰减(消失)或增长(爆炸)的趋势，从而导致权重更新失效或训练过程不稳定[1]。这一问题并非 SNN 所独有，而是深度神经网络(包括传统 ANN)中普遍存在的现象[50]。然而，在 SNN 中，由于时间维度的引入，这一问题进一步加剧。

BPTT 将传统反向传播算法扩展至 SNN 的时间动力学领域，通过计算脉冲序列的时间导数并沿时间维度反向传播误差来优化网络权重。SpikeProp 算法[51]作为首个基于脉冲发放时刻的监督学习规则，借鉴了第二代 ANN 的误差反向传播思想，适用于多层 SNN 的分类任务，其误差函数定义为期望脉冲时间与实际输出脉冲时间的均方误差：

$$E=\frac{1}{2}\sum_{j}(t_{j,\ \text{actual}}-t_{j,\ \text{desired}})^2 \tag{15}$$

式中，$t_{j,\ \text{actual}}$ 和 $t_{j,\ \text{desired}}$ 分别为实际与期望的脉冲时间。该误差信号通过网络层反向传播以调整突触权重。为提升原算法性能，研究者提出了多种改进版本，包括 QuickProp[52]、弹性传播[52]以及带动量的反向传播[53]。

3.2.3 EMS-YOLO：直接训练的 SNN 目标检测框架

继文献[2, 51-53]之后，Su 等[7]针对 SNN 在目标检测任务中依赖 ANN-SNN 转换导致的精度损失、推理延迟高及能效优化不足问题，提出了一种直接训练的 SNN 目标检测框架。该方法通过能效膜电位捷径残差网络(Energy-efficient Membrane-Shortcut Residual Network, EMS-ResNet)和极低时间步训练策略，构建了高性能、低功耗的脉冲驱动检测模型。

在网络结构设计阶段，构建了全脉冲化的 EMS-ResNet 主干网络，引入膜电位捷径机制来优化能量动态。具体实现过程中，通过膜电位短路连接来替代传统残差结构中的恒等映射，在脉冲域直接传递膜电位信息，从而减少来脉冲发放频率。这一设计不仅降低了突触操作的能耗，还通过理论分析验证了其避免梯度消失或爆炸的能力，可支持深层 SNN 的稳定训练。在训练策略优化阶段，采用直接训练方法替代 ANN-SNN 转换，通过梯度替代函数来实现端到端优化。模型仅需 4 个时间步即可完成推理，显著压缩了时序冗余。同时，基于 LIF 神经元膜电位特性来动态调整阈值参数，使脉冲发放与目标检测的复杂特征提取需求相匹配，从而实现了检测精度与能效之间的平衡。

实验结果表明，EMS-YOLO 在 COCO 和 Gen1 数据集上达到了与同结构 ANN 相当的检测精度，同时能耗降至原来的 1/5.83，并在仅 4 个时间步下实现了实时推理。然而，该模型对高分辨率多尺度目标的检测鲁棒性仍弱于成熟的 ANN 模型，且直接训练所需的梯度替代函数在复杂场景中的稳定性仍需进一步优化，这也是其未来在实际应用中将面临的关键挑战。

3.2.4 Meta-SpikeFormer：Transformmer SNN 架构

Yao 等[8]针对 SNN 在复杂视觉任务中性能不足、多任务泛化能力弱及硬件部署效率低的问题，提出了一种元 Transformer SNN 架构(Meta-SpikeFormer)。该方法通过宏观模块设计与微观算子优化，构建了高性能、低功耗且支持多任务的脉冲驱动 Transformer 模型，并为下一代神经形态芯片的设计提供了指导。

