Optimization Method for Surrogate Function in Spiking Neural Networks Based on Membrane Potential Distribution

Qi Sun, Zhen Cao*, Kaige Geng, Ziyi Zhang, Biao Hou*

The School of Artificial Intelligence, Xidian University, Xi'an, 710126

Abstract

Spiking Neural Networks (SNNs) offer promising energy efficiency and temporal sparsity for edge intelligence, but their training remains difficult due to gradient mismatch, membrane potential drift, and discretization errors. In this paper, we propose a membrane potential-guided surrogate optimization(MPO) framework that dynamically aligns the surrogate function with the membrane potential distribution to enhance the gradient propagation. Specifically, we introduce a KL-divergence-based regularization to stabilize membrane potential dynamics, and an adaptive width constraint to synchronize the surrogate gradient range with neural activity statistics. Additionally, we design a spike discretization error metric and a correction strategy to mitigate temporal discretization effects. Experiments on CIFAR-10, CIFAR-100, and ImageNet show our method achieves 94.76%, 74.20%, and 65.70% top-1 accuracy respectively, while improving gradient stability and energy efficiency. This work provides a principled optimization scheme for robust and scalable SNN training in practical neuromorphic systems.

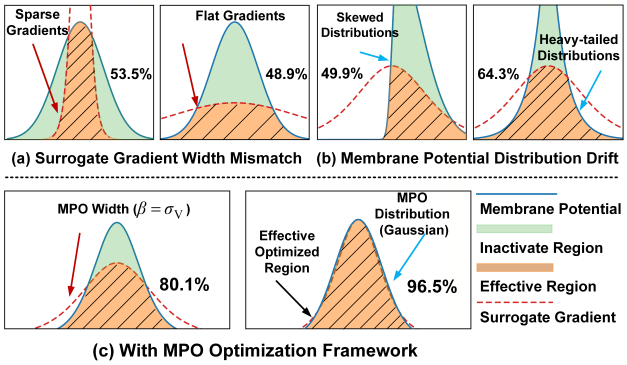

Figure 1: Illustration of how surrogate gradient width and distribution mismatch limits surrogate optimization. (a): Improper surrogate width (too flat or sparse) restricts effective optimized regions. (b): Distributional drift also reduces regions. (c): MPO adaptively aligns width and membrane potential distribution, expanding coverage to 96.5%.

Introduction

Inspired by the nature of biological neurons, Spiking Neural Networks (SNNs) leverage sparse, event-driven spike communication to achieve low-latency, energy-efficient processing—properties that make them attractive for edge and neuromorphic applications (Tavanaei et al. 2019a; Zhang, Jia, and Wang 2022). Unlike conventional artificial neural networks (ANNs), SNNs encode information in binary spike events, which reduces redundant computation and enables real-time operation on power-constrained hardware.

However, effective training of SNNs remains a core challenge due to the non-differentiable nature of spike generation and the complex temporal dynamics involved. Two mainstream strategies have emerged to address this: ANN-to-SNN conversion (Deng and Gu 2021; Bu et al. 2023), which maps pre-trained ANN weights onto spiking architectures, and surrogate gradient (SG) learning (Ponghiran and Roy 2022; Lian et al. 2023), which introduces smooth, differentiable approximations to enable backpropagation through time (BPTT). While conversion-based methods struggle with temporal resolution and latency, SG-based training provides direct optimization pathways and greater flexibility for network design.

Despite its growing popularity, SG learning is often hindered by unstable gradient propagation, especially in deep or large-scale SNNs (Lian et al. 2023; Zhou et al. 2024). A key bottleneck arises from the mismatch between the surrogate gradient's effective range and the membrane potential distribution. When the surrogate width is too narrow, neurons become insensitive to membrane potential fluctuations, impairing adaptability; when too wide, gradient magnitude becomes unbounded, introducing training instability. Fig. 1 shows these failure modes, where surrogate misalignment leads to either vanishing gradients or uncontrolled membrane potential excursions.

Several recent studies attempt to alleviate these issues by introducing auxiliary constraints or adaptive mechanisms. For example, membrane normalization (Ding et al. 2025), gradient clipping (Liang et al. 2025; Xue et al. 2025), and adaptive surrogate functions (Jiang and Zhang 2024) have been proposed to enhance gradient flow. However, these approaches often impose global or static rules that fail to capture the dynamic, layer-specific characteristics of the membrane potential distributions. Moreover, most methods over-

*Zhen Cao and Biao Hou are the co-corresponding authors

look the impact of temporal discretization errors—arising from time-step sampling—which further degrades gradient fidelity in long-sequence learning.

In light of these challenges, we propose a membrane potential-guided surrogate optimization (MPO) framework that dynamically couples membrane potential distributions with surrogate function design. Our method regularizes membrane potential statistics towards a stable Gaussian distribution through a KL divergence loss, adapts the surrogate gradient width based on real-time membrane potential variations, and introduces a discretization-aware gradient degradation metric to quantify temporal signal loss. Building on this metric, we develop an adaptive correction strategy to enhance temporal gradient propagation and improve training stability. The main contributions of this work are summarized as follows:

- We theoretically analyze how surrogate width mismatch and membrane potential shift impair the effective optimization region, and formally link distributional properties to gradient variance and training stability.
- We propose the membrane potential-guided surrogate optimization (MPO) framework that dynamically aligns surrogate width with membrane statistics to mitigate gradient mismatch.
- We achieve SOTA results on CIFAR-10, CIFAR-100, and ImageNet, improving accuracy, energy efficiency, and convergence speed over prior comparison methods.

