Late Breaking Results: Improving Deep SNNs with Gradient Clipping and Noise Exploitation in Neuromorphic Devices

Seongsik Park*, Jongkil Park, Hyun Jae Jang, Jaewook Kim, YeonJoo Jeong,
Gyu Weon Hwang, Inho Kim, Jong-Keuk Park, Kyeong Seok Lee, and Suyoun Lee
Center for Semiconductor Technology, *Korea Institute of Science and Technology*, Seoul, Korea

***Abstract*—Deep spiking neural networks (SNNs) have shown remarkable progress due to improvements, such as training algorithms. However, most of them have not considered the features of neuromorphic devices. Their improvement has relied on soft resets, which are computationally expensive and unsuitable for neuromorphic devices. To address this, this paper proposes gradient clipping in hard reset-based deep SNNs and explores how device noise enhances learning performance. According to our experiments on various datasets and models, the proposed approach improved the training performance of deep SNNs with hard reset. These findings bridge gaps between SNN algorithms and hardware constraints, paving the way for efficient neuromorphic computing.**

***Index Terms*—spiking neural networks, neuromorhpic device, hard reset, gradient clipping**

I. Introduction

Neuromorphic computing, a next-generation computing paradigm inspired by human neural behavior, has gained significant attention for its potential to enable low-power artificial intelligence [1]. To realize ultra-low-power neuromorphic computing, operations for spiking neurons and synapses with neuromorphic devices are crucial. However, neuromorphic devices have faced challenges, such as imprecise computations and device variations that affect key parameters like threshold voltage and reset potential [2].

Recently, with improvements in training algorithms, such as spatio-temporal backpropagation (STBP) with surrogate gradient [3], deep spiking neural networks (SNNs) have achieved remarkable performance in various tasks [4], [5]. However, most deep SNNs have overlooked the characteristics of neuromorphic devices. Many of them have employed a soft reset mechanism, which introduces significant overhead with subtraction operations on the neuromorphic devices. Moreover, the impact of device variations and noise on training performance remains under-explored.

In this work, we address these gaps by analyzing the limitations of the hardware-friendly hard reset and proposing gradient clipping as an effective method to improve training stability and performance. Additionally, we have demonstrated that the noise in neuromorphic devices can enhance the training performance of deep SNNs. Our approach has built on STBP with surrogate gradients [3] and tdBN [5]. Experimental results on various datasets and model architectures have validated the effectiveness of the proposed methods, and further analysis has revealed how the proposed approach improved training performance.

*corresponding author: Seongsik Park (seong.sik.park@kist.re.kr)

II. Preliminary and Related Works

Deep SNNs have mostly adopted simplified neuron models, such as the leaky integrated-and-fire (LIF) neurons. The neuronal dynamics of an LIF neuron are defined as:

$$u_i^l[t] = 1/\tau(v_i^l[t-1] - v_{\text{reset},i}^l + \sum\nolimits_j w_{ji}^l s_j^{l-1}[t]), \qquad (1)$$

where $u_i^l[t]$ is a membrane potential at i-th neuron in l-th layer on time step t. τ, v, v_{reset}, and w_{ji} are a leaky constant, intermediate membrane, reset potential, and synaptic weight, respectively. Spike s is determined by $s_i^l[t] = H(u_i^l[t] - v_{\text{th},i}^l)$, where H is the Heaviside step function and V_{th} is a threshold voltage. After a spike, the membrane potential resets as:

$$v_i^l[t] = u_i^l[t](1 - s_i^l[t]) + g(u_i^l[t], v_{\text{reset},i}^l)s_i^l[t], \qquad (2)$$

where $g(\cdot)$ is a reset function.