在宏观架构设计阶段，构建了卷积基与 Transformer 基混合的 SNN 模块，全程采用稀疏加

法运算以降低计算开销。具体实现过程中, 通过模块化组合来兼顾局部特征提取与全局依赖建模, 并引入短路连接以确保梯度流动, 避免深层网络训练中的梯度消失或爆炸问题。这一设计不仅提升了模型表达能力, 还增强了其对不同视觉任务的适应性。在微观算子创新阶段, 提出了多种脉冲卷积方法、脉冲 Query/Key/Value(QS/KS/VS)生成机制, 以及 3 种不同计算复杂度的脉冲驱动自注意力算子。通过动态调整算子结构与脉冲发放模式, 模型能够根据任务需求平衡精度与能效, 尤其在资源受限场景下仍能保持高效推理。

实验结果表明, Meta-SpikeFormer 在 ImageNet-1K 分类任务中首次将 SNN 准确率提升至 80%, 较 SOTA 基线提高 3.7%, 且参数量减少了 17%。同时, 在 COCO 检测、ADE20K/VOC2012 分割及 HAR-DVS 事件识别任务中均达到了 SNN 领域的最优性能。此外, 其架构的硬件友好特性显著优于传统卷积 SNN, 为 Transformer 基神经形态芯片的开发提供了关键技术支撑。然而, 该模型在极低时间步(如 4 步)下的实时推理能力仍有待进一步优化, 当前在多任务并行处理时仍需较高的脉冲计算密度, 这也是未来实际部署中需要突破的瓶颈之一。

3.2.5　SpikeYOLO: 整数值训练与脉冲驱动推理优化

2025 年 Luo 等[9] 提出的 SpikeYOLO 采用了一种结合 YOLO 宏观架构与 SNN 微观模块的设计策略。该研究认为, YOLO 系列中的复杂模块虽在 ANN 中有效, 却不适用于 SNN, 易导致脉冲退化现象, 表现为深层神经元脉冲发放几乎停止。因此, SpikeYOLO 在设计中简化了模块结构, 仅保留 YOLOv8 的宏观框架, 并采用 Meta-SpikeFormer[8] 中的元 SNN 模块作为基础构建单元, 以进行微观层面的再设计。

针对 SNN 在目标检测任务中存在的训练效率低、硬件部署复杂度高及能效比不足的问题, 提出了整数值训练与脉冲驱动推理协同优化的 SNN 框架。该方法通过整数值权重约束和事件驱动计算机制, 实现了高精度、低功耗的脉冲检测模型。在整数值训练阶段, 对网络权重施加整数值离散化约束, 将全精度权重映射至低比特整数空间。具体实现时, 通过可微量化函数在训练中同步优化权重整数值与脉冲神经元参数, 结合梯度裁剪与直通估计器保持训练稳定性。这一操作不仅大幅降低了权重存储与传输开销, 还使模型天然适配整数运算单元, 为神经形态硬件的部署奠定了基础。在脉冲驱动推理阶段, 基于整数值权重构建纯事件驱动的推理流水线, 避免浮点运算。通过动态膜电位积分与阈值触发机制, 仅在输入脉冲到达时更新神经元状态, 显著减少了冗余计算。同时, 采用时序脉冲编码压缩特征信息, 使模型在极低时间步内完成高置信度检测。

实验结果表明, 该方法在 PASCAL VOC 和 COCO 数据集上分别达到了与同结构 ANN 相当的 mAP 性能, 同时能耗降低了两个数量级, 在 8 位整数量化下仅需 16 个时间步即可实现实时检测。然而, 整数值训练对超参数的敏感性较高, 且低比特量化在复杂场景中可能引发梯度失配问题, 当前仍需依赖部分浮点运算进行训练校准, 这些是未来实现端到端整数化部署的关键挑战。

3.2.6　小结

总之, 基于直接训练的梯度替代方法通过引入连续可微的代理函数(如分段线性函数)来近似模拟脉冲发放函数的导数, 克服了脉冲神经元在时间域上的不可微性, 使误差能够通过时间反向传播在 SNN 中有效传递。这一方法将 SNN 建模为递归网络, 并利用随机梯度下降或 Adam 等优化算法实现端到端训练, 显著简化了流程并提升了可扩展性。该方法在目标检测任务中展现了能效与实时性优势, 但仍受限于梯度传播效率、架构适配性与数据稀缺。

