HLML-SNN: Fast Continual Learning in Spiking Neural Networks Achieved via Hebbian Learning-Driven Meta-Learning

Jiangshuai Xu [1*], Peiyun Xue [1*], Jiacheng Song[1], Xuhui Huang[1†], Qingshan Hou[1†]

[1]College of Electronic Information Engineering, Taiyuan University of Technology, China
huangxuhui@tyut.edu.cn, houqingshancv@gmail.com

Abstract

Catastrophic forgetting remains a fundamental barrier to artificial continual learning (CL) - a capability innate to humans. Existing CL methods often incur prohibitive computational costs in resource-constrained scenarios. Spiking neural networks (SNNs), with their biological plausibility and energy efficiency, offer distinct advantages for CL. Inspired by cortico-hippocampal memory mechanisms, we propose a spiking neural network framework integrating Hebbian plasticity with meta-learning, named HLML-SNN. This architecture emulates a dual-phase task-incremental CL process: (1) In the short-term phase, sample-level Hebbian learning rapidly adapts to new inputs through local synaptic updates; (2) In the long-term phase, task-level meta-learning optimizes cross-task parameters using consolidated synaptic weights, mimicking cortical memory integration to refine shared representations and initialize subsequent Hebbian learning. HLML-SNN incrementally transforms short-term adaptations into stable long-term knowledge, where the synergy of rapid synaptic updates and meta-driven global optimization enables efficient continual learning while balancing stability and plasticity. Empirical results establish HLML-SNN's state-of-the-art performance across split-MNIST/CIFAR10/CIFAR100/TinyImageNet while markedly reducing training time compared to existing methods, demonstrating substantial practical potential for rapid deployment scenarios. *The code and appendix are available on https://github.com/JiangshuaiXu/HLML_SNN.*

Introduction

Humans possess an innate capacity for continual learning, enabling ongoing knowledge acquisition across the lifespan without catastrophic forgetting (Kudithipudi et al. 2022). This sustained retention of prior knowledge during new learning stems from a critical balance between stability and plasticity in the nervous system (Turrigiano and Nelson 2004). Such capability is indispensable for high-stakes applications like autonomous driving (Mei et al. 2024) and AI-assisted healthcare(Ye et al. 2024; Perkonigg et al. 2021), making its efficient deployment in resource-constrained settings essential.

Current intelligent systems remain highly susceptible to catastrophic forgetting—a phenomenon where critical parameters encoding prior knowledge are overwritten during new task learning, resulting in abrupt performance deterioration on previously mastered tasks. So far, there are three primary CL paradigms (Van de Ven, Tuytelaars, and Tolias 2022) for addressing this issue: (1) Weight constrained updating methods, which include regularization based methods (Zenke, Poole, and Ganguli 2017; Zhao et al. 2024; Kirkpatrick et al. 2017) with protecting important weights of prior tasks via penalty and orthogonalization based methods (Zeng et al. 2019; Farajtabar et al. 2020) by updating the new task parameters on the orthogonal direction of the prior task parameters; (2) Replay methods (Lopez-Paz and Ranzato 2017; Chaudhry et al. 2019), which store and rehearse prior task samples during new training; (3) Dynamic architecture methods (Masse, Grant, and Freedman 2018; Douillard et al. 2022), which assign dynamically dedicated neurons or special network structure to new tasks. These algorithms share a core principle: leveraging past knowledge to restrict model updates on new tasks, enabling simultaneous forward/backward knowledge transfer. Currently, meta-learning (Finn et al. 2019; Hospedales et al. 2021; Vettoruzzo et al. 2024a) has been extensively validated as a pivotal approach to achieving continual learning, by extracting cross-task generalizable strategies to enable scratch-free adaptation and prevent catastrophic forgetting. Multiple meta-learning variants have been developed for the above CL paradigms: Self-Attention Meta-Learner (SAM) (Sokar, Mocanu, and Pechenizkiy 2021) constructs generalizable priors, using self-attention to filter task-relevant representations and reduce cross-task interference; Consistent MetaReg (Tian et al. 2021) stabilizes few-shot adaptation against intra-task shifts via consistent meta-regularization; Meta Reusable Knowledge (Hurtado, Raymond, and Soto 2021) builds meta-knowledge repositories with task-specific masks for selective weight activation, promoting reuse over overwriting; Transformer meta-optimizers (Vettoruzzo et al. 2024b) leverage attention to generate task-adaptive update rules from parameter dependencies; Meta continual learning via variance reduction (VR-MCL) (Wu et al. 2024) employs meta-learned variance reduction to stabilize cross-task updates for robust knowledge integration. However, these methods still suffer from

[*]These authors contributed equally.
[†]Corresponding authors.

architectural complexity, weak generalization, and hyperparameter sensitive stability-plasticity trade-off, which make it challenging to achieve fast continual learning to support swift deployment in resource-constrained scenarios.

The brain enables learning and memory through complementary multi-scale mechanisms (Hassabis et al. 2017): At the systems level, the cortico-hippocampal circuit (Sun et al. 2023) implements dual-timescale processing—hippocampal rapid encoding for transient storage and neocortical gradual consolidation into stable representations—exhibiting computational isomorphism with meta-learning's dual-loop mechanisms. This neurobiological blueprint for learning-to-learn provides directly transferable principles for neuro-inspired designs. Concurrently at the synaptic level, Hebbian plasticity (Kempter, Gerstner, and Van Hemmen 1999; Gerstner and Kistler 2002) achieves millisecond-precise feature extraction of novel stimuli through spike-timing-dependent plasticity (STDP) (Caporale and Dan 2008; Song, Miller, and Abbott 2000), where dynamic strengthening of co-activated synapses facilitates instantaneous contextual adaptation via selective pathway potentiation. This integrated architecture, linking macroscopic memory systems with synaptic rapid-feature detection, inspires catastrophic forgetting-resilient CL agents.

