CATFormer: When Continual Learning Meets Spiking Transformers With Dynamic Thresholds

Vaishnavi Nagabhushana, Kartikay Agrawal, Ayon Borthakur

SustainAI Lab, MFSDS&AI, IIT Guwahati, Assam, India
{n.vaishnavi, a.kartikay, ayon.borthakur}@iitg.ac.in

Abstract

Although deep neural networks perform extremely well in controlled environments, they fail in real-world scenarios where data isn't available all at once, and the model must adapt to a new data distribution that may or may not follow the initial distribution. Previously acquired knowledge is lost during subsequent updates based on new data. a phenomenon commonly known as catastrophic forgetting. In contrast, the brain can learn without such catastrophic forgetting, irrespective of the number of tasks it encounters. Existing spiking neural networks (SNNs) for class-incremental learning (CIL) suffer a sharp performance drop as tasks accumulate. We here introduce CATFormer (Context Adaptive Threshold Transformer), a scalable framework that overcomes this limitation. We observe that the key to preventing forgetting in SNNs lies not only in synaptic plasticity but also in modulating neuronal excitability. At the core of CATFormer is the Dynamic Threshold Leaky Integrate-and-Fire (DTLIF) neuron model, which leverages context-adaptive thresholds as the primary mechanism for knowledge retention. This is paired with a Gated Dynamic Head Selection (G-DHS) mechanism for task-agnostic inference. Extensive evaluation on both static (CIFAR-10/100/Tiny-ImageNet) and neuromorphic (CIFAR10-DVS/SHD) datasets reveals that CATFormer outperforms existing rehearsal-free CIL algorithms across various task splits, establishing it as an ideal architecture for energy-efficient, true-class incremental learning.

Introduction

Progress in physical AI holds immense promise for enhancing real-world capabilities across robotics, near-sensor edge devices, and autonomous systems. A critical challenge on these platforms is learning and predicting cyclically across extended deployments with minimal resource utilisation. Model updates are often essential due to sequentially arriving data and distributional shifts (Chaudhry et al. 2018; Wang et al. 2024a). But naively training standard deep neural networks (MLPs, CNNs, or even modern transformers) from scratch repeatedly typically results in *catastrophic forgetting* of previously acquired knowledge. In battery-operated, memory or bandwidth-constrained physical agents, data rehearsal (i.e., storage and replay of past data during training on new data) is often infeasible due to energy, onboard memory, privacy, or regulatory constraints (Lesort et al. 2020).

Energy-efficient learning architectures are essential in these applications, and **Spiking neural networks (SNNs)** have become a well-established solution, offering event-driven sparse computations. Recent advances have brought class-incremental learning (CIL) to SNNs with early efforts mainly on small-scale tasks (e.g., CIFAR-10) and, more recently, on larger, more realistic benchmarks like CIFAR-100 using various incremental task regimes (Ni et al. 2025; Han et al. 2023). However, prior efforts in SNN-based CL have predominantly relied on convolutional (CNN) architectures. Recently, transformers and their variants have dominated performance in numerous AI applications. Vision transformers (Dosovitskiy et al. 2020) can also capture global dependencies and contextual understanding. And, SpikFormer (Zhou et al. 2023) extends the standard vision transformer paradigm to SNNs by combining its strengths with energy efficiency. Yet, its potential for continual learning remains largely untapped. Hence, we here design CATFormer, a SNN-based transformer for long class incremental learning. CATFormer is inspired by the brain, where resistance to forgetting has been proposed to be closely linked to neuromodulation (Masse, Grant, and Freedman 2018; Beaulieu et al. 2020). Neuromodulators such as acetylcholine, dopamine, serotonin, and norepinephrine mediate changes in neural circuit behaviour by altering plasticity or excitability. For instance, acetylcholine plays a central role by modulating membrane excitability and synaptic plasticity in hippocampal and cortical networks, thereby enabling the rapid encoding of new memories while transiently lowering neuronal firing thresholds to suppress interference from prior information (Hasselmo and Barkai 1995; Grossberg 2017). Similar cholinergic modulation of plasticity is also observed in rodent piriform cortex (Hasselmo 2006). These processes in the brain dynamically regulate neuronal excitability and firing thresholds (Liu et al. 2022; Oh and Disterhoft 2015), enabling selective pathway activation and memory consolidation while preventing interference (Xu et al. 2005; Farmer and Thompson 2012). Computational models and experimental studies further demonstrate that neuromodulators broadly demonstrate task and context-specific routing of information by reshaping network dynamics (Tsuda et al. 2026; Masse, Grant, and Freedman 2018; Hammouamri, Masquelier, and Wilson 2022). Although we don't claim that CATFormer is completely biologically plausible, these