Related Work

Surrogate Gradient Training Methods

Surrogate gradient (SG) techniques have become the dominant paradigm for end-to-end SNN training (Lian et al. 2023). By replacing the non-differentiable spike activation with smooth surrogate functions, SG enables backpropagation through time. Tavanaei et al. (Tavanaei et al. 2019b) systematically study the impact of surrogate shapes on gradient flow, revealing that Sigmoid-like surrogates tend to cause gradient vanishing in deep networks. Liang et al. (Liang et al. 2025) propose an adaptive gradient modulation method that dynamically adjusts the slope of the surrogate function according to layer-wise feature distributions, improving gradient stability by 37% in temporal tasks. Jiang et al. (Jiang and Zhang 2024) introduce KLIF neurons with learnable surrogate shapes and thresholds, significantly outperforming standard LIF models. However, most approaches apply static, globally uniform surrogate parameters, overlooking the layer-dependent dynamics of membrane potentials—potentially leading to gradient misalignment and convergence inefficiency.

Membrane Potential Distribution and Dynamic Regulation

The distribution of membrane potentials directly influences neuronal firing behavior and gradient dynamics. Shen et al. (Shen et al. 2024) demonstrate that learnable initial membrane potentials enrich firing patterns and improve static task performance through an IMP adjustment strategy. Duan et al. (Duan et al. 2022) introduce temporal effective batch normalization, applying weighted scaling to pre-activation inputs at each timestep to smooth temporal membrane potential distributions, stabilizing gradients and achieving state-of-the-art performance on multiple datasets. Xue et al. (Xue et al. 2025) designe a channel-parallel spiking neuron model that leverages temporal correlations to enlarge receptive fields and reduces computation with multiplication-free shift operations, boosting temporal information utilization. While these methods enhance static regularization and architecture, they lack online adaptive strategies to co-optimize surrogate gradients and membrane dynamics during training. This limits their ability to resolve gradient-range mismatches and discretization errors, impeding scalability to complex tasks.

Preliminary

Leaky Integrate-and-Fire Model

The Leaky Integrate-and-Fire (LIF) neuron is a widely used model that captures essential spiking dynamics. Its membrane potential $V_l(t)$ evolves according to the continuous dynamics:

$$\tau \frac{dV_l(t)}{dt} = -V_l(t) + I_l(t) \tag{1}$$

where τ is the membrane time constant. In discrete-time simulations with timestep Δt, Eq. (1) is approximated as:

$$V_l(t+1) = \beta V_l(t) + (1-\beta) I_l(t) - S_l(t) V_{th} \tag{2}$$

where $\beta = \exp(-\Delta t/\tau)$ and the spike output is:

$$S_l(t) = \Theta(V_l(t) - V_{th}) \tag{3}$$

$\Theta(\cdot)$ is the Heaviside step function. Upon spiking, the membrane potential resets via subtracting the threshold V_{th}.

Surrogate Gradient Method

The Heaviside spike function is non-differentiable:

$$\frac{\partial S(t)}{\partial V(t)} = \delta(V(t) - V_{th}) \tag{4}$$

which prohibits direct gradient computation. Surrogate gradient methods address this by replacing (4) with smooth approximations. For example:

$$\frac{\partial \tilde{S}(t)}{\partial V(t)} = \frac{\alpha}{1 + \alpha^2 (V(t) - V_{th})^2} \tag{5}$$

where α controls gradient sharpness. Alternatively, the piecewise linear surrogate gradient (PLGrad) is:

$$\frac{\partial \tilde{S}(t)}{\partial V(t)} = \begin{cases} \lambda, & |V(t) - V_{th}| \le \gamma \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

with λ the gradient magnitude and γ the width of the linear region. Using surrogate gradients, the weight gradients are computed via chain rule:

$$\frac{\partial L}{\partial W_l} = \sum_t \frac{\partial L}{\partial S_l(t)} \cdot \frac{\partial \tilde{S}_l(t)}{\partial V_l(t)} \cdot \frac{\partial V_l(t)}{\partial W_l} \tag{7}$$

This surrogate-based approach mitigates gradient sparsity and improves training stability.

Method

This paper proposes a membrane potential-guided surrogate optimization method. The overall framework is illustrated in Figure 2, and its core components are detailed in the following subsections.

Problem Formulation and Theoretical Motivation

SNNs commonly employ surrogate gradients to approximate the non-differentiable spiking activation function, facilitating gradient-based training. A widely used surrogate adopts a sigmoid form with a tunable width parameter β:

$$o(u) = \frac{1}{1 + \exp(-\beta(u - V_{\text{thresh}}))} \tag{8}$$

where u is the membrane potential and V_{thresh} is the spiking threshold. The gradient of this surrogate function:

$$o'(u) = \beta \cdot \frac{e^{-\beta(u - V_{\text{thresh}})}}{(1 + e^{-\beta(u - V_{\text{thresh}})})^2} \tag{9}$$

The gradient $o'(u)$ is unimodal and peaks at $u = V_{\text{thresh}}$, with a maximal slope of $\beta/4$. The surrogate width β crucially modulates gradient flow: smaller β yields oversmoothed gradients, while larger β sharply narrows the active region, risking vanishing updates. To formalize this relationship, we propose Proposition 1 to characterize the optimal β range for stable gradient localization.

Proposition 1. *The surrogate gradient aligns most effectively with membrane dynamics when the width parameter β matches the standard deviation σ_V of the membrane potential distribution, i.e., $\beta \approx \sigma_V$.*

Proof 1. We define the *effective support region* $\mathcal{S}_\epsilon$ of the surrogate gradient as the domain where $|o'(u)| > \epsilon$ for a small threshold ϵ. Under the assumption that membrane potentials follow a Gaussian distribution centered at threshold V_{thresh} with standard deviation σ_V (as supported by the Central Limit Theorem, CLT), this region is:

$$\mathcal{S}_\epsilon = \{u \mid |o'(u)| > \epsilon\}, \quad |\mathcal{S}_\epsilon| \approx \frac{2}{\beta} \ln\left(\frac{\beta}{4\epsilon}\right) \tag{10}$$

To ensure the surrogate gradient aligns with the typical range of u, we require:

- **Coverage**: $\mathcal{S}_\epsilon$ spans the bulk of the membrane potential distribution, ideally $[V_{\text{thresh}} - \sigma_V, V_{\text{thresh}} + \sigma_V]$;
- **Locality**: $\mathcal{S}_\epsilon$ should not extend far beyond $\pm\sigma_V$ to avoid gradient noise in low-density regions.