Therefore, we propose a Hebbian learning-driven meta-learning framework for spiking neural networks (HLML-SNN). Our approach addresses the fundamental task-incremental CL dilemma between rapid task adaptation and stable knowledge retention through a biologically-inspired dual-phase learning mechanism. The framework operates through complementary learning phases: rapid synaptic adaptation via Hebbian learning for millisecond-scale new-task encoding, and incremental meta-learning for long-term memory consolidation. This design mimics the hippocampocortical memory consolidation process, where short-term memory of Hebbian synaptic changes guide progressive meta-parameter accumulation across tasks while providing optimal initialization for subsequent Hebbian adaptations. The main contributions of this paper include:

• We develop a novel SNN framework that emulates cortico-hippocampal memory dynamics, achieving rapid adaptation without catastrophic forgetting or requiring memory replay/network expansion, enabling resource-efficient deployment in constrained environments.

• We introduce a synergistic mechanism where Hebbian plasticity drives rapid task-specific adaptations, which then inform incremental meta-parameter updates. This bidirectional coupling accelerates training convergence while maintaining stable long-term memory consolidation.

• Our HLML-SNN achieves state-of-the-art results for the task-incremental CL on split-MNIST, split-CIFAR10, split-CIFAR100, and split-TinyImageNet benchmarks while significantly reducing training time, demonstrating strong potential for real-world applications requiring rapid model updates and deployment.

Related Works

Spiking neural networks (Zhou et al. 2024; Maass 1997; Taherkhani et al. 2020; Guo, Huang, and Ma 2023) can realistically mimic biological features of neurons and synapses (Izhikevich 2003), offering inherent sparse coding and low-power advantages (Roy, Jaiswal, and Panda 2019). Currently, SNNs have gradually been explored for serving as a potential solution for efficient CL, representative approaches include: Neuromodulation-assisted credit assignment (Zhang et al. 2023) mitigates forgetting via global neuromodulator-mediated synaptic plasticity; Hebbian learning based orthogonal (Xiao et al. 2024) projects new-task weights orthogonally to principal neural subspaces to reduce interference. The SNNs model with selective activation (Shen et al. 2024) employs trace-based k-WTA sparsity and adaptive thresholds for task-specific neuronal segregation. The ALADE-SNN framework (Ni et al. 2025) balances feature representations of old and new tasks via dynamically expandable network structure, adaptive logit alignment, and OtoN suppression. Active Dendrite SNN (Pes et al. 2024) implements dynamic gating through TTFS dendritic delay modulation; A cortico-hippocampal hybrid neural network (Shi et al. 2025) models integrates ANNs (cross-task regularities) with SNNs (event encoding) with memory replay for knowledge consolidation. Although these methods advance continual learning in SNNs, core challenges persist: Insufficient guidance and adaptive regulatory capabilities from bio-inspired mechanisms; Meta-learning approaches lack incremental accumulation properties, failing to support robust incremental knowledge consolidation during long-term CL; Lack of mechanisms enabling fast continual learning.

Comparing with the above works, we propose the HLML-SNN, the Hebbian learning-driven meta-learning in SNN framework that mimics cortico-hippocampal memory consolidation. By dynamically accumulating meta-parameters for simultaneous task-adaptive initialization and short-term to long-term knowledge transfer, our method preserves SNNs' efficiency while bridging Hebbian plasticity and meta-learning continuity gaps, enabling task-incremental CL in resource-constrained scenarios.

Methodology

The proposed HLML-SNN is a neural network framework designed for fast task-incremental continual learning. Drawing inspiration from the long-term and short-term memory mechanisms of the cortico-hippocampal system, this framework achieves dual-phase collaboration through local synaptic plasticity via Hebbian learning and cross-task global parameter optimization via meta-learning. This design ensures robust continual learning performance while significantly reducing training time. The overall architecture is illustrated in Figure 1, comprising a pretrained ResNet feature extractor with enhanced convolution kernels and an HLML-SNN module responsible for continual learning classification, among them, the SNN model uses LIF neurons (Hunsberger and Eliasmith 2015). The detailed mathematical formulations of the LIF neuron dynamics are provided in the appendix for reference.

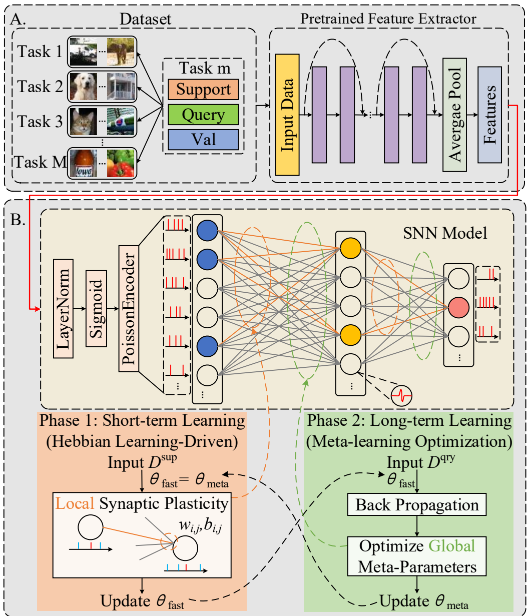

Figure 1: The task-incremental continual learning architecture of HLML-SNN.(A) Data task flow: The dataset is divided into sequential tasks, each containing a support set D_m^{sup} for Hebbian-based rapid synaptic adaptation, a query set D_m^{qry} for meta-learning and global parameter optimization, and a validation set D_m^{val} for performance evaluation. A pretrained modified ResNet serves as the feature extractor, and its outputs are fed into the SNN module for continual learning. (B) CL framework of the SNN: Extracted features are transformed into spike trains via Poisson encoding. During short-term learning (Hebbian phase), co-occurrence of presynaptic $\text{pre}(t)$ and postsynaptic $\text{post}(t)$ spikes strengthens synapses (orange links), producing fast-adapted parameters $\theta_{\text{fast}}^{(m)}$. During long-term learning (meta-learning phase), gradients (green arrows) from the query loss L_m update meta-parameters θ_{meta}, resetting the initial state for the next Hebbian update.