3.3　主流方法性能对比

表 1 系统总结了基于转换的 SNN 方法与基于直接训练的梯度替代方法在 COCO 2017 val 数据集[54] 上的性能对比结果。其中, $T\times D$ 表示总时间步数 T 与每时间步展开次数 D 的乘积。在传统 SNN 与 ANN 模型中, D 通常默认取 1。ANN 与 SNN 的能耗可分别通过以下公式进行估算:

$$E_{\mathrm{ANN}} = O^2 C_{\mathrm{in}} C_{\mathrm{out}} k^2 E_{\mathrm{MAC}} \tag{16}$$

$$E_{\mathrm{SNN}} = (T\times D) f_{\mathrm{r}} O^2 C_{\mathrm{in}} C_{\mathrm{out}} k^2 E_{\mathrm{AC}} \tag{17}$$

式 (16) 和 (17) 中, O 为特征输出大小, C_{in} 和 C_{out} 分别为输入通道和输出通道的数量, k 为核大小, f_{r} 为平均脉冲触发率。本文采用 SNN 领域最常用的能耗评估方法[12]。所有操作均假定采用 45 nm 技术

表 1 在 COCO 2017 val 数据集上的最新结果

Table 1 The latest results on the COCO 2017 validation dataset

Architecture	Model	Parameter count/10^7	Power/mJ	$T\times D$	mAP@50/%	mAP@50:95/%
ANN	PVT-tiny[10]	3.29	520.3	1	59.2	36.7
	DETR-DC5[11]	4.10	197.8	1	62.4	42.0
	YOLOv5[12]	2.12	112.5	1	64.1	45.4
ANN-to-SNN	Spike-YOLO[3]	1.02	—	3 500	—	25.7
	Spike Calib[4]	1.71	—	512	45.3	—
	Fast-SNN[5]	2.51	—	15	46.4	—
Directly-trained	Spike Retina[6]	1.13	—	4×1	28.5	—
	EMS-YOLO[7]	2.69	29.0	4×1	50.1	30.1
	Meta-SpikeFormer[8]	1.68	70.7	4×1	50.3	30.2
	SpikeYOLO[9]	2.31	34.6	1×4	62.3	45.5
		6.88	84.2	1×4	66.2	48.9

的 32 位浮点来实现，其中，E_{MAC} = 4.6 pJ，E_{AC} = 0.9 pJ[55]。由公式 (16) 和 (17) 可知，SNN 的低功耗特性源于其稀疏加法运算。

如表 1 所示，基于直接训练的 SpikeYOLO[9] 模型在 COCO 数据集上显著提升了 SNN 的性能上限，其 mAP@50 和 mAP@50:95 分别达到 66.2% 和 48.9%，较先前的最优 SNN 方法[8] 分别提升了 15.9% 和 18.7%。在参数量与能耗方面，基于直接训练的 SpikeYOLO 同样表现出明显优势，SpikeYOLO 与 EMS-YOLO[7] 的参数量分别为 2.31×10^7 和 2.69×10^7，mAP@50 分别为 62.3% 和 50.1%，mAP@50:95 分别为 45.5% 和 30.1%，能耗分别为 34.6 mJ、29.0 mJ。此外，SNN 与 ANN 之间的性能差距显著缩小，在参数量相近（2.31×10^7 与 2.12×10^7）的情况下，基于直接训练的 SpikeYOLO 与 YOLOv5[12] 性能相当，且后者的能效为前者的 3.3 倍。