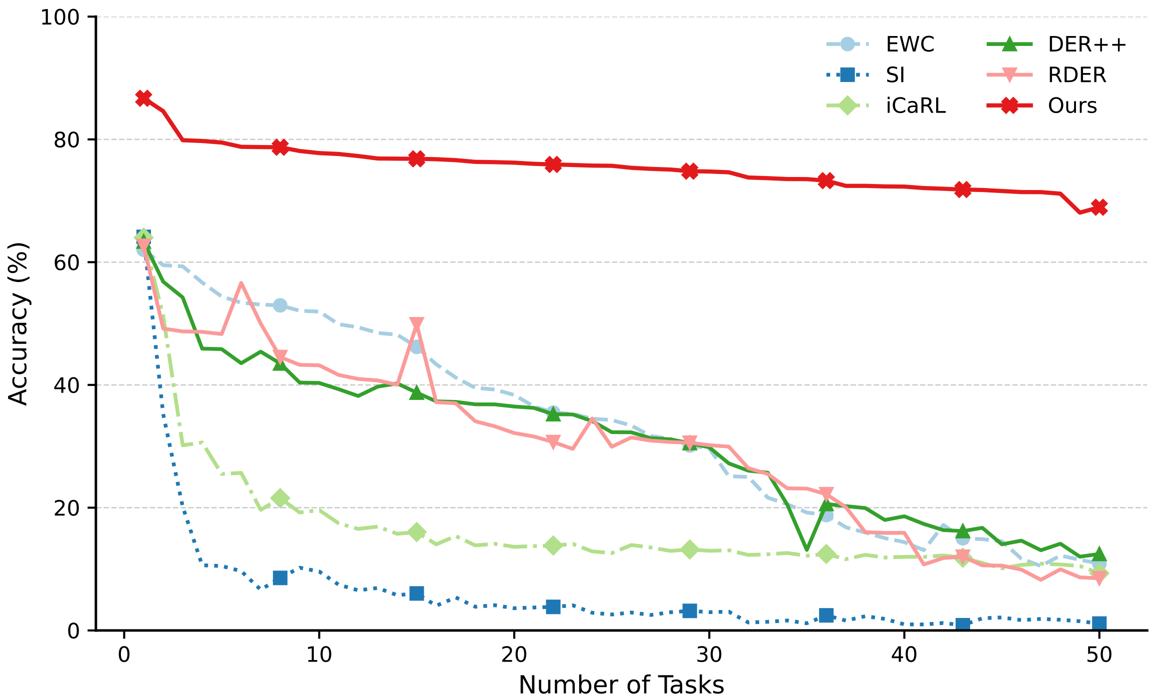

Figure 1: Test performance variation with respect to the progress in the number of trained tasks (for a maximum of 50 tasks). CATFormer (ours) maintains consistent performance with other existing CIL methods when implemented on a Spiking Transformer. All methods are evaluated on CIFAR 100.

multi-scale neuromodulatory effects inspire our approach to implement adaptive thresholds within CATFormer, serving as an analogy to support plasticity in exemplar-free continual learning.

In this work, we systematically study class incremental learning in a spiking vision transformer (Zhou et al. 2023). We analyse how biologically inspired, dynamic spiking thresholds influence continual learning and propose mechanisms that foster robustness across encountered tasks. This is crucial for *physical AI and robotics*: real-world robots and edge agents must adapt over months or years, often encountering dozens or even hundreds of different skills or operating regimes (Lesort et al. 2020). Thus, we rigorously evaluate our approach at unprecedented scales, including challenging 50 and 100-task sequences, providing a rigorous testbed for long-term continual learning relevant to autonomous robotics and physical edge applications. Moreover, our data rehearsal-free protocol ensures that results directly reflect core algorithmic advances, rather than storage-based workarounds. This can be observed in Figure 1, where the model maintains its accuracy across longer task sequences. We observe a phenomenon we call *reverse forgetting*, where the model actually learns more effectively when exposed to fewer classes per task. This scenario is more reflective of real-world settings—for example, in robotics or lifelong learning, the model is unlikely to encounter 20 or 50 new classes all at once (Lesort et al. 2020; Masse, Grant, and Freedman 2018).