Framework of HLML-SNN Continual Learning

The lifelong learning capability of nervous systems stems from the complementary roles of long-term and short-term memory in the cortico-hippocampal circuit (Sun et al. 2023). Short-term memory enables rapid acquisition of new knowledge, while long-term memory is responsible for stable storage and integration of information across experiences. HLML-SNN leverages this biological principle to construct a dual-phase collaborative framework: in the short-term phase, local synaptic weights are adjusted by Hebbian learning, based on co-firing patterns of neural spikes, to achieve rapid adaptation to new tasks; in the long-term phase, global memory are regulated through meta-learning, optimizing cross-task general parameters to integrate short-term adaptations into stable long-term memory, to achieve collaborative optimization between new task acquisition and old knowledge retention ultimately.

Fast Adaptation via Hebbian Learning Hebbian learning dynamically adjusts local synaptic weights by capturing temporal correlations in synaptic activities. The core principle follows the well-established rule that synaptic connections are strengthened when presynaptic and postsynaptic neurons fire together (Caporale and Dan 2008; Song, Miller, and Abbott 2000), while prioritizing recent activities to minimize excessive modification of established connections. Specifically, we record the presynaptic spike sequence $\{\text{pre}(1), \text{pre}(2), \dots, \text{pre}(T)\}$ and postsynaptic spike sequence $\{\text{post}(1), \text{post}(2), \dots, \text{post}(T)\}$ of the SNN over T consecutive time steps, where T represents the temporal window of the SNN. The mean firing rate within this time window is computed as:

$$\overline{\text{pre}} = \frac{1}{T}\sum_{t=1}^{T}\text{pre}(t), \quad \overline{\text{post}} = \frac{1}{T}\sum_{t=1}^{T}\text{post}(t) \tag{1}$$

where $\overline{\text{pre}}$ and $\overline{\text{post}}$ represent the mean firing rates of presynaptic and postsynaptic neurons, respectively. This temporal mean firing rates provide a foundation for activity-dependent weight updates.

To balance rapid adaptation to new tasks with parameter stability, we employ a learning rate decay strategy during the short-term learning phase of each task. The Hebbian learning rate decays exponentially with training steps:

$$\eta_h(k) = \eta_{h0} \cdot \gamma^k, \quad k = 0, 1, \dots, K-1 \tag{2}$$

where η_{h0} is the initial Hebbian learning rate, γ is the decay factor, and k is the short-term learning step. This design enables rapid weight adjustment toward task-critical features in early phases, followed by fine-tuning and convergence in later phases, preventing excessive updates that could damage established connections. The weight update rule strictly adheres to the co-firing strengthening principle. The weight update at step k is given by:

$$\Delta w(k) = \eta_h(k) \cdot \overline{\text{post}}^{\top}\overline{\text{pre}} \cdot \frac{1}{B} \tag{3}$$

where $\overline{\text{post}}^{\top}\overline{\text{pre}}$ represents the correlation matrix between presynaptic and postsynaptic activities, and B is the batch size. The bias parameter update implements dynamic regulation of neuronal excitability based on average postsynaptic activity:

$$b(k+1) = b(k) + \eta_h(k) \cdot \overline{\text{post}} \cdot \frac{1}{B} \tag{4}$$

When postsynaptic neurons are frequently activated, the bias update amplitude increases to reduce activation thresholds and accelerate task adaptation. Conversely, when activation is sparse, the update amplitude decreases accordingly.

Cross-Task Parameter Optimization via Meta-Learning Meta-learning provides global guidance for Hebbian learning by optimizing meta-parameters, endowing the system with cross-task generalization capabilities. The fundamental role of meta-parameters is to learn the general structural patterns across all tasks rather than task-specific knowledge (Son, Lee, and Kim 2024; Chi et al. 2022; Javed and White 2019), which is crucial for achieving catastrophic forgetting mitigation.

The meta-parameters $\theta_{\text{meta}} = \theta_{\text{lin}}$ encompass the linear layer parameters $\theta_{\text{lin}} = \{w^{(l)}, b^{(l)}\}_{l=1}^{L}$, where $w^{(l)}$ and $b^{(l)}$ represent the weight and bias of the l-th layer, respectively. These meta-parameters provide high-quality initialization for Hebbian learning. For the m-th task $\mathcal{T}_m$, fast parameters $\theta_{\text{fast}}^{(m)}$ are obtained by cloning the current meta-parameters, which are then adaptively adjusted on the support set $\mathcal{D}_m^{\text{sup}}$ through Hebbian learning. The adapted parameters $\theta_{\text{fast}}^{(m)}$ are subsequently applied to the query set $\mathcal{D}_m^{\text{qry}}$ to compute the loss:

$$\mathcal{L}_m = \text{CrossEntropy}(\text{SNN}(\theta_{\text{fast}}^{(m)}, x_{\text{qry}}), y_{\text{qry}}) \tag{5}$$

where $x_{\text{qry}} \in \mathcal{D}_m^{\text{qry}}$ and y_{qry} is the corresponding ground-truth label. Meta-learning optimizes θ_{meta} through backpropagation based on this loss evaluation, with the update rule $\theta_{\text{meta}} \leftarrow \theta_{\text{meta}} - \eta_{\text{meta}} \cdot \nabla_\theta \mathcal{L}_m$, where η_{meta} is the meta-learning rate, updated using cosine annealing $\eta_{\text{meta}}(t) = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \cdot \left(1 + \cos\left(\pi \cdot \frac{t}{T_{\max}}\right)\right)$. The gradient $\nabla_\theta \mathcal{L}_m$ is computed using surrogate gradients for SNNs to address the non-differentiability of spike functions.

Synergistic Mechanism between Hebbian Learning and Meta-Learning

Hebbian learning and meta-learning achieve collaboration through a closed-loop process of "parameter initialization $\rightarrow$ rapid fine-tuning $\rightarrow$ global optimization." The core mechanism involves meta-parameters providing a reusable initialization foundation for Hebbian learning, while Hebbian learning provides task-specific feedback for meta-parameter optimization. This bidirectional interaction achieves dynamic balance between old and new knowledge through explicit parameter transmission and gradient correlation.