直接训练 SNN 通过端到端的梯度优化框架，有效规避了 ANN-to-SNN 转换过程中存在的分布偏移问题，从而显著提升了模型在复杂任务（如目标检测、多模态推理）中的精度。该方法的能效优势尤为突出，在神经形态硬件上的能耗可降至等效 ANN 的 1/70，特别适用于实时边缘计算场景（如自动驾驶感知和无人机视觉）。此外，直接训练支持动态神经元机制，能够根据输入特征动态调整脉冲发放模式，增强时空特征表达能力。尽管直接训练 SNN 表现出在能效和精度方面的优势，但其训练过程仍存在显著的复杂性。代理梯度函数的引入虽然解决了脉冲不可微问题，但可能引入梯度偏差，导致训练不稳定或收敛困难。同时，超参数（如代理函数宽度 γ 和神经元阈值）对性能敏感，需针对不同任务进行精细调优。此外，直接训练目前缺乏成熟的标准化工具链，开发者需自行实现时空梯度传播并适配硬件平台，提高了开发成本和门槛。

ANN-to-SNN 转换方法通过复用预训练 ANN 的权重和架构，无须处理时序梯度传播或代理函数设计，极大地简化了 SNN 的训练流程并降低了开发复杂度。该方法在充分长的时序步下能逼近 ANN 的精度，转换过程完全兼容现有 ANN 工具链（如 PyTorch 和 TensorFlow），支持对经典模型（如 ResNet、VGG）的直接部署，降低了算法迁移成本。此外，转换后的 SNN 模型在理论上保留了 ANN 的特征表示能力，适用于对精度要求极高但对延迟不敏感的应用场景。ANN-to-SNN 转换方法的核心缺陷在于高延迟和能效瓶颈。为了模拟 ANN 的连续激活值，需要大量时序步进行脉冲发放积分，导致推理速度大幅下降，甚至在能耗方面可能反超等效 ANN。此外，转换过程还面临分布偏移问题，即 ANN 的激活分布与 SNN 的脉冲发放频率之间存

在统计差异, 需通过额外的权重归一化、阈值校准等后处理手段进行补偿, 且校准效果对复杂任务(如目标检测和分割)的泛化性较差。转换后的 SNN 脉冲模式固定, 难以充分利用神经形态硬件的时空稀疏性和事件驱动特性, 在芯片部署中的能效提升有限(通常仅 10~20 倍), 远不如直接训练 SNN 的 60~100 倍的能效优势。

3.4　SNN 在神经形态硬件部署中的核心挑战

尽管 SNN 在能效和事件驱动计算方面具备理论优势, 但其在神经形态硬件上的实际部署仍面临多重挑战。以下结合 Loihi[13] 与 TrueNorth[14] 芯片的实践案例展开分析。

SNN 的核心优势在于其事件驱动的稀疏计算特性, 这种稀疏性对内存访问模式提出了特殊挑战。由于脉冲激活的神经元位置在时序上的动态变化, 内存访问地址呈现不连续、不规则的特征, 难以充分利用传统硬件中基于空间局部性的优化策略(如缓存预取机制)。以 Loihi 芯片为例, 其通过分层核心结构与分布式权重内存应对这一挑战: 每个神经核心在本地存储突触权重, 仅在与脉冲事件相关的核心之间触发数据交换, 从而有效降低全局内存带宽的压力。相比之下, TrueNorth 采用交叉开关(Crossbar)硬件结构来映射突触连接, 通过事件路由表将脉冲直接分发至目标神经元, 避免了显式的内存寻址操作, 但其静态连接模式限制了动态稀疏性的利用效率。

SNN 中神经元模型(如 LIF)的微分方程运算需要持续更新膜电位, 这对硬件计算单元的设计提出了较高要求。TrueNorth 为每个神经元分配了专用逻辑单元, 但固定硬件资源难以灵活适配不同网络结构的动态计算需求, 导致资源利用率低下或扩展性受限。Loihi 则通过可编程微码引擎支持用户的自定义神经元动力学, 在灵活性与计算密度之间寻求平衡, 但同时增加了硬件设计的复杂度。另一方面, 神经元状态的数值精度对硬件实现至关重要。Loihi 支持可配置的膜电位精度(1~24 位), 但低比特量化可能引发梯度失配问题, 例如整数值训练中常见的量化误差累积。