We make the following advances in this work:

- To the best of our knowledge, we present the first CIL framework designed for spiking vision transformers, closing a major gap in the field.
- We propose a novel, biologically-inspired adaptation mechanism where a frozen backbone learns new tasks primarily through task-specific dynamic thresholds. This approach serves as the core mechanism for preventing catastrophic forgetting without requiring network growth or storing raw data exemplars.
- We demonstrate state-of-the-art performance among exemplar-free SNNs and, critically, show that CATFormer is uniquely scalable to long task sequences. Unlike prior methods that degrade, our model maintains, or even improves, its accuracy on challenging benchmarks comprising up to 100 incremental tasks, i.e., exhibiting a **reverse forgetting** trend.

Related Work

Continual Learning Paradigms

Continual learning methods often fall into three main approaches (Van de Ven and Tolias 2019). **Regularisation-based** methods mitigate forgetting by constraining updates to parameters deemed important for previous tasks, thus preventing overwriting of past knowledge (Kirkpatrick et al.

2017; Zenke, Poole, and Ganguli 2017). **Rehearsal-based** methods maintain a buffer of stored previous data samples either as original images or as representations for replay during training, improving memory retention but incurring increased storage and privacy concerns (Rebuffi et al. 2017; Buzzega et al. 2020; Zhu et al. 2022; Ye and Bors 2025). **Architecture-based** methods adapt the network structure dynamically(Han et al. 2023; Wang et al. 2025), such as by adding modules or selecting task-specific sub-networks, balancing plasticity and stability at the cost of computational overhead (Rusu et al. 2016; Fernando et al. 2017). While these architecture-based continual learning methods are effective, they have scalability limitations and memory overhead as the task count increases.

In the brain, context-dependent signals from regions like the prefrontal cortex project across cortical areas, allowing neural circuits to adaptively process information based on the task at hand (Engel, Fries, and Singer 2001; Miller and Cohen 2001). Previous techniques leveraged EWC (Kirkpatrick et al. 2017) with a gating mechanism to stabilise training for feedforward and recurrent architectures (Masse, Grant, and Freedman 2018).

Transformers in Continual Learning

Transformers, with their self-attention mechanisms, have recently emerged as a powerful alternative to convolutional neural networks (CNNs) for continual learning due to their ability to model long-range dependencies and the ease with which pretrained architectures can be extended (Wang et al. 2025; Liang and Li 2024). Recent work has focused on parameter-efficient fine-tuning techniques like Low-Rank Adaptation (LoRA) (He, Duan, and Zhu 2025; Liang and Li 2024) to update large pre-trained transformers continually with limited overhead (Wang et al. 2025; Li et al. 2025). However, these standard vision transformers require attention computations, which in turn lead to energy inefficiency due to their heavy matrix multiplications. Hence, we move towards spiking vision transformers (3.31 times more energy efficient) (Zhou et al. 2023), which can eliminate these heavy computations. Continual learning research on spiking vision transformers remains unexplored. Our work presents a data rehearsal-free continual learning on spiking vision transformers trained from scratch, addressing these challenges and expanding the landscape of SNN continual learning.

Spiking Neural Networks for Continual Learning and Neuromodulation Inspiration

Early SNN continual learning methods adapted classical regularisation and rehearsal methods that are limited to CNNs (Han et al. 2023; Lin et al. 2025). Recent work (Ni et al. 2025) shows significant improvement by incorporating rehearsal buffers into a method inspired by DER++ (Buzzega et al. 2020) to enhance performance, but it violates data privacy and is memory inefficient. Preliminary work on converting the idea of neuromodulation to the circuit level was demonstrated by (Hammouamri, Masquelier, and Wilson 2022), where the thresholds of the current layer were modulated based on the previous layer's excitation. Our model introduces adaptive, task-specific dynamic thresholds. This mimics neuromodulatory circuits that release context signals, modulating downstream neurons' responses, corresponding to our task-ID routing mechanism that dynamically tunes neuron thresholds per task to enable data-replay-free continual adaptation.