These two objectives are simultaneously optimized when $\beta \approx \sigma_V$, balancing spread and sharpness. If $\beta \ll \sigma_V$, gradients become too flat to drive updates; if $\beta \gg \sigma_V$, gradients are sparse and unstable. The detailed derivation is given in Appendix A.

To promote alignment across layers and neurons, we encourage each learnable surrogate width $\beta_i^{(l)}$ to match the empirical standard deviation $\sigma_{V,i}^{(l)}$ of membrane potentials by Eq (11).This scale-level alignment provides a stable foundation for surrogate gradient flow.

$$\min_{\beta} \sum_l \sum_i \left|\beta_i^{(l)} - \sigma_{V,i}^{(l)}\right| \tag{11}$$

While Proposition 1 ensures alignment between the surrogate gradient's support and the variance of membrane potentials, variance matching alone is insufficient for training stability. Skewed, heavy-tailed, or high-kurtosis distributions can still cause gradient dispersion, despite matched scale.

Given that a neuron's membrane potential arises from aggregating many weakly correlated inputs, its distribution approximately follows a Gaussian under mild conditions, by the CLT (Dayan and Abbott 2005; Gerstner et al. 2014). We thus introduce Proposition 2, linking gradient stability to the distributional similarity between the normalized membrane potential and the standard Gaussian.

Proposition 2. *Gradient fluctuation magnitude increases monotonically with the distributional dissimilarity between the normalized membrane potential distribution P_Z and the standard Gaussian Φ, thereby impairing gradient stability.*

Proof 2. Let the normalized membrane potential be defined as

$$Z := \frac{u - V_{\text{thresh}}}{\sigma_V}, \quad \text{with distribution } P_Z, \tag{12}$$

and let $g(z) := o'(\sigma_V z + V_{\text{thresh}})$ denote the surrogate gradient under width matching ($\beta \approx \sigma_V$). Since g is bounded and Lipschitz continuous, its fluctuation can be quantified by the deviation probability

$$\mathbb{P}\left(|g(Z) - \mathbb{E}[g(Z)]| > \kappa\right), \tag{13}$$

For a fixed threshold $\kappa \in (0, g(0))$. When $P_Z = \Phi$, this quantity admits tight exponential bounds via classical concentration inequalities such as Poincaré or log-Sobolev inequalities, leading to

$$\mathbb{P}\left(|g(Z) - \mathbb{E}[g(Z)]| > \kappa\right) \leq 2\exp\left(-\frac{c\kappa^2}{L^2}\right), \tag{14}$$

With $L = \sup|g'| \leq \sigma_V/4$ and $c > 0$ depending on Φ. In contrast, when P_Z deviates from Φ—e.g., by developing heavier tails or asymmetric profiles—such concentration bounds deteriorate, and the fluctuation probability admits only loose lower bounds.

Let $\mathfrak{D}(P_Z, \Phi)$ be any distributional dissimilarity measure satisfying: (i) $\mathfrak{D}(P_Z, \Phi) = 0$ iff $P_Z = \Phi$; (ii) $\mathfrak{D}$ increases under tail expansion, skewness, or excess kurtosis; (iii) $\mathfrak{D}$ is weakly lower semicontinuous. Then there exist non-decreasing functions $h, \psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(0) = \psi(0) = 0$, such that:

$$\mathbb{P}\left(|g(Z) - \mathbb{E}[g(Z)]| > \kappa\right) \geq h\big(\mathfrak{D}(P_Z, \Phi)\big) \tag{15}$$

$$\mathrm{Var}[g(Z)] \geq \psi\big(\mathfrak{D}(P_Z, \Phi)\big) \tag{16}$$

That is, deviations of P_Z from Gaussianity incur non-negligible gradient variance and fluctuation probability. Hence, aligning the overall shape of membrane potential distributions with the Gaussian not only supports the CLT-based neuronal modeling assumption but also ensures a more stable and concentrated gradient landscape.