Meta-parameters initialize the starting point for Hebbian learning. The initial parameters for Hebbian learning on the m-th task are set as $\theta_{\text{fast}}^{(m,0)} = \theta_{\text{lin}}$, including initial weights $w_{\text{fast}}^{(m,0)} = w^{(l)}$ and biases $b_{\text{fast}}^{(m,0)} = b^{(l)}$. During the inner short-term learning phase, Hebbian learning performs rapid fine-tuning within the local parameter space. Starting from the initial parameters $\theta_{\text{fast}}^{(m,0)}$, Hebbian learning executes K steps of fine-tuning on the support set $\mathcal{D}_m^{\text{sup}}$, generating task-adapted parameters through the decay-based update rule:

$$\theta_{\text{fast}}^{(m,k)} = \theta_{\text{fast}}^{(m,k-1)} + \Delta\theta_{\text{heb}}(k) \tag{6}$$

where $\Delta\theta_{\text{heb}}(k) = \{\Delta w(k), \Delta b(k)\}$ represents the Hebbian update at step k, determined by the mean presynaptic and postsynaptic spike activities $\overline{\text{pre}}$, $\overline{\text{post}}$ and the decaying learning rate $\eta_h(k)$. Since $\eta_h(k)$ decays exponentially with steps, the fine-tuning amplitude gradually diminishes, ensuring that only synaptic connections strongly correlated with the current task are reinforced, thereby reducing interference with parameters from previous tasks. In the outer long-term learning phase, meta-learning optimizes global parameters based on Hebbian learning outcomes. The loss $\mathcal{L}_m$ of the fine-tuned parameters $\theta_{\text{fast}}^{(m,K)}$ on the query set $\mathcal{D}_m^{\text{qry}}$ serves as the optimization signal for meta-learning. Meta-parameter updates are realized through gradient descent, where the gradient depends on the sensitivity of $\theta_{\text{fast}}^{(m,K)}$ to θ_{meta}:

$$\theta_{\text{meta}} \leftarrow \theta_{\text{meta}} - \eta_{\text{meta}} \cdot \nabla_{\theta_{\text{meta}}} \mathcal{L}_m \tag{7}$$

where the loss is defined as $\mathcal{L}_m = \text{CrossEntropy}(\text{SNN}(\theta_{\text{fast}}^{(m,K)}, x_{\text{qry}}), y_{\text{qry}})$. The gradient $\nabla_{\theta_{\text{meta}}} \mathcal{L}_m$ is computed using the chain rule:

$$\nabla_{\theta_{\text{meta}}} \mathcal{L}_m = \nabla_{\theta_{\text{fast}}^{(m,K)}} \mathcal{L}_m \cdot \frac{\partial \theta_{\text{fast}}^{(m,K)}}{\partial \theta_{\text{meta}}} \tag{8}$$

Since $\theta_{\text{fast}}^{(m,K)} = \theta_{\text{lin}} + \sum_{k=0}^{K-1} \Delta\theta_{\text{heb}}(k)$ and $\Delta\theta_{\text{heb}}(k)$ depends only on spike activities, we have $\frac{\partial \theta_{\text{fast}}^{(m,K)}}{\partial \theta_{\text{meta}}} = I$. Therefore, the gradient simplifies to $\nabla_{\theta_{\text{meta}}} \mathcal{L}_m = \nabla_{\theta_{\text{fast}}^{(m,K)}} \mathcal{L}_m$. This simplification ensures that meta-learning can directly optimize the initial parameter space through Hebbian learning outcomes, creating an efficient feedback loop for continual learning.

Experiments

To comprehensively validate the effectiveness of the proposed HLML-SNN framework for fast task-incremental CL, this section conducts detailed experimental analysis from multiple dimensions: first clarifying experimental setup details including dataset partitioning, model configuration, and evaluation metrics; then conducting comparative analysis between HLML-SNN and classical continual learning methods as well as recent advanced methods on multiple benchmark datasets; exploring the impact of key parameters on model performance through hyperparameter sensitivity analysis; designing ablation studies to verify the necessity of each framework component; and finally conducting in-depth analysis of the framework's underlying mechanisms. All experiments are conducted on a workstation equipped with a single NVIDIA RTX4090 GPU, implemented based on the PyTorch framework to ensure reproducibility of results.

Experimental Setup

Datasets and Task Configuration. This paper conducts experimental validation on four benchmark datasets: split-MNIST, split-CIFAR10, split-CIFAR100, and split-TinyImageNet. split-MNIST and split-CIFAR10 are each partitioned into 5 tasks (2 classes per task), split-CIFAR100 is partitioned into 10/20 tasks (10/5 classes per task respectively), and split-TinyImageNet is partitioned into 10/20 tasks (20/10 classes per task respectively). For each task, training data is split into support and query sets at a 2:8 ratio.

Network Architecture. All experiments employ single hidden layer spiking neural network structures. The number of hidden layer neurons is adjusted according to dataset complexity: 256 neurons for split-MNIST and split-CIFAR10, and 1024 neurons for split-CIFAR100 and split-TinyImageNet.

Pre-Feature Extraction. For split-MNIST, split-CIFAR10, and split-CIFAR100, the pretrained feature extractor (PFE) is directly initialized using the official ImageNet pretrained weights. For split-TinyImageNet, since all of its 200 categories are included in ImageNet, two pretraining configurations are adopted to control the influence of category overlap: 1) Non-isolation setting (N): The PFE directly uses the official ImageNet pretrained weights, which naturally include the 200 overlapping categories. 2) Isolation setting (W): To remove potential category leakage, the 200 overlapping classes are excluded from ImageNet, and the PFE is retrained on the remaining 800 classes before being applied to split-TinyImageNet.

Hyperparameter Configuration and Evaluation Metrics. The initial Hebbian learning rate is set to $\eta_h(0) = 0.00001$ with decay factor $\gamma = 0.95$, and inner short-term iteration steps $k = 3$. The meta-learning rate $\eta_{\text{meta}} = 0.0001$ with cosine annealing strategy for learning rate scheduling. Time steps T are set according to dataset complexity: 4 for split-MNIST and split-CIFAR10, and 10 for split-CIFAR100 and split-TinyImageNet. Two metrics are employed for comprehensive evaluation: (1) Average Accuracy to measure performance in continual learning; (2) Training Time to assess computational efficiency. The final result is the mean and variance of 5 sets of random seeds.