Dynamic Thresholds in Spiking Neural Networks

Adaptive firing thresholds have been shown to improve temporal precision and robustness in SNNs (Ding et al. 2022; Huang et al. 2016; Wei et al. 2023). Although many mechanisms have been extensively studied in neuroscience, only a handful of works have investigated bioinspired dynamic threshold rules to improve SNN generalisation. Existing approaches include a dynamic threshold update rule that adaptively scales firing thresholds to prevent excessive activity cite hao2020biologically, double exponential functions for threshold decay (Shaban, Bezugam, and Suri 2021), and predefined target firing counts (Kim et al. 2021). BDETT computes thresholds via average membrane potentials for neuronal homeostasis (Ding et al. 2022), but excessive spiking from highly active neurons reduces sparsity. However, prior work has predominantly focused on single-task adaptation rather than task-specific thresholding for continual learning. By freezing synaptic weights after the initial task and relying solely on learnable, per-task thresholds for plasticity, our approach introduces a novel mechanism to prevent catastrophic forgetting in a rehearsal-free framework. Together with lightweight task gating, this enables stable continual learning across up to 100 tasks on static and neuromorphic datasets, setting a new standard in memory and energy-efficient SNN continual learning.

Methodology

We introduce a data rehearsal-free framework for class-incremental learning (CIL) in Spiking Neural Networks (SNNs) that robustly accommodates new classes over time without catastrophic forgetting. Unlike existing systems, our entire system is trained without a separate exemplar representation buffer from previous learning across tasks (Ni et al. 2025; Buzzega et al. 2020). The proposed architecture leverages two key innovations: *task-specific (Context Adaptive) dynamic neuronal thresholds* and a *gated inference mechanism*, combined through a two-stage training protocol.

Problem Formulation and Notation

In CIL, the model is exposed to T number of tasks $\{\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_{T-1}\}$ sequentially, each with dataset $\mathcal{D}^k = \{(x_i^k, y_i^k)\}_{i=1}^{N_k}$ and disjoint label sets $\mathcal{Y}_k$ (i.e., $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$ for $i \neq j$). At each task k, the model must classify samples over the cumulative label space $\mathcal{Y}_{1:k}$, having access solely to the current task's train data during, and *without the any task oracle or previous task samples* at test time.

Context Adaptive Dynamic Threshold Neurons

Dynamic Threshold LIF Neuron Model: To enable task-adaptive spiking dynamics, we extend the standard