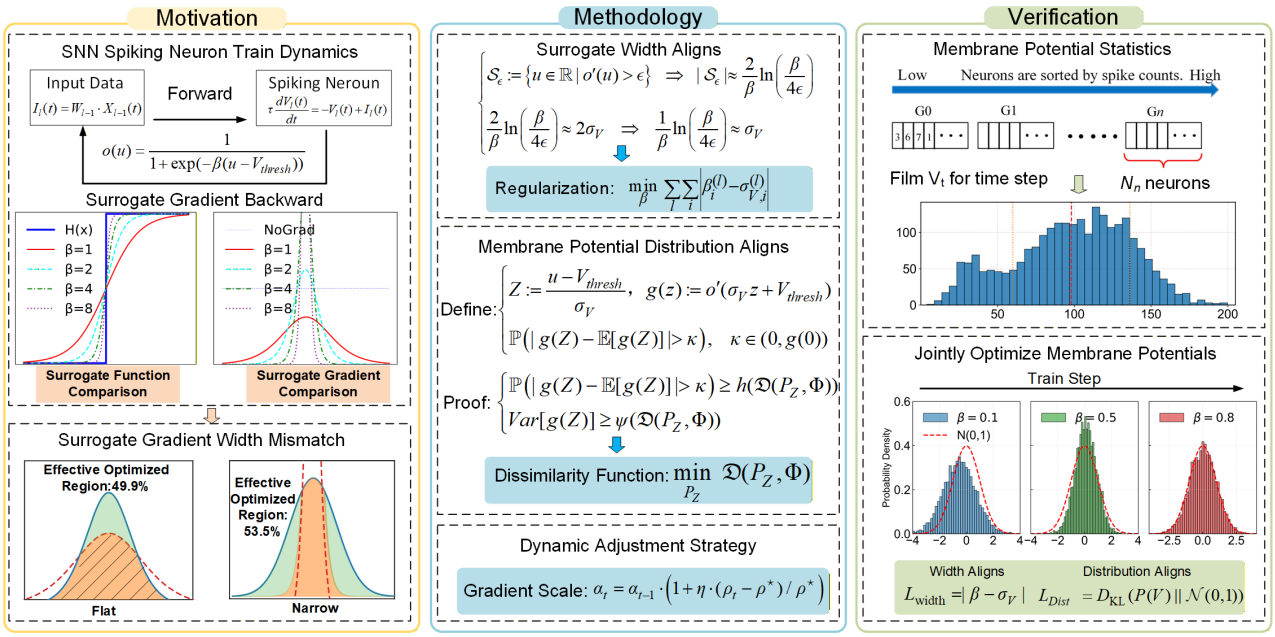

Figure 2: Overview of the Membrane Potential-Guided Optimization Framework. Left: Mismatched surrogate widths reduce training efficiency. Center: Our method aligns gradient width with membrane variance, minimizes distributional divergence, and adaptively scales gradients. Right: These strategies improve membrane regularity and gradient stability.

Surrogate Width Adaptation and Membrane Potential Distribution Regularization

Building on the theoretical foundations in previous section, we derive practical constraints from the statistics of membrane potentials to guide surrogate gradient shaping and stabilize neuronal dynamics. Specifically, we translate the theoretical insights from Proposition 1 and Proposition 2 into two loss components: one for aligning the surrogate width with membrane potential variability, and another for regularizing the distributional shape toward Gaussianity.

We first define the time-averaged membrane potential for each neuron n as:

$$U(n) = \frac{1}{T}\sum_{t=1}^{T} u_t^{(n)} \tag{17}$$

where T is the total simulation length. The global mean μ_V and standard deviation σ_V of membrane potentials across all N neurons are computed as:

$$\mu_V = \frac{1}{N}\sum_{n=1}^{N} U(n) \tag{18}$$

$$\sigma_V = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(U(n)-\mu_V)^2} \tag{19}$$

Surrogate Width Alignment Loss. Proposition 1 demonstrates that stable gradient propagation is achieved when the surrogate width β matches the standard deviation of the membrane potential. To enforce this alignment during training, we introduce a width regularization loss:

$$L_{\text{width}} = |\beta - \sigma_V| \tag{20}$$

This adaptive constraint dynamically adjusts the gradient's support region to match the membrane activity scale, avoiding both over-smoothing and over-concentration of surrogate gradients.

Membrane Potential Distribution Loss. Proposition 2 establishes that gradient stability is also influenced by higher-order discrepancies between the normalized membrane potential distribution and a standard Gaussian. Since individual membrane potentials are approximately Gaussian under the CLT, we explicitly encourage this alignment by minimizing the Kullback–Leibler divergence:

$$\begin{aligned} L_{\text{Dist}} &= D_{\text{KL}}(P(V) \parallel \mathcal{N}(0,1)) \\ &= \frac{1}{2}\left(\mu_V^2 + \sigma_V^2 - \log\sigma_V^2 - 1\right) \end{aligned} \tag{21}$$

This closed-form expression penalizes deviations in both the mean and variance, promoting symmetric and stable membrane potential distributions with low tail risk. For detailed derivation, refer to Appendix C.

Joint Objective. By jointly optimizing L_{Dist} and L_{width}, the model adapts both the surrogate gradient scale and the statistical shape of membrane potentials. This dual regularization improves training robustness by reducing gradient variance and preventing pathological membrane dynamics, aligning closely with theoretical conditions for stability and biological plausibility.

Dynamic Adjustment Strategy for Surrogate Function

Biological neurons maintain membrane activity within homeostatic bounds through homeostatic plasticity to avoid

saturation or inactivity. Building on this mechanism, we introduce a dynamic adjustment strategy that adaptively rescales the surrogate gradient by monitoring the membrane potential distribution during training, thereby enhancing training stability and gradient flow.

Specifically, we track the proportion ρ_t of membrane potentials falling within an effective dynamic range $\mathcal{I} = [\mu_V - \gamma\sigma_V, \mu_V + \gamma\sigma_V]$, where μ_V and σ_V denote the mean and standard deviation of V over a mini-batch. If ρ_t significantly deviates from a target ratio $\rho^\star$, we adjust the gradient scale α_t of the surrogate function accordingly to steer the membrane potential back into a well-behaved regime.

$$\alpha_t = \alpha_{t-1} \cdot \left(1 + \eta \cdot \frac{\rho_t - \rho^\star}{\rho^\star}\right) \tag{22}$$

where η is a sensitivity hyperparameter. This dynamic mechanism improves training robustness by preventing gradient vanishing or explosion caused by overly concentrated or dispersed potentials.

Algorithmic implementation is outlined in Appendix D.2, and statistical details of interval partitioning and the weight function design are discussed in Appendix D.1.

Training Framework and Optimization

The unified training objective combines task supervision with biologically inspired regularizers:

$$L = L_{\text{task}} + \lambda_2 L_{\text{width}} + \lambda_1 L_{\text{Dist}} \tag{23}$$

where L_{task} drives task-specific accuracy, L_{width} enforces alignment between surrogate gradient width β and membrane potential standard deviation σ_V, and L_{Dist} regularizes membrane potentials towards a stable Gaussian distribution. Coefficients λ_1, λ_2 balance regularizations.