Comparative Analysis with SOTA Methods

To validate the performance advantages of the HLML-SNN method, this paper conducts comprehensive comparative studies, selecting two categories of baseline methods covering classical continual learning paradigms and recent advanced methods. Classical continual learning methods include regularization based methods (SI (Zenke, Poole, and Ganguli 2017), LWF (Li and Hoiem 2017)), memory replay methods (ER (Chaudhry et al. 2019), A-GEM (Lopez-Paz and Ranzato 2017)), and structural expansion methods (XdG (Masse, Grant, and Freedman 2018), SepNet (Van de Ven, Tuytelaars, and Tolias 2022)). Recent methods include SNN-based methods (HLOP (Xiao et al. 2024), CH-HNN (Shi et al. 2025)) and ANN-based methods (EsaCL (Ren and Honavar 2024), HALRP (Li et al. 2024), RKR (Mazumder et al. 2022)). Experimental results are shown in Tables 1, 2.

As shown in Table 1, HLML-SNN demonstrates significantly superior performance compared to classical methods on complex datasets. On the split-CIFAR10 dataset, our method achieves 93.1% accuracy, improving by 8.59% compared to the second-best method LWF (84.51%), while the training time is reduced by about 3 times. On the 10-split-CIFAR100 dataset, the 90.2% accuracy substantially surpasses LWF (78.59%), and training time is reduced by about 7.5 times compared to ER (152.46s). Although performance on the simple split-MNIST dataset is comparable to the best method LWF (99.10%), our method's advantages are more prominent on complex tasks. This trend is more intuitively

Methods	split-MNIST		split-CIFAR10		10-split-CIFAR100	
	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)
SI	98.86	8.87	77.14	38.35	74.84	75.08
ER	98.77	15.92	82.18	39.54	76.43	152.46
XdG	98.76	7.90	80.73	25.98	69.86	60.97
SepNet	98.94	7.13	83.82	24.62	76.83	59.48
LWF	99.10	7.40	84.51	27.97	78.59	62.05
A-GEM	98.75	15.62	83.05	42.24	73.30	126.21
Ours	**99.1**	**4.73**	**93.1**	**7.91**	**90.2**	**22.02**
Ours^{+}	**98.96±0.19**	**4.87±0.19**	**92.56±0.78**	**7.90±0.25**	**89.52±0.93**	**21.73±0.59**

Ours^{+} indicates the average results over 5 times of random-seed training.

Table 1: Comparison Results with Classical Continual Learning Methods

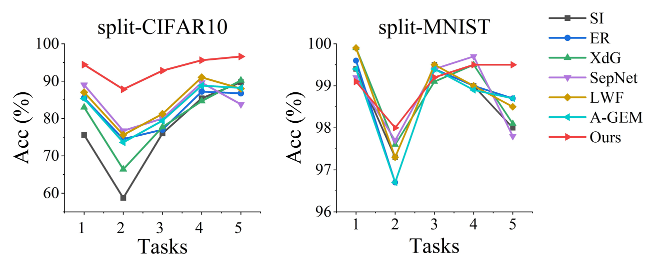

Figure 2: Performance comparison of the proposed model with other baseline models on the split-CIFAR10 (left) and split-MNIST (right) datasets.

demonstrated in Figure 2, the left panel shows that performance gaps among various methods are relatively small on the split-MNIST dataset, but HLML-SNN still maintains stable performance; the right panel on the split-CIFAR10 dataset indicates that HLML-SNN's accuracy curve consistently and significantly outperforms other methods, particularly as the number of tasks increases (such as tasks 4 and 5), where its performance degradation amplitude is much lower than comparative methods, validating our method's anti-forgetting capability in complex scenarios.

The comparison results in Table 2 demonstrate that on the 10-split-CIFAR100 dataset, HLML-SNN (90.2%) outperforms CH-HNN (87.45%) and RKR (88.49%), surpassing current advanced results for both ANNs and SNNs. On the 20-split-CIFAR100 dataset, it also achieves the best performance, proving our method's good adaptability to large-scale datasets. Regarding training efficiency, our PFE-based method leads to significant acceleration. Since HLOP trains end-to-end without a PFE, our method is approximately 420× faster on the 10-split task and 770× faster on the 20-split task. Compared with CH-HNN—which also uses a PFE—our method still achieves about 7–8× speedup while attaining higher accuracy. Furthermore, as shown in Table 5, this paper conducts comparisons with current advanced methods on the TinyImageNet dataset under both data isolation (W) and non-isolation (N) settings. On the 10-split-TinyImageNet (W/N) task, our method achieves an accuracy of 67.0%/80.1%, substantially higher than the representative ANN-based method EsaCL (53.16%). On the 20-split-TinyImageNet (W/N) task, it reaches 71.5%/73.3%, outperforming HALRP (70.09%) by a significant margin. These results validate the availability of our method on more

Methods	Type	PFE	10-split-CIFAR100		20-split-CIFAR100	
			Acc(%)	Time(s)	Acc(%)	Time(s)
EsaCL	ANN	—	73.62	—	—	—
HALRP	ANN	—	—	—	73.71	—
RKR	ANN	—	88.49	—	—	—
HLOP	T=20	✗	78.58	9274	70.33	22450
CH-HNN	T=10	✓	87.45	167	86.62	214
Ours	T=10	✓	**90.2**	**22**	**88.2**	**29**
Ours^{+}	T=10	✓	**89.52±0.93**	**21.73±0.59**	**87.24±1.08**	**28.19±0.50**

PFE indicates pretrained feature extractor.