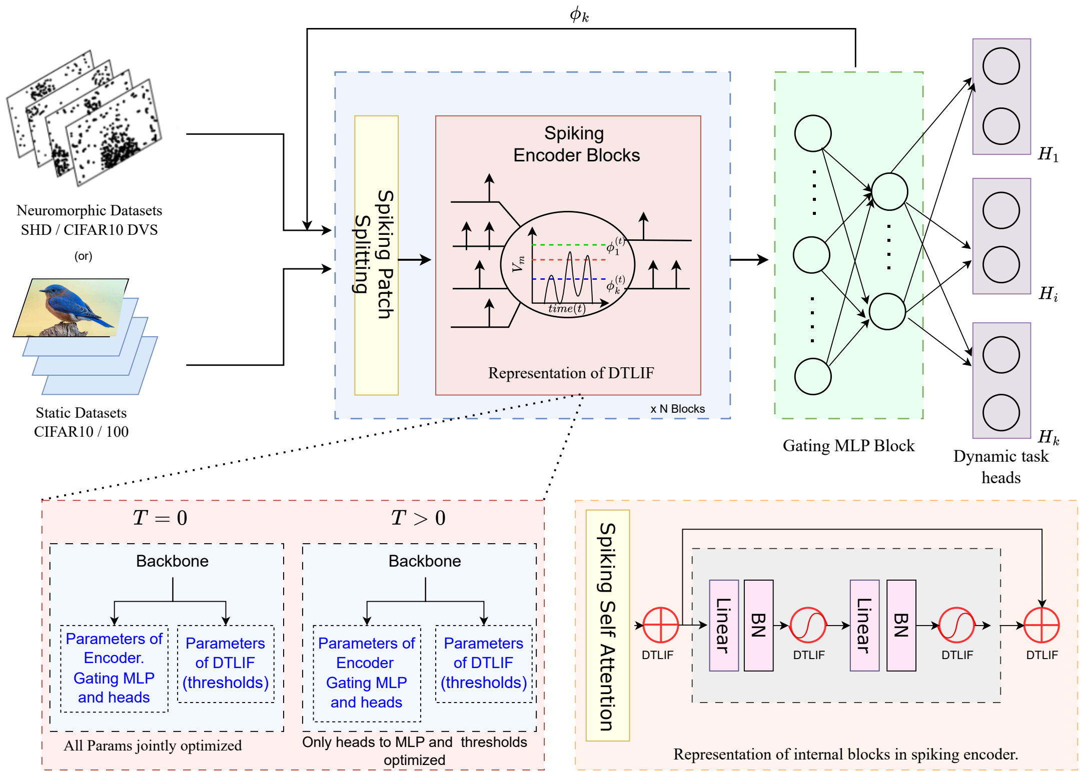

Figure 2: The diagram depicts the full architecture and workflow of CATFormer.

Algorithm 1: CATFormer Training Protocol

Input: Task sequence $\{\mathcal{T}_k, \mathcal{D}^k\}_{k=1}^{T}$
Initialize: Backbone θ, base threshold $\phi_{init} = 0.5$, gating MLP $\mathcal{G}$, buffers $\mathcal{F}_{\text{gate}}, \mathcal{L}_{\text{gate}}$
for $k = 0$ to $T - 1$ **do**
    Add head W_k (Xavier init), set $\phi^{(k)} \leftarrow \phi_{init}$
    if $k = 0$ **then**
        Train $\{\theta, \phi^{(0)}, W_0\}$ jointly with $\mathcal{L}_{\text{CE}}$
    else
        all previous parameters are frozen; initialize W_k for optimization.
        Jointly optimize $\{W_k, \phi^{(k)}\}$, i.e $\min_{\{\phi^{(k)}, W_k\}} \mathbb{E}_{(x,y)\sim\mathcal{D}^k} \left[\mathcal{L}_{\text{CE}}(W_k \cdot f(x; \theta, \phi^{(k)}), y)\right]$
    end if
    Extract features $\mathbf{f}(x)$ using ϕ_{init}; add $(\mathbf{f}(x), k)$ to buffers **(local scope not across tasks)**.
    Train gating MLP $\mathcal{G}$ on accumulated data with $\mathcal{L}_{\text{CE}}$
end for

Algorithm 2: Gated Inference

Input: Test sample x, trained model $\{\theta, \Phi, \{W_0 \cdots W_k\}, \mathcal{G}\}$, seen tasks k
Output: Predicted class $\hat{y}$
if $T = 0$ **then**
    Set thresholds to $\phi^{(0)}$ and reset SNN state
    Extract features: $\mathbf{f}(x) \leftarrow \text{SpikFormer}(x)$
    return $\arg\max(W_0\mathbf{f}(x))$
else
    Task Prediction:
        Set all thresholds to base ϕ_{init} and reset SNN state
        Extract base features: $\mathbf{f}_{\text{base}}(x) \leftarrow \text{SpikFormer}(x)$
        Predict task: $k^* \leftarrow \arg\max(\mathcal{G}(\mathbf{f}_{\text{base}}(x)))$
    Classification(once the head is selected):
        Set thresholds to $\phi^{(k^*)}$ and reset SNN state
        Extract task-specific features:
        $\mathbf{f}_{k^*}(x) \leftarrow \text{SpikFormer}(x)$
        return $\arg\max(W_{k^*}\mathbf{f}_{k^*}(x))$
end if