To mitigate discretization errors in surrogate gradient training, the method incorporates two key mechanisms:

- **Surrogate Width Adaptation:** The surrogate width parameter β dynamically tracks the evolving membrane potential distribution, ensuring surrogate gradients remain informative and aligned with neural activation.
- **Neuron-wise Gradient Scaling:** Each neuron's weight updates are modulated by a scaling factor α_i, computed from the neuron's membrane potential distribution to reduce bias and variance in gradient estimates caused by temporal discretization.

These mechanisms adaptively tune both gradient shape and update scales, enhancing convergence stability and biological plausibility. Algorithm 1 outlines the training procedure, where potential statistics govern β and α_i updates per iteration.

Algorithm 1: Membrane Potential Guided Surrogate Optimization

Input: Training data $\mathcal{D}$, weights W, width β, learning rate η, batch count B, timesteps T, loss weights λ_1, λ_2
Output: Optimized weights W^*

Forward Pass:
for $\texttt{batch} = 1$ **to** B **do**
  Propagate V_t for $t = 1, \ldots, T$
  Record $V_{\text{final}} \leftarrow \{V_T^{(i)}\}_{i=1}^N$
  Compute $\mu_V, \sigma_V \leftarrow \text{stats}(V_{\text{final}})$
  Estimate $P(V) \leftarrow \text{histogram}(V_{\text{final}})$
  Loss Computation:
  $L_{\text{width}} \leftarrow |\beta - \sigma_V|$ *Eq.*(20)
  $L_{\text{Dist}} \leftarrow D_{\text{KL}}\big(P(V) \| \mathcal{N}(0,1)\big)$ *Eq.*(21)
  $L \leftarrow L_{\text{task}} + \lambda_1 L_{\text{dist}} + \lambda_2 L_{\text{width}}$ *Eq.*(23)
  Backward Pass:
  for each neuron i **do**
    $\alpha_i \leftarrow \phi(\sigma_V^{(i)}), \quad \Delta W_i \leftarrow \eta \cdot \alpha_i \cdot \frac{\partial L}{\partial W_i}$ *Eq.*(22)
  end for
  $W \leftarrow W - \Delta W, \quad \beta \leftarrow \beta - \eta \nabla_\beta L_{\text{width}}$
end for
return $W^* \leftarrow W$

Experimental

We comprehensively evaluate our method in terms of classification accuracy, robustness under corruption, and energy efficiency across multiple datasets and metrics.

Experiment Setup

Datasets We evaluate our approach on three widely-used benchmarks: CIFAR-10, CIFAR-100, and ImageNet. These datasets vary in scale and complexity, providing a comprehensive evaluation for accuracy and efficiency. Standard splits/preprocessing applied, with statistics in E.1.

Implementation Details All models are implemented in PyTorch with identical settings across variants. VGG-11 and ResNet-18 backbones integrated with our spiking modules. Training uses the Adam optimizer and cosine annealing schedule. Simulation timesteps are varied to assess temporal sensitivity. Full training configurations are listed in Appendix E.2.

Evaluation Metrics The Top-1 accuracy is used as the main performance metric. To account for robustness and efficiency, we additionally report Relative Robustness (RR) and the Simulated Energy Efficiency Index (SEEI), capturing model behavior under input corruption and energy constraints. Formal definitions are provided in Appendix E.3.

Experimental Results

To evaluate the effectiveness of the proposed MPO strategy, we conduct comprehensive experiments on CIFAR-10, CIFAR-100, and ImageNet. As shown in Table 1, MPO achieves the best or comparable performance across all datasets. On CIFAR-10 and CIFAR-100, MPO achieves a new state-of-the-art accuracy of 94.76% and 74.20%, surpassing RecDis-SNN (94.55%, 72.10%) and significantly outperforming other methods such as Dspike (93.13%,71.68%) and Hybrid Training (92.22%). On ImageNet, MPO attains 65.70%, closely approaching the best-performing RecDis-SNN (67.33%) while outperforming

Model	Method	Backbone Network	CIFAR-10	CIFAR-100	ImageNet
SpikeNorm (Sengupta et al. 2019)	ANN-to-SNN	VGG-16	91.55%	-	-
RMP-SNN (Han, Srinivasan, and Roy 2020)	ANN-to-SNN	VGG-16	93.63%	70.93%	-
SpikeCNN (Esser et al. 2016)	ANN-to-SNN	ConvNet	89.32%	65.48%	-
SpikeCompress (Rueckauer et al. 2017)	ANN-to-SNN	ConvNet	88.82%	-	-
BinarySNN (Wang et al. 2020)	ANN-to-SNN	ConvNet	90.19%	62.02%	-
CBP-QSNN (Yoo and Jeong 2023)	ANN-to-SNN	VGG-16&ResNet-34	91.51%	66.53%	60.10%
ALBSNN (Pei et al. 2023)	Direct training	ConvNet	92.12%	69.55%	-
BitSNNs (Hu, Zheng, and Pan 2024)	Direct training	ResNet-18	-	-	62.06%
ADMM-SNN (Deng et al. 2021)	Direct training	ConvNet	89.01%	55.95%	-
STBP (Wu et al. 2019)	Direct training	CIFARNet	90.53%(12)	-	-
TSSL-BP (Zhang and Li 2020)	Direct training	CIFARNet	91.41%(5)	-	-
RecDis-SNN (Guo et al. 2022)	Direct training	ResNet-19&ResNet-34	94.55%(6)	72.10%(6)	**67.33%**(6)
Dspike (Li et al. 2021)	Direct training	ResNet-18&VGG-19	93.13%(6)	71.68%(6)	-
Gradmax (Evci et al. 2022)	Direct training	WRN-28-1	91.10%(12)	66.80%(18)	-
Diet-SNN (Rathi and Roy 2020)	Direct training	CIFARNet&Mobilenet-V1	91.59%(5)	69.67%(5)	-
SpikeProp (Lee et al. 2020)	Direct training	ResNet-11&VGG-9	90.95%(100)	-	-
Hybrid Training (Rathi et al. 2020a)	Hybrid training	ResNet-20, ResNet-34	92.22%(250)	67.78%(125)	61.48%
STBP-tdBP (Rathi et al. 2020b)	Direct training	ResNet-19&ResNet-34	93.16%(5)	-	63.72%(6)
This work (MPO/CNN)	Direct training	ResNet-19&ResNet-34	**94.76%**(10)	**74.20%**(16)	**65.70%**(18)