To train deep SNNs, we adopted a gradient-based direct training approach, including STBP with surrogate gradients [3] and tdBN [5]. According to the STBP, gradients are given by

$$\frac{\partial L}{\partial u_j^l[t]} = \frac{\partial L}{\partial s_j^l[t]}\frac{\partial s_j^l[t]}{\partial u_i^l[t]} + \frac{\partial L}{\partial u_j^l[t+1]}\frac{\partial u_j^l[t+1]}{\partial u_j^l[t]}, \qquad (3)$$

$$\frac{\partial L}{\partial s_j^l[t]} = \sum_k \frac{\partial L}{\partial u_k^{l+1}[t]}\frac{\partial u_k^{l+1}[t]}{\partial s_j^l[t]} + \frac{\partial L}{\partial u_j^l[t+1]}\frac{\partial u_j^l[t+1]}{\partial s_j^l[t]}. \qquad (4)$$

III. Methods

The reset function $g(\cdot)$ (Eq. 2) is categorized into two types

$$g(u_i^l[t], v_{\text{reset},i}^l) = \begin{cases} u_i^l[t] - v_{\text{th},i}^l[t], & \text{(soft reset)} \\ v_{\text{reset},i}^l, & \text{(hard reset).} \end{cases} \qquad (5)$$

Based on this definition with Eqs. 1 and 2, the impact of the reset function on the gradient calculation in Eq. 4 is given as

$$\frac{\partial u_j^l[t+1]}{\partial s_j^l[t]} = \begin{cases} -(1/\tau)v_{\text{th},j}^l, & \text{(soft reset)} \\ -(1/\tau)(u_j^l[t] - v_{\text{reset},j}^l), & \text{(hard reset).} \end{cases} \qquad (6)$$

This term represents the change in the membrane potential due to the reset operation when a spike is fired. As shown in Eq. 6,

978-3-9826741-0-0/DATE25/

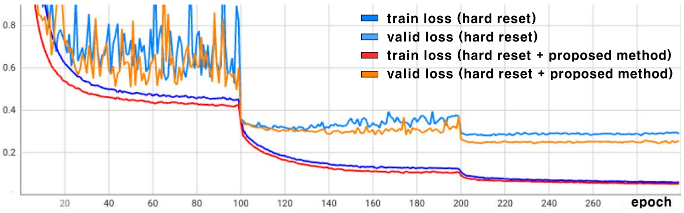


Fig. 1: Training curves - loss (ResNet20, CIFAR10)

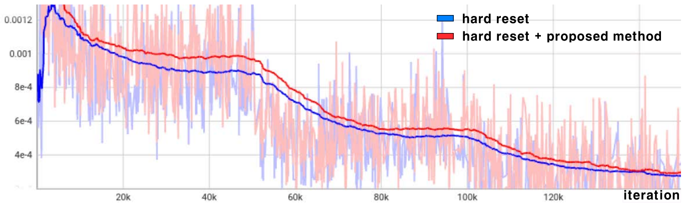


Fig. 2: Gradient standard deviation (ResNet20, CIFAR10)

the hard reset mechanism can cause abrupt changes in gradient, which may negatively impact training performance compared to the soft reset mechanism. To address this issue and prevent sudden gradient changes that destabilize training, we propose gradient clipping as follows

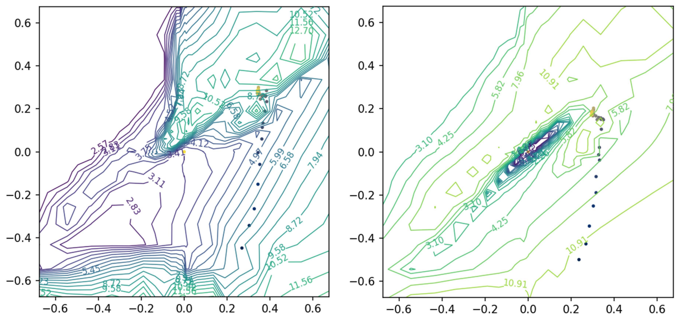
$$\frac{\partial u_j^l[t+1]}{\partial s_j^l[t]} = -(1/\tau)\min(u_i^l[t], v_{\text{th},j}^l). \tag{7}$$

By limiting the gradient changes induced by the reset, gradient clipping enables more stable training. Furthermore, to investigate the effects of neuromorphic device variations, we modeled variations in v_{th} and v_{reset} by introducing noise. Random noise was added to these parameters, following a normal distribution $N(0, \sigma)$, where σ was determined empirically.