Leaky Integrate-and-Fire (LIF) neuron with *context adaptive, learnable thresholds*. Updation of thresholds :$\tilde{V}_j^{(t)} = \left(1 - \frac{1}{\tau}\right)V_j^{(t-1)} + \frac{1}{\tau}\, I_j^{(t)}$ where τ is the membrane time constant, $I_j^{(t)}$ is the input current. $S_j^{(t)} = \Theta\big(\tilde{V}_j^{(t)} - \phi_j^{(k)}\big)$ here $\Theta(\cdot)$ is the Heaviside step function and $S_j^{(t)}$ is the spike output. We use soft reset Mechanism $V_j^{(t)} = \tilde{V}_j^{(t)} - S_j^{(t)}\,\phi_j^{(k)}$.

Updation of Dynamic thresholds: During training on task k, the threshold $\phi_c^{(k)}$ is updated via gradient descent: $\phi_j^{(k)} \leftarrow \phi_j^{(k)} - \eta \frac{\partial \mathcal{L}}{\partial \phi_j^{(k)}}$ where η is the learning rate and $\mathcal{L}$

is the loss function. This mechanism allows each channel to adjust its firing threshold for different tasks, supporting task-adaptive spiking in continual learning.

Two-Stage Training Protocol

Our training protocol balances plasticity and stability by freezing and updating components, using threshold adaptation as the key mechanism to prevent catastrophic forgetting. Algorithm 1 outlines the complete training mechanism.

Inference via Dynamic Head Routing

During inference, we employ our Gated Dynamic Head Selection (G-DHS) mechanism to efficiently route inputs to appropriate task-specific heads.

Gating MLP Architecture. The gating network is used as the Gated Dynamic Head Selection (G-DHS), which consists of a two-layer MLP that maps feature embeddings to task predictions: $\mathcal{G}(\mathbf{f}) = \text{Linear}(\text{ReLU}(\text{Linear}(\mathbf{f})))$, where $\mathbf{f} \in \mathbb{R}^D \to \mathbb{R}^{D/4} \to \mathbb{R}^k$. Given an input x and D is the dimension of the feature vector, the inference mechanism is done for this as described by the algorithm 2

Results

Dataset and Experimental Setup

To evaluate the effectiveness of our context-adaptive dynamic threshold mechanism in a spiking Transformer (CATFormer), we conducted extensive experiments across a range of static and neuromorphic datasets. We evaluate on CIFAR-10/100 (10/100 classes) (Krizhevsky, Hinton et al. 2009): standard 32×32 RGB image benchmarks, tiny-ImageNet (200 classes) (Le, Yang et al. 2015): subset of ImageNet with 64×64 images, CIFAR10-DVS (10 classes) (Li et al. 2017): a neuromorphic, event-based version of CIFAR-10 captured with Dynamic Vision Sensors, and SHD (20 classes) (Cramer et al. 2020): a neuromorphic auditory dataset of spiking event sequences preprocessed into fixed-length frames.

Experimental Observations

Performance in Extended Task Sequences on static datasets We compared CATFormer with state-of-the-art methods for class incremental learning in spiking neural networks. For a fair comparison, we evaluated other methods on the same SpikFormer backbone and with respect to reverse forgetting. Further compared Tiny-ImageNet with a non-spiking benchmark with the same SpikFormer backbone.