Table 2: Comparison Results with Recent Advanced Continual Learning Methods

η_{meta}	1e-2	1e-3	1e-4	1e-5
Task1	87.1%	89.4%	93.9%	96.0%
Task2	70.7%	75.0%	83.9%	90.4%
Task3	87.4%	88.1%	92.8%	94.0%
Task4	98.1%	97.9%	96.5%	90.1%
Task5	97.9%	97.0%	96.6%	85.5%
Average Acc	88.2%	90.2%	**93.1%**	91.2%

Table 3: Experimental Results under Different Meta-Learning Rates

T	split-MNIST		split-CIFAR10	
	Acc(%)	Time(s)	Acc(%)	Time(s)
2	98.8	**3.21**	91.7	**6.21**
4	**99.1**	4.73	**93.1**	7.91

Table 4: Performance of split-MNIST and split-CIFAR10 under Different Time Steps

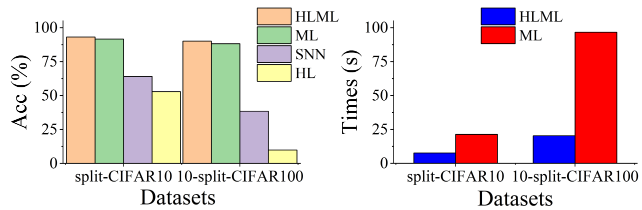

Figure 3: Comparison of accuracy and training time in the ablation study. (Left) Accuracy performance of different methods on two datasets. (Right) Corresponding training time comparison.

complex tasks.

Hyperparameter Sensitivity Analysis

Impact of Meta-Learning Rate The meta-learning rate is a key parameter controlling the magnitude of global parameter updates, significantly affecting the balance between model plasticity and stability. This paper sets the meta-learning rate to $\eta_{\text{meta}} \in 1e-2, 1e-3, 1e-4, 1e-5$ and tests its impact on the CIFAR10 dataset, with results shown in Table 3.

According to the results in Table 3, when the meta-learning rate is set too high, the model tends to excessively update the meta-parameters to adapt to new tasks, which leads to significant forgetting of previously learned knowledge. Conversely, when the meta-learning rate is too low, the model becomes overly conservative, favoring the retention of old knowledge while struggling to adapt to new tasks. Experimental results show that when the meta-learning rate falls within an optimal range (with an average accuracy of 93.1%), the model achieves a balanced parameter update that enables effective learning of new tasks while minimizing the forgetting of prior knowledge. Further details can be found in the appendix.

Impact of Time Steps Time steps determine the SNN's capability to capture temporal features. To explore the impact of time steps (T) on model performance, this paper conducts comparative experiments with T $\in$ 2, 4, 10 on different datasets, with results shown in Tables 4 and 5. The results indicate a significant correlation between the time step length and model performance. As shown in Table 4 for the split-MNIST and the split-CIFAR10, the longer time step length, the higher accuracy, while the slower convergence of the training process. Similarly, as shown in Table 5, in 10-split-TinyImageNet, the case T=10 (80.1%) improves upon the case T=4 (78.9%) by 1.2%. These results suggest that longer time steps provide the model with more sufficient parameter optimization space, facilitating the capture of inter-task feature correlations. However, increasing the time step also leads to a notable rise in total continual learning time. For instance, in 10-split-CIFAR100, the case T=10 takes 22.02 seconds, which is approximately 2.3 times longer than the case T=4 (9.54 seconds). Therefore, it is necessary to balance performance and efficiency based on task complexity.

Ablation Study

To evaluate the synergistic effect between Hebbian Learning (HL) and meta-learning (ML) in the HLML-SNN framework, we conducted three ablation experiments on the split-CIFAR10 and 10-split-CIFAR100 datasets:

1) Retaining only Hebbian learning while removing meta-learning (HL); 2) Retaining meta-learning while replacing Hebbian learning with standard update mechanisms (ML); 3) Combining both Hebbian learning and meta-learning (HLML). 4) In addition, a baseline SNN model without HL or ML was included for comparison.

The results in Figure 3 reveal that removing meta-learning (HL) causes the accuracy on split-CIFAR10 to drop sharply from 93.1% to 52.8%, and on 10-split-CIFAR100 from 90.2% to 9.9%, approaching random guess levels. This demonstrates that meta-learning plays a critical role in alleviating catastrophic forgetting by optimizing the global parameter space and providing a well-initialized starting point for task adaptation. When Hebbian learning is replaced with standard gradient descent (ML), the model still maintains high accuracy on both datasets, whereas, the training time is significantly increased by 3 to 5 times (as shown in the right subfigure), confirming the acceleration benefits of Hebbian learning. Overall, the HLML configuration outperforms the single mechanisms in both accuracy (93.1% / 90.2%) and training time (7.66s / 21.33s), and significantly exceeds the performance of the vanilla SNN model, validating the ef-

T	CIFAR100				TinyImageNet											
	10-split		20-split		10-split (W)		20-split (W)		10-split (N)		20-split (N)					
	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)				
4	88.7	**9.54**	86.8	**16.55**	64.1	**34.61**	68.7	**39.76**	78.9	**33.49**	69.6	**40.25**				
10	**90.2**	22.02	**88.2**	28.91	**67.0**	57.12	**71.5**	70.05	**80.1**	56.60	**73.3**	69.73				

W'/'N' indicate results with/without data sample isolation during pre-training.

Table 5: Comparison of split-CIFAR100 and split-TinyImageNet under Different Time Steps and Dataset Isolation Settings

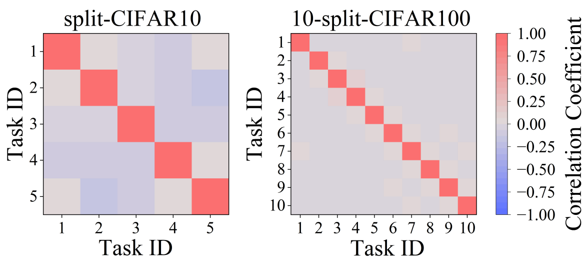

Figure 4: Correlation heatmaps of inter-task parameter update directions on the split-CIFAR10 (left) and 10-split-CIFAR100 datasets (right).

fectiveness and complementarity of the dual-phase learning mechanism.

Mechanism Analysis

To validate HLML-SNN's ability to mitigate catastrophic forgetting through parameter orthogonality—a key mechanism in continual learning—we quantitatively analyzed inter-task parameter correlations and intra-task update direction orthogonality on split-CIFAR10 and 10-split-CIFAR100 datasets. Results are shown in Figures 4 and 5.