4　总结

本文系统地综述了面向视觉目标检测任务的脉冲神经网络的核心方法与架构研究进展。在训练方法层面, 基于转换的 SNN 通过复用预训练 ANN 的权重和架构, 显著降低了开发复杂度, 并可在长时序步下逼近 ANN 的精度, 但其对高时序步的依赖导致推理延迟的大幅增加, 同时能效提升有限, 且存在激活分布与脉冲发放频率不匹配的固有问题。相比之下, 基于直接训练的梯度替代方法通过代理梯度函数和 BPTT 实现端到端优化, 可在低时序步下达到高精度和极致能效, 但面临训练内存占用高、代理函数超参数敏感及缺乏标准化工具链的挑战。在网络架构方面, SNN 从传统 CNN 架构扩展至 Transformer 架构, 并通过简化模块设计缓解脉冲退化问题。能耗分析表明, SNN 的能效优势源于事件驱动的稀疏计算, 但其实际性能高度依赖于训练方法与硬件协同优化水平。

5　展望

未来 SNN 研究需聚焦算法、硬件与应用的深度融合, 以突破当前瓶颈并拓展其实际部署潜力。在训练方法上, 应探索混合范式(如直接训练初始化结合转换微调)和自适应代理梯度机制(如动态宽度 γ 函数), 以在训练效率与精度之间实现更好的平衡; 在架构设计上, 需开发脉冲专用模型(如稀疏注意力 Transformer 和多尺度动态神经元), 避免 ANN 模块移植导致的脉冲退化, 并增强对时序事件数据的建模能力。硬件协同设计是值得关注的关键研究方向之一, 需通过神经形态芯片的联合优化(如适配脉冲稀疏性的内存布局与路由策略)和标准化能效评估体系, 以实现算法到硬件的端到端高效映射。应用场景需超越分类与检测, 向语义分割、多模态融合(视觉-语言)及高实时需求领域(如安防、自动驾驶、无人机导航、医学影像)拓展。同时, 必须建设集成 BPTT、代理梯度与部署功能的开源工具链并开发跨平台部署方案, 降低开发门槛并促进生态成熟, 最终推动 SNN 成为具有低功耗、高实时性、强泛化能力的下一代神经形态智能核心。

参考文献:

[1] NEFTCI E O, MOSTAFA H, ZENKE F. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks[J]. IEEE Signal Processing Magazine, 2019, 36(6): 51-63.

[2] CAO Y, CHEN Y, KHOSLA D. Spiking deep convolutional neural networks for energy-efficient object recognition[J]. International Journal of Computer Vision, 2015, 113(1): 54-66.

[3] KIM S, PARK S, NA B, et al. Spiking-YOLO: Spiking neural network for energy-efficient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11270-11277.

[4] LI Y, HE X, DONG Y T, et al. Spike calibration: Fast and accurate conversion of spiking neural network for object detection and segmentation[EB/OL]. (2022-07-01). https://arxiv.org/pdf/2207.02702.

[5] HU Y, ZHENG Q, JIANG X, et al. Fast-SNN: Fast spiking neural network by converting quantized ANN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(12): 14546-14562.

[6] ZHANG H, LI Y, HE B, et al. Direct training high-performance spiking neural networks for object recognition and detection[J]. Frontiers in Neuroscience, 2023, 17: 1229951.

[7] SU Q, CHOU Y, HU Y, et al. Deep directly-trained spiking neural networks for object detection[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2023: 6555-6565.

[8] YAO M, HU J K, HU T X, et al. Spike-driven transformer V2: Meta spiking neural network architecture inspiring the design of next-generation neuromorphic chips[EB/OL]. (2024-02-15). https://arxiv.org/pdf/2404.03663.

[9] LUO X, YAO M, CHOU Y, et al. Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection[C]// Proceedings of the European Conference on Computer Vision. Cham: Springer, 2025.