Table 1: Accuracy comparison of proxy gradient-related methods

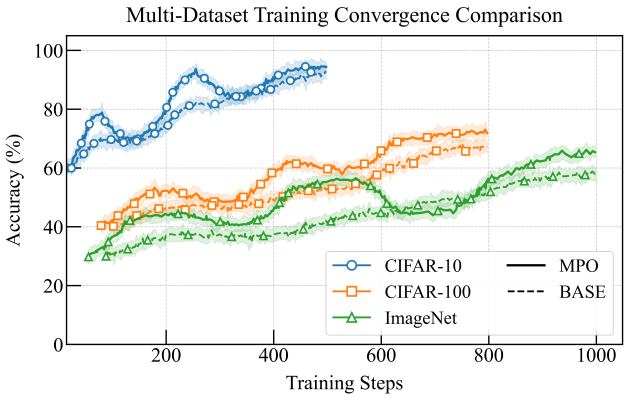

Figure 3: Training accuracy progressions.

other competitive methods such as STBP-tdBP (63.72%) and CBP-QSNN (60.10%).

Beyond accuracy, MPO achieves these results with relatively low simulation steps(10/16/18) on CIFAR-10, CIFAR-100, and ImageNet, respectively—demonstrating strong trade-offs between accuracy and temporal efficiency. In contrast, hybrid training strategies often require 100–250 steps to reach lower performance, and RecDis-SNN, while slightly higher on ImageNet, relies on delicate tuning and fewer steps, often at the cost of stability. Our method achieves stable convergence and avoids overfitting or saturation in deeper architectures such as ResNet-34, as further substantiated in training dynamics visualizations (Figure 3). These results collectively demonstrate that MPO provides a unified and scalable framework that generalizes well across datasets, offering both accuracy and efficiency advantages over existing SNN training paradigms.

Configuration	CIFAR-10	CIFAR-100	ImageNet	Avg. SEEI
Full MPO	**94.76%**	**74.20%**	**65.70%**	**8.7**
L_{Dist}	92.75% (↓1.4%)	70.5% (↓1.7%)	64.2% (↓1.5%)	7.9
L_{width}	92.35% (↓1.8%)	70.2% (↓2.0%)	64.0% (↓1.7%)	8.1
DAS	93.65% (↓0.5%)	71.8% (↓0.4%)	65.4% (↓0.3%)	7.3
SFH	93.35% (↓0.8%)	71.6% (↓0.6%)	65.1% (↓0.6%)	7.0
Baseline	91.55% (↓2.6%)	69.1% (↓3.1%)	63.3% (↓2.4%)	7.0

Table 2: Ablation results of the MPO method.

Ablation Study

Ablation of Key Modules To evaluate the necessity of each component in MPO, we conduct ablation experiments by individually removing the distribution loss, width regularization, surrogate function height (SFH), and dynamic adjustment strategy (DAS). As reported in Table 2, eliminating any of these modules consistently degrades classification accuracy and increases energy consumption (SEEI). Among them, removing the distribution loss leads to the largest average drop of 1.5% in accuracy, followed closely by width regularization. Although DAS and SFH result in smaller accuracy declines, they noticeably reduce SEEI, indicating their auxiliary role in efficient optimization. The full model achieves the highest accuracy and SEEI (8.7). These results confirm that each module—regardless of individual impact—contributes synergistically to MPO's performance-efficiency trade-off.

Membrane Dynamics and Training Stability To examine how distribution constraints enhance training stability, we analyze the co-evolution of membrane dynamics and surrogate adaptation. As training progresses, the membrane po-

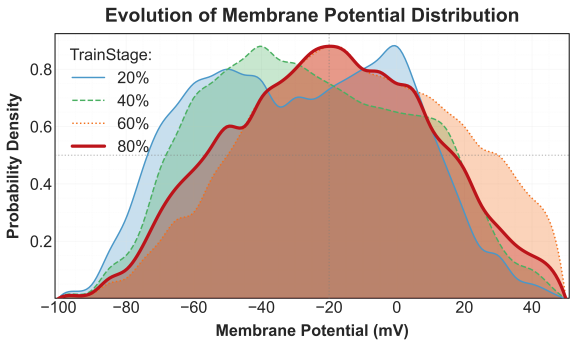

Figure 4: Schematic of Membrane Potential Distribution Evolution

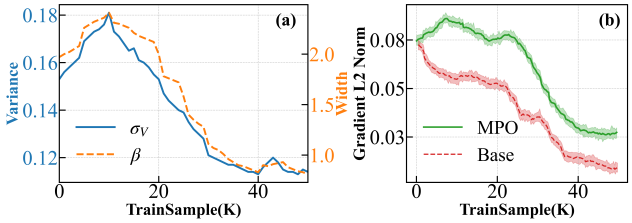

Figure 5: (a) Surrogate width β vs. membrane σ^2 (blue/orange); (b) Gradient L_2 norm comparison (MPO/Base)

tential distribution evolves from bimodal to near-Gaussian (Figure 4), suppressing outliers and enhancing state regularity. Concurrently, the surrogate width β adapts proportionally to membrane variance σ_V (Figure 5a), aligning gradient sensitivity with activation variability. This coordination leads to smoother, lower-magnitude gradients (Figure 5b), confirming the role of distribution-aware constraints in stabilizing convergence. Further analysis is provided in Appendix F.