At the inter-task level, the correlation analysis of parameter update directions (Figure 4) shows that in both datasets, the meta-parameter changes across tasks exhibit notably low correlation. Most of the off-diagonal elements in the correlation matrices fall within the range of [-0.05, 0.05], and are centered around zero. This indicates that the parameter updates for different tasks are nearly orthogonal, meaning that modifications for one task rarely affect the critical connections of another. This reflects successful task-specific and orthogonal parameter updates. From an intra-task fine-grained perspective, we further analyzed the angle between the parameter update direction at each training epoch and the accumulated parameter vector of previously learned tasks (Figure 5). The results show that as training progresses, the angle between the current update and the existing task parameters increases significantly, with most values clustering around 90°, indicating near-orthogonality. This suggests that even in early training phases, updates for new tasks remain largely non-invasive to previously learned knowledge.

Our results demonstrate that HLML-SNN achieves automatic parameter orthogonality across tasks and training stages without explicit constraints, enabling efficient new task acquisition while preserving prior knowledge. This inherent orthogonal learning mechanism provides the core foundation for HLML-SNN's continual learning performance. Validation experiments with explicit orthogonality

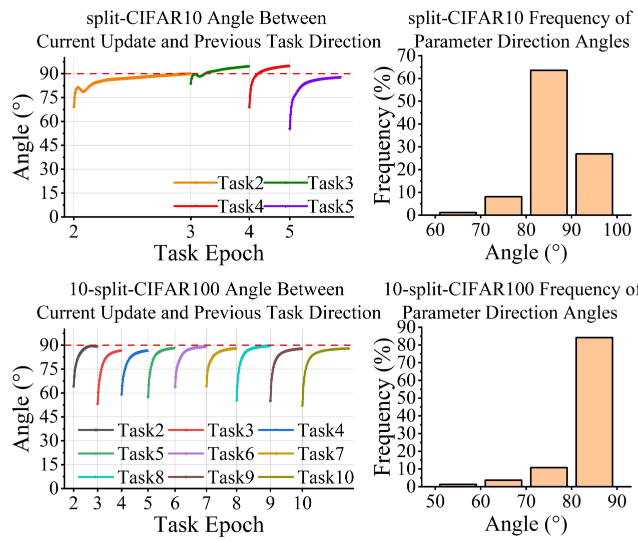

Figure 5: Curves and frequency distributions of angles between parameter update directions at each epoch of the current task and the accumulated weight vector directions of previously learned tasks, on the split-CIFAR10 (top) and 10-split-CIFAR100 (bottom) datasets.

constraints showed no performance gains but 4-fold increased training time, confirming the efficiency of our natural orthogonal learning approach. More experimental analysis can be found in the appendix.

Conclusion

Inspired by the cortico-hippocampal memory system, the proposed HLML-SNN framework integrates dual-phase neurocomputational mechanisms to achieve rapid adaptation with stable retention, enhancing its biological plausibility. The architecture embodies a hierarchical memory process, where Hebbian learning enables fast synaptic adaptation for short-term memory, meta-learning consolidates cross-task knowledge for long-term representation, and localized Hebbian updates mitigate interference through emergent parameter isolation without explicit regularization/orthogonality constaints. Experimental results demonstrate that HLML-SNN achieves state-of-the-art task-incremental CL performance while significantly reducing training iterations compared to existing methods. By unifying Hebbian local plasticity with meta-learning global optimization, HLML-SNN effectively addresses the stability–plasticity dilemma, establishing a new paradigm for fast and efficient task-incremental CL in resource-constrained environments.

References

Caporale, N.; and Dan, Y. 2008. Spike timing–dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.*, 31(1): 25–46.

Chaudhry, A.; Rohrbach, M.; Elhoseiny, M.; Ajanthan, T.; Dokania, P. K.; Torr, P. H.; and Ranzato, M. 2019. On tiny episodic memories in continual learning. *arXiv preprint arXiv:1902.10486*.

Chi, Z.; Gu, L.; Liu, H.; Wang, Y.; Yu, Y.; and Tang, J. 2022. Metafscil: A meta-learning approach for few-shot class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14166–14175.

Douillard, A.; Ramé, A.; Couairon, G.; and Cord, M. 2022. Dytox: Transformers for continual learning with dynamic token expansion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9285–9295.

Farajtabar, M.; Azizan, N.; Mott, A.; and Li, A. 2020. Orthogonal gradient descent for continual learning. In *International conference on artificial intelligence and statistics*, 3762–3773. PMLR.

Finn, C.; Rajeswaran, A.; Kakade, S.; and Levine, S. 2019. Online meta-learning. In *International conference on machine learning*, 1920–1930. PMLR.

Gerstner, W.; and Kistler, W. M. 2002. Mathematical formulations of Hebbian learning. *Biological cybernetics*, 87(5): 404–415.

Guo, Y.; Huang, X.; and Ma, Z. 2023. Direct learning-based deep spiking neural networks: a review. *Frontiers in Neuroscience*, 17: 1209795.

Hassabis, D.; Kumaran, D.; Summerfield, C.; and Botvinick, M. 2017. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2): 245–258.

Hospedales, T.; Antoniou, A.; Micaelli, P.; and Storkey, A. 2021. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9): 5149–5169.

Hunsberger, E.; and Eliasmith, C. 2015. Spiking deep networks with LIF neurons. *arXiv preprint arXiv:1510.08829*.

Hurtado, J.; Raymond, A.; and Soto, A. 2021. Optimizing reusable knowledge for continual learning via metalearning. *Advances in Neural Information Processing Systems*, 34: 14150–14162.

Izhikevich, E. M. 2003. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6): 1569–1572.

Javed, K.; and White, M. 2019. Meta-learning representations for continual learning. *Advances in neural information processing systems*, 32.

Kempter, R.; Gerstner, W.; and Van Hemmen, J. L. 1999. Hebbian learning and spiking neurons. *Physical Review E*, 59(4): 4498.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13): 3521–3526.

Kudithipudi, D.; Aguilar-Simon, M.; Babb, J.; Bazhenov, M.; Blackiston, D.; Bongard, J.; Brna, A. P.; Chakravarthi Raja, S.; Cheney, N.; Clune, J.; et al. 2022. Biological underpinnings for lifelong learning machines. *Nature Machine Intelligence*, 4(3): 196–210.