[10] WANG W, XIE E, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 548-558.

[11] YAO M, ZHAO G, ZHANG H, et al. Attention spiking neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 9393-9410.

[12] Ultralytics. YOLOv5[CP/OL]. (2022-11-22). https://github.com/ultralytics/yolov5.

[13] DAVIES M, SRINIVASA N, LIN T H, et al. Loihi: A neuromorphic manycore processor with on-chip learning[J]. IEEE Micro, 2018, 38(1): 82-99.

[14] MEROLLA P A, ARTHUR J V, ALVAREZ-ICAZA R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. Science, 2014, 345(6197): 668-673.

[15] SCHUETZE S M. The discovery of the action potential[J]. Trends in Neurosciences, 1983, 6: 164-168.

[16] GERSTNER W, KISTLER W M. Spiking neuron models: Single neurons, populations, plasticity[M]. Cambridge: Cambridge University Press, 2002.

[17] JAHN R, SUDHOF T C. Synaptic vesicles and exocytosis[J]. Annual Review of Neuroscience, 1994, 17: 219-246.

[18] BROWN D A, SIHRA T S. Presynaptic signaling by heterotrimeric G-proteins[C]//Pharmacology of neurotransmitter release. Berlin, Heidelberg: Springer, 2008: 207-260.

[19] FENG J . Computational neuroscience: A comprehensive approach[M]. Boca Raton: CRC Press, 2003.

[20] DAYAN P, ABBOTT L F. Theoretical neuroscience: Computational and mathematical modeling of neural systems[M]. Cambridge: The MIT Press, 2005.

[21] LI C, JONES E G, FURBER S. Unleashing the potential of spiking neural networks with dynamic confidence[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2023: 13304-13314.

[22] IZHIKEVICH E M. Which model to use for cortical spiking neurons?[J]. IEEE Transactions on Neural Networks, 2004, 15(5): 1063-1070.

[23] HODGKIN A L, HUXLEY A F. A quantitative description of membrane current and its application to conduction and excitation in nerve[J]. The Journal of Physiology, 1952, 117(4): 500-544.

[24] HAN B, SRINIVASAN G, ROY K. RMP-SNN: Residual membrane potential neuron for enabling deeper high-accuracy and low-latency spiking neural network[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 13555-13564.

[25] NIU L Y, WEI Y, LIU Y. Event-driven spiking neural network based on membrane potential modulation for remote sensing image classification[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106322.

[26] DIEHL P U, NEIL D, BINAS J, et al. Fast-classifying, high-

accuracy spiking deep networks through weight and threshold balancing[C]//Proceedings of the 2015 International Joint Conference on Neural Networks. New York: IEEE, 2015: 1-8.

[27] QIU X, ZHU R J, CHOU Y, et al. Gated attention coding for training high-performance and efficient spiking neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2024, 38: 601-610.

[28] RUECKAUER B, LUNGU I A, HU Y, et al. Theory and tools for the conversion of analog to spiking convolutional neural networks[EB/OL]. (2016-12-15). https://doi.org/10.48550/arXiv.1612.04052.

[29] DONATI E, PAYVAND M, RISI N, et al. Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network[J]. IEEE Transactions on Biomedical Circuits and Systems, 2019, 13(5): 795-803.

[30] DIEHL P U, COOK M. Unsupervised learning of digit recognition using spike-timing-dependent plasticity[J]. Frontiers in Computational Neuroscience, 2015, 9: 99.

[31] RUECKAUER B, LUNGU I A, HU Y, et al. Conversion of continuous-valued deep networks to efficient event-driven networks for image classification[J]. Frontiers in Neuroscience, 2017, 11: 682.

[32] KHERADPISHEH S R, GANJTABESH M, THORPE S J, et al. STDP-based spiking deep convolutional neural networks for object recognition[J]. Neural Networks, 2018, 99: 56-67.