Table 1 presents our class incremental learning (CIL) performance on Split CIFAR-100. Classical regularisation-based methods like EWC (Kirkpatrick et al. 2017), MAS (Aljundi et al. 2018), and SI (Zenke, Poole, and Ganguli 2017) suffer from severe catastrophic forgetting. Rehearsal-based approaches such as iCaRL (Rebuffi et al. 2017) and DER++ (Buzzega et al. 2020) offer better retention, but their memory requirements are often constrained on hardware like the Lakemont x86 processors and neuromorphic cores in Loihi 2 (Shrestha et al. 2024), making even the 2000 samples proposed by ALADE-SNN (Ni et al. 2025) pretty difficult. Among data rehearsal-free SNN baselines, previous state-of-the-art DSD-SNN (Han et al. 2023) achieves moderate performance but exhibits a consistent forgetting pattern. The accuracy consistently degrades from **60.47%** at 10 tasks to **50.55%** at 50 tasks. To confirm the trend, we extended the original DSD-SNN repository[1] to generate 25 and 50 task results. On the contrary, **CATFormer fundamentally breaks this degradation pattern.** Our model not only surpasses the classical CIL baselines by substantial margins but also demonstrates the unprecedented behaviour of *improving* accuracy with increasing tasks from **68.33%** at 10 tasks to **75.66%** at 50 tasks. This is a counter-intuitive 'reverse forgetting' trend, as illustrated in Figure 3. This effect stems from our dynamic threshold adaptation with the gating mechanism, which optimises neuronal firing for new tasks without overwriting prior knowledge. This trend toward long task sequences is crucial for real-world applications, as highlighted in robotics and embodied AI research (Lesort et al. 2020; Hajizada et al. 2022), which involve continuous adaptation and data drift. We evaluated our model performance on the 10-class count CIFAR-10 dataset. The comparative results in Table 2, with the best-performing model, DSD-SNN (Shen et al. 2024), serving as the baseline, demonstrate CATFormer's superior performance on Split CIFAR-10, achieving an accuracy of **89.29%** for the 5-task split.

Comparison of parameter updates at k^{th} task In terms of model size, our proposed architecture is parameter-efficient. The base SpikFormer (Zhou et al. 2023) has a base size of 9.32M parameters, while our model has a base size of 10.5M, of which 1.2M is attributed to the routing mechanism. Although there is a minimal increase in the number of parameters, it is comparable to other methods using the same SpikFormer backbone. For instance, a task 0 CATFormer is comparable to most of the current training paradigms presented in Table 1. Our approach becomes significantly parameter-efficient once we move to tasks T^k, where $k > 0$, where we approximately train only 1.4 million parameters, while the other model actually updates the entire model's parameters and also utilises a rehearsal buffer in some cases (which can also lead to unwanted privacy breaches). The total parameter count for our model scales efficiently with the term count (ϕ^k), where approximately 16,032 thresholds per task are stored, requiring no more than 64.2 KB of memory when FP32 is used for storage. This can eventually be reduced to FP16, since we hardly need many decimal places, indicating that only a small number of task-specific parameters are added as new tasks are learned, while achieving state-of-the-art performance.

Performance on tiny-ImageNet Dataset We further evaluate CATFormer's performance on the 200-class count Tiny-ImageNet dataset. Since CIFAR-100 allows up to 50 task training, we test the efficacy of our model on 100 tasks that were not possible earlier. Due to the unavailability of a previous similar CIL implementation of SNNs on ImageNet, we