Discrete Error Verification To assess the impact of discretization error, we compare performance with and without our adjustment strategy. As shown in Table 3, unadjusted networks exhibit higher membrane potentials near the threshold, leading to more near-fire neurons and a larger performance gap (11.3% between T=20 and T=8). After adjustment, the voltage distribution shifts away from the threshold, reducing near-fire activity and narrowing the gap to 6.9%. These results confirm the method's effectiveness and are consistent with the theoretical analysis in Appendix G, which explains how discretization alters membrane variance and degrades surrogate gradient coverage.

Robustness Experiment We evaluate MPO's robustness under Gaussian and salt-and-pepper noise on CIFAR-10. As shown in Figure 6, MPO consistently surpasses the baseline across noise levels, retaining 91.97% accuracy at σ=0.2 and over 91% at σ=0.3, where the baseline drops below 90.03%. This robustness arises from membrane potential optimization, which emphasizes stable features and mitigates noise-induced drift. Extended results are in Appendix H.

Condition	T	Avg. V_{end}	Near-fire	Accuracy
Without adjustment	8	-58mV	37%	80.9%
	12	-60mV	34%	86.3%
	16	-64mV	30%	90.1%
	20	-67mV	26%	92.2%(↑11.3%)
With adjustment	8	-62mV	25%	85.8%
	12	-64mV	23%	87.2%
	16	-66mV	20%	91.0%
	20	-68mV	19%	92.7%(↑6.9%)

Table 3: Discretization Error Optimization

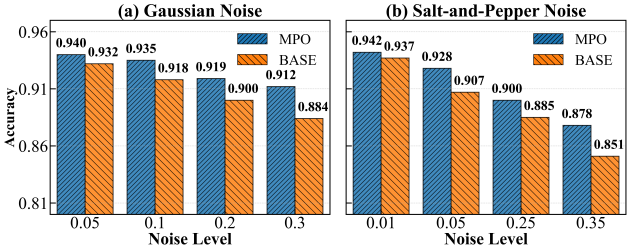

Figure 6: Robustness under different noise conditions

Power Consumption

We evaluate MPO's energy efficiency using Spike Density, SynOps, and SEEI. As shown in Table 4, MPO consistently improves accuracy while reducing spike activity and synaptic operations on both VGG-16 and ResNet-18. For example, Spike Density drops from 0.53 to 0.41 on VGG-16 and from 0.65 to 0.47 on ResNet-18, yielding higher SEEI. These improvements stem from MPO's suppression of redundant spikes and optimized membrane potential dynamics, enabling superior accuracy-efficiency trade-offs.

Method	SNN Architecture	Accuracy	Spike Density	SynOps	SEEI
BASE	VGG-16	91.11%	0.53	23.0M	7.8
	ResNet-18	92.35%	0.65	26.7M	8.1
MPO	VGG-16	92.95%	0.41	19.5M	9.4
	ResNet-18	94.12%	0.47	20.1M	10.2

Table 4: Power Consumption Performance Comparison

Conclusions

This work proposes a membrane potential-guided surrogate optimization framework to tackle key challenges in SNN training, including gradient mismatch and spike discretization errors. By dynamically aligning the surrogate gradient width with membrane potential statistics, and introducing a Dynamic Adjustment Strategy, the method significantly improves training efficiency and model performance. Experiments show superior results on CIFAR-10, CIFAR-100, and ImageNet, along with better gradient stability and energy efficiency. The proposed MPO framework demonstrates strong robustness through comprehensive ablation studies. Overall, this study offers a unified optimization strategy that harmonizes biological plausibility and engineering requirements, laying a solid foundation for advancing efficient, low-power SNN-based intelligence.

Acknowledgments

This work was supported by the Siyuan Fund Project, China under Grant No. SY240101203.

References

Bu, T.; Fang, W.; Ding, J.; Dai, P.; Yu, Z.; and Huang, T. 2023. Optimal ANN-SNN conversion for high-accuracy and ultra-low-latency spiking neural networks. *arXiv preprint arXiv:2303.04347*.

Dayan, P.; and Abbott, L. F. 2005. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press.

Deng, L.; Wu, Y.; Hu, Y.; Liang, L.; Li, G.; Hu, X.; Ding, Y.; Li, P.; and Xie, Y. 2021. Comprehensive snn compression using admm optimization and activity regularization. *IEEE transactions on neural networks and learning systems*, 34(6): 2791–2805.

Deng, S.; and Gu, S. 2021. Optimal conversion of conventional artificial neural networks to spiking neural networks. *arXiv preprint arXiv:2103.00476*.

Ding, Y.; Zuo, L.; Jing, M.; He, P.; and Deng, H. 2025. Rethinking spiking neural networks from an ensemble learning perspective. *arXiv preprint arXiv:2502.14218*.

Duan, C.; Ding, J.; Chen, S.; Yu, Z.; and Huang, T. 2022. Temporal effective batch normalization in spiking neural networks. *Advances in Neural Information Processing Systems*, 35: 34377–34390.

Esser, S. K.; Merolla, P. A.; Arthur, J. V.; Cassidy, A. S.; Appuswamy, R.; Andreopoulos, A.; Berg, D. J.; Mckinstry, J. L.; Melano, T.; and Barch, D. R. 2016. Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing. *arXiv e-prints*.