[33] MASQUELIER T, THORPE S J. Unsupervised learning of visual features through spike timing dependent plasticity[J]. PLOS Computational Biology, 2007, 3(2): e31.

[34] YANG M, LIU S C, DELBRUCK T. A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding[J]. IEEE Journal of Solid-State Circuits, 2015, 50(9): 2149-2160.

[35] ZHOU S, LI X, CHEN Y, et al. Temporal-coded deep spiking neural network with easy training and robust performance[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11143-11151.

[36] PAN Z, WU J, ZHANG M, et al. Neural population coding for effective temporal classification[C]//Proceedings of the 2019 International Joint Conference on Neural Networks. New York: IEEE, 2019: 1-8.

[37] LICHTSTEINER P, POSCH C, DELBRUCK T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor[J]. IEEE Journal of Solid-State Circuits, 2008, 43(2): 566-576.

[38] DENG L. The MNIST database of handwritten digit images for machine learning research[J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.

[39] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009.

[40] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2009: 248-255.

[41] ORCHARD G, JAYAWANT A, COHEN G K, et al. Converting static image datasets to spiking neuromorphic datasets using saccades[J]. Frontiers in Neuroscience, 2015, 9: 437.

[42] AMIR A, TABA B, BERG D, et al. A low power, fully event-based gesture recognition system[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 7388-7397.

[43] PÉREZ-CARRASCO J A, ZHAO B, SERRANO C, et al. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing: Application to feedforward ConvNets[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2706-2719.

[44] DE TOURNEMIRE P, NITTI D, PEROT E, et al. A large scale event-based detection dataset for automotive[EB/OL]. (2020-01-21). https://doi.org/10.48550/arXiv.2001.08499.

[45] NEGRI P, SOTO M, LINARES-BARRANCO B, et al. Scene context classification with event-driven spiking deep neural networks[C]//Proceedings of the 25th IEEE International Conference on Electronics, Circuits and Systems. New York: IEEE, 2018: 569-572.

[46] REBECQ H, GEHRIG D, SCARAMUZZA D. ESIM: An open event camera simulator[C]//Proceedings of the 2nd Conference on Robotics Learning. Zürich: PMLR, 2018.

[47] KAISER J, VASQUEZ TIECK J C, HUBSCHNEIDER C, et al. Towards a framework for end-to-end control of a simulated vehicle with spiking neural networks[C]//Proceedings of the 2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots. New York: IEEE, 2016: 127-134.

[48] BI Y, ANDREOPOULOS Y. PIX2NVS: Parameterized

conversion of pixel-domain video frames to neuromorphic vision streams[C]//Proceedings of the 2017 IEEE International Conference on Image Processing. New York: IEEE, 2017: 1990-1994.

[49] GEHRIG D, GEHRIG M, HIDALGO-CARRIO J, et al. Video to events: Recycling video datasets for event cameras[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 3583-3592.

[50] MENG Q, XIAO M, YAN S, et al. Training high-performance low-latency spiking neural networks by differentiation on spike representation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 12444-12453.

[51] BOHTE S M, KOK J N, LA POUTRÉ J A. SpikeProp: Backpropagation for networks of spiking neurons[C]// Proceedings of the European Symposium on Artificial Neural Networks. Bruges: d-side publications, 2000: 11-16.

[52] MCKENNOCH S, LIU D, BUSHNELL L G. Fast modifications of the SpikeProp algorithm[C]//Proceedings of the 2006 IEEE International Joint Conference on Neural Network. New York: IEEE, 2006: 3970-3977.

[53] XIN J, EMBRECHTS M J. Supervised learning with spiking neural networks[C]//Proceedings of the IJCNN'01. International Joint Conference on Neural Networks. New York: IEEE, 2001: 1772-1777.

[54] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference. Berlin, Heidelberg: Springer: 740-755.

[55] HOROWITZ M. 1.1 computing's energy problem (and what we can do about it)[C]//Proceedings of the 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers. New York: IEEE, 2014: 10-14.