[1] https://github.com/BrainCog-X/Brain-Cog.git

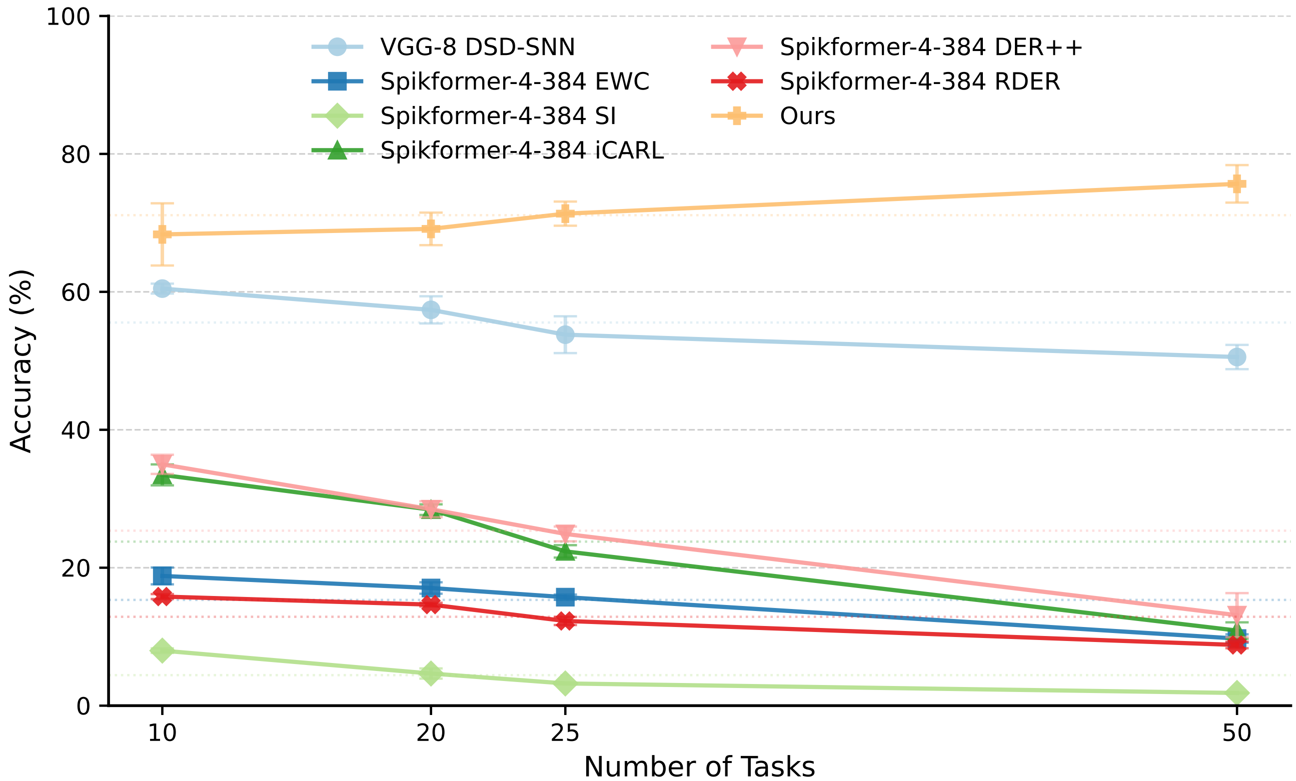

Figure 3: Reverse Forgetting vs Catastrophic Forgetting Trend comparison (No. of Task vs Accuracy(in %)) of CATFormer against DSD-SNN (Han et al. 2023) and same SpikFormer on (Kirkpatrick et al. 2017; Zenke, Poole, and Ganguli 2017; Rebuffi et al. 2017; Buzzega et al. 2020; Wang et al. 2024b). The dotted line represents the average accuracy across tasks as the number of tasks increases.

Backbone	**Methods**	**Number of Tasks**				**Parameters (M)**	
Traditional		**1 Task (Full dataset)**				**Total**	
VGG-11	Hybrid	67.87				9.27	
ResNet-19	TET	74.47				12.63	
SpikFormer-4-384	BPTT	77.86				9.32	
Class incremental Learning		**10 Tasks**	**20 Tasks**	**25 Tasks**	**50 Tasks**	**Task 0**	**Task k**
VGG-8	DSD-SNN	60.47 ± 0.72	57.39 ± 1.97	53.79 ± 2.67	50.55 ± 1.76	14.2	14.2
SpikFormer-4-384	EWC	18.81 ± 1.22	17.06 ± 0.83	15.73 ± 0.38	9.73 ± 0.62	9.32	18.64
	SI	7.98 ± 0.33	4.66 ± 0.74	3.22 ± 0.14	1.84 ± 0.09	9.32	27.96
	iCARL	33.46 ± 1.52	28.42 ± 0.77	22.37 ± 0.90	10.89 ± 1.19	9.32	9.32
	DER++	34.99 ± 1.39	28.48 ± 1.16	24.9 ± 1.07	13.12 ± 3.2	9.32	9.32
	RDER	15.82 ± 0.36	14.65 ± 0.27	12.28 ± 0.59	8.8 ± 0.47	11.06	11.06
	Ours	$\mathbf{68.33 \pm 4.51}$	$\mathbf{69.13 \pm 2.36}$	$\mathbf{71.34 \pm 1.75}$	$\mathbf{75.66 \pm 2.72}$	10.5	**1.4**

Table 1: Comparison of standard CIL accuracy (%) on Split CIFAR-100 across different task granularities. Reported average test accuracy after all tasks. DSD-SNN results (25/50 tasks)(Han et al. 2023); other CIL baselines (Kirkpatrick et al. 2017; Zenke, Poole, and Ganguli 2017; Rebuffi et al. 2017; Buzzega et al. 2020; Wang et