IV. Experimental Results

We evaluated the performance of the proposed method using VGG16 and ResNet20 for static image classification datasets (CIFAR10 and CIFAR100) and VGG11 for the neuromorphic dataset (CIFAR10-DVS). For noise modeling, we introduced Gaussian noise with standard deviations σ of 0.001 for v_{th} and 0.007 for v_{reset}. To ensure accurate comparisons, we reported the average results over four repetitions. The training curves (loss) of the gradient clipping method and the conventional method with the hard reset are shown in Fig. 1. Across all training and validation scenarios, the proposed method consistently demonstrates lower loss than the conventional model. This indicates that gradient clipping helps achieve a more optimized solution by preventing abrupt gradient changes caused by the reset operation in spiking neurons.

To analyze the effect of the proposed method on gradients, we measured the standard deviation of gradients during training (Fig. 2). The results show that gradient clipping leads to more diverse gradients, which contributes to improved training. We investigated the loss landscape to further understand the improvements. Fig. 3 shows that the loss landscape becomes smoother when the proposed method is applied, indicating more stable training. The overall experimental results are in Table I. Experiments across various datasets and models demonstrate that gradient clipping improves the training of hard reset-based SNNs. Furthermore, the noise in the neuromorphic device can enhance the training performance and robustness.



(a) Hard reset (b) Hard reset + proposed method

Fig. 3: Loss landscape (ResNet20, CIFAR10)

TABLE I: Test accuracy and the number of spikes

Dataset	Model	Method	Reset	Noise	Acc.(%)	Spk (K)
CIFAR10	VGG16	tdBN	Soft	X	93.39±0.09	148±24
		tdBN	Hard	X	93.18±0.18	143±2.1
		Ours	Hard	X	93.53±0.23	148±1.9
		Ours	Hard	O	93.55±0.11	148±1.9
	ResNet20	tdBN	Soft	X	95.02±0.10	496±29
		tdBN	Hard	X	93.59±0.14	476±8.2
		Ours	Hard	X	94.41±0.48	493±3.4
		Ours	Hard	O	94.71±0.12	492±9.4
CIFAR100	VGG16	tdBN	Soft	X	69.30±0.11	169±4.2
		tdBN	Hard	X	68.39 ±0.09	162±0.8
		Ours	Hard	X	69.52±0.34	172±25
		Ours	Hard	O	69.89±0.22	174±0.4
	ResNet20	tdBN	Soft	X	75.09±0.38	652±5.0
		tdBN	Hard	X	72.22±0.25	563±3.1
		Ours	Hard	X	74.22±0.11	584±6.5
		Ours	Hard	O	74.28±0.03	583±5.5
C10-DVS	VGG11	tdBN	Soft	X	77.57±0.27	153±4.0
		tdBN	Hard	X	74.80±0.26	248±0.7
		Ours	Hard	X	75.30±0.24	199±4.2
		Ours	Hard	O	75.50±0.77	201±3.2

V. Conclusion

This work proposed gradient clipping to improve the training performance of deep SNNs with a hard reset. Experiments demonstrated that the proposed approach prevents abrupt gradient changes and enhances optimization. Additionally, the noise of neuromorphic devices could be shown to positively impact the training. These results contribute to efficient and hardware-compatible SNN training, advancing neuromorphic computing for real-world applications.

Acknowledgment

This work was supported in part by the Korea Institute of Science and Technology (KIST) through 2E33561 and the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) [NRF-2021R1C1C2010454].

References

[1] C. D. Schuman *et al.*, "Opportunities for neuromorphic computing algorithms and applications," *Nat. Comput. Sci.*, vol. 2, no. 1, pp. 10–19, 2022.

[2] S. Park, D. Lee, and S. Yoon, "Noise-robust deep spiking neural networks with temporal information," in *DAC*, 2021.

[3] Y. Wu *et al.*, "Spatio-temporal backpropagation for training high-performance spiking neural networks," *Front. in Neurosci.*, vol. 12, p. 331, 2018.

[4] S. Kim, S. Park, B. Na, and S. Yoon, "Spiking-yolo: spiking neural network for energy-efficient object detection," in *AAAI*, 2020.

[5] H. Zheng *et al.*, "Going deeper with directly-trained larger spiking neural networks," in *AAAI*, 2021.