Evci, U.; van Merrienboer, B.; Unterthiner, T.; Vladymyrov, M.; and Pedregosa, F. 2022. Gradmax: Growing neural networks using gradient information. *arXiv preprint arXiv:2201.05125*.

Gerstner, W.; Kistler, W. M.; Naud, R.; and Paninski, L. 2014. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Guo, Y.; Tong, X.; Chen, Y.; Zhang, L.; Liu, X.; Ma, Z.; and Huang, X. 2022. Recdis-snn: Rectifying membrane potential distribution for directly training spiking neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 326–335.

Han, B.; Srinivasan, G.; and Roy, K. 2020. Rmp-snn: Residual membrane potential neuron for enabling deeper high-accuracy and low-latency spiking neural network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13558–13567.

Hu, Y.; Zheng, Q.; and Pan, G. 2024. BitSNNs: revisiting energy-efficient spiking neural networks. *IEEE Transactions on Cognitive and Developmental Systems*, 16(5): 1736–1747.

Jiang, C.; and Zhang, Y. 2024. Klif: An optimized spiking neuron unit for tuning surrogate gradient function. *Neural Computation*, 36(12): 2636–2650.

Lee, C.; Sarwar, S. S.; Panda, P.; Srinivasan, G.; and Roy, K. 2020. Enabling spike-based backpropagation for training deep neural network architectures. *Frontiers in neuroscience*, 14: 497482.

Li, Y.; Guo, Y.; Zhang, S.; Deng, S.; Hai, Y.; and Gu, S. 2021. Differentiable spike: Rethinking gradient-descent for training spiking neural networks. *Advances in neural information processing systems*, 34: 23426–23439.

Lian, S.; Shen, J.; Liu, Q.; Wang, Z.; Yan, R.; and Tang, H. 2023. Learnable Surrogate Gradient for Direct Training Spiking Neural Networks. In *IJCAI*, 3002–3010.

Liang, Y.; Wei, W.; Belatreche, A.; Cao, H.; Zhou, Z.; Wang, S.; Zhang, M.; and Yang, Y. 2025. Towards Accurate Binary Spiking Neural Networks: Learning with Adaptive Gradient Modulation Mechanism. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 1402–1410.

Pei, Y.; Xu, C.; Wu, Z.; Liu, Y.; and Yang, Y. 2023. ALBSNN: ultra-low latency adaptive local binary spiking neural network with accuracy loss estimator. *Frontiers in Neuroscience*, 17: 1225871.

Ponghiran, W.; and Roy, K. 2022. Spiking neural networks with improved inherent recurrence dynamics for sequential learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 8001–8008.

Rathi, N.; and Roy, K. 2020. Diet-snn: Direct input encoding with leakage and threshold optimization in deep spiking neural networks. *arXiv preprint arXiv:2008.03658*.

Rathi, N.; Srinivasan, G.; Panda, P.; and Roy, K. 2020a. Enabling deep spiking neural networks with hybrid conversion and spike timing dependent backpropagation. *arXiv preprint arXiv:2005.01807*.

Rathi, N.; Srinivasan, G.; Panda, P.; and Roy, K. 2020b. Enabling deep spiking neural networks with hybrid conversion and spike timing dependent backpropagation. *arXiv preprint arXiv:2005.01807*.

Rueckauer, B.; Lungu, I.-A.; Hu, Y.; Pfeiffer, M.; and Liu, S.-C. 2017. Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Frontiers in neuroscience*, 11: 682.

Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; and Roy, K. 2019. Going deeper in spiking neural networks: VGG and residual architectures. *Frontiers in neuroscience*, 13: 95.

Shen, H.; Zheng, Q.; Wang, H.; and Pan, G. 2024. Rethinking the membrane dynamics and optimization objectives of spiking neural networks. *Advances in Neural Information Processing Systems*, 37: 92697–92720.

Tavanaei, A.; Ghodrati, M.; Kheradpisheh, S. R.; Masquelier, T.; and Maida, A. 2019a. Deep learning in spiking neural networks. *Neural networks*, 111: 47–63.

Tavanaei, A.; Ghodrati, M.; Kheradpisheh, S. R.; Masquelier, T.; and Maida, A. 2019b. Deep learning in spiking neural networks. *Neural networks*, 111: 47–63.

Wang, Y.; Xu, Y.; Yan, R.; and Tang, H. 2020. Deep spiking neural networks with binary weights for object recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3): 514–523.

Wu, Y.; Deng, L.; Li, G.; Zhu, J.; Xie, Y.; and Shi, L. 2019. Direct training for spiking neural networks: Faster, larger, better. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 1311–1318.

Xue, P.; Fang, W.; Ma, Z.; Huang, Z.; Zhou, Z.; Tian, Y.; Masquelier, T.; and Zhou, H. 2025. Channel-wise Parallelizable Spiking Neuron with Multiplication-free Dynamics and Large Temporal Receptive Fields. *arXiv preprint arXiv:2501.14490*.

Yoo, D.; and Jeong, D. S. 2023. Cbp-qsnn: Spiking neural networks quantized using constrained backpropagation. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 13(4): 1137–1146.

Zhang, D.; Jia, S.; and Wang, Q. 2022. Recent advances and new frontiers in spiking neural networks. *arXiv preprint arXiv:2204.07050*.

Zhang, W.; and Li, P. 2020. Temporal spike sequence learning via backpropagation for deep spiking neural networks. *Advances in neural information processing systems*, 33: 12022–12033.

Zhou, C.; Zhang, H.; Yu, L.; Ye, Y.; Zhou, Z.; Huang, L.; Ma, Z.; Fan, X.; Zhou, H.; and Tian, Y. 2024. Direct training high-performance deep spiking neural networks: a review of theories and methods. *Frontiers in Neuroscience*, 18: 1383844.