Li, J.; Lai, Y.; Wang, R.; Shui, C.; Sahoo, S.; Ling, C. X.; Yang, S.; Wang, B.; Gagné, C.; and Zhou, F. 2024. Hessian aware low-rank perturbation for order-robust continual learning. *IEEE Transactions on Knowledge and Data Engineering*, 36(11): 6385–6396.

Li, Z.; and Hoiem, D. 2017. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12): 2935–2947.

Lopez-Paz, D.; and Ranzato, M. 2017. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30.

Maass, W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9): 1659–1671.

Masse, N. Y.; Grant, G. D.; and Freedman, D. J. 2018. Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proceedings of the National Academy of Sciences*, 115(44): E10467–E10475.

Mazumder, P.; Singh, P.; Rai, P.; and Namboodiri, V. P. 2022. Rectification-based knowledge retention for task incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3): 1561–1575.

Mei, J.; Ma, Y.; Yang, X.; Wen, L.; Cai, X.; Li, X.; Fu, D.; Zhang, B.; Cai, P.; Dou, M.; et al. 2024. Continuously learning, adapting, and improving: A dual-process approach to autonomous driving. *arXiv preprint arXiv:2405.15324*.

Ni, W.; Shen, J.; Xu, Q.; and Tang, H. 2025. ALADE-SNN: Adaptive Logit Alignment in Dynamically Expandable Spiking Neural Networks for Class Incremental Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 19712–19720.

Perkonigg, M.; Hofmanninger, J.; Herold, C. J.; Brink, J. A.; Pianykh, O.; Prosch, H.; and Langs, G. 2021. Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging. *Nature communications*, 12(1): 5678.

Pes, L.; Luiken, R.; Corradi, F.; and Frenkel, C. 2024. Active dendrites enable efficient continual learning in time-to-first-spike neural networks. In *2024 IEEE 6th International Conference on AI Circuits and Systems (AICAS)*, 41–45. IEEE.

Ren, W.; and Honavar, V. G. 2024. EsaCL: An efficient continual learning algorithm. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*, 163–171. SIAM.

Roy, K.; Jaiswal, A.; and Panda, P. 2019. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784): 607–617.

Shen, J.; Ni, W.; Xu, Q.; and Tang, H. 2024. Efficient spiking neural networks with sparse selective activation for continual learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 611–619.

Shi, Q.; Liu, F.; Li, H.; Li, G.; Shi, L.; and Zhao, R. 2025. Hybrid neural networks for continual learning inspired by corticohippocampal circuits. *Nature Communications*, 16(1): 1272.

Sokar, G.; Mocanu, D. C.; and Pechenizkiy, M. 2021. Self-attention meta-learner for continual learning. *arXiv preprint arXiv:2101.12136*.

Son, J.; Lee, S.; and Kim, G. 2024. When meta-learning meets online and continual learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*.

Song, S.; Miller, K. D.; and Abbott, L. F. 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature neuroscience*, 3(9): 919–926.

Sun, W.; Advani, M.; Spruston, N.; Saxe, A.; and Fitzgerald, J. E. 2023. Organizing memories for generalization in complementary learning systems. *Nature neuroscience*, 26(8): 1438–1448.

Taherkhani, A.; Belatreche, A.; Li, Y.; Cosma, G.; Maguire, L. P.; and McGinnity, T. M. 2020. A review of learning in biologically plausible spiking neural networks. *Neural Networks*, 122: 253–272.

Tian, P.; Qi, L.; Dong, S.; Shi, Y.; and Gao, Y. 2021. Consistent MetaReg: alleviating intra-task discrepancy for better meta-knowledge. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2718–2724.

Turrigiano, G. G.; and Nelson, S. B. 2004. Homeostatic plasticity in the developing nervous system. *Nature reviews neuroscience*, 5(2): 97–107.

Van de Ven, G. M.; Tuytelaars, T.; and Tolias, A. S. 2022. Three types of incremental learning. *Nature Machine Intelligence*, 4(12): 1185–1197.

Vettoruzzo, A.; Bouguelia, M.-R.; Vanschoren, J.; Rögnvaldsson, T.; and Santosh, K. 2024a. Advances and challenges in meta-learning: A technical review. *IEEE transactions on pattern analysis and machine intelligence*, 46(7): 4763–4779.

Vettoruzzo, A.; Vanschoren, J.; Bouguelia, M.-R.; and Rögnvaldsson, T. 2024b. Learning to learn without forgetting using attention. *arXiv preprint arXiv:2408.03219*.

Wu, Y.; Huang, L.-K.; Wang, R.; Meng, D.; and Wei, Y. 2024. Meta continual learning revisited: Implicitly enhancing online hessian approximation via variance reduction. In *The Twelfth international conference on learning representations*, volume 2.

Xiao, M.; Meng, Q.; Zhang, Z.; He, D.; and Lin, Z. 2024. Hebbian learning based orthogonal projection for continual learning of spiking neural networks. *arXiv preprint arXiv:2402.11984*.

Ye, Y.; Xie, Y.; Zhang, J.; Chen, Z.; Wu, Q.; and Xia, Y. 2024. Continual self-supervised learning: Towards universal multi-modal medical data representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11114–11124.

Zeng, G.; Chen, Y.; Cui, B.; and Yu, S. 2019. Continual learning of context-dependent processing in neural networks. *Nature Machine Intelligence*, 1(8): 364–372.

Zenke, F.; Poole, B.; and Ganguli, S. 2017. Continual learning through synaptic intelligence. In *International conference on machine learning*, 3987–3995. PMLR.

Zhang, T.; Cheng, X.; Jia, S.; Li, C. T.; Poo, M.-m.; and Xu, B. 2023. A brain-inspired algorithm that mitigates catastrophic forgetting of artificial and spiking neural networks with low computational cost. *Science Advances*, 9(34): eadi2947.

Zhao, X.; Wang, H.; Huang, W.; and Lin, W. 2024. A statistical theory of regularization-based continual learning. *arXiv preprint arXiv:2406.06213*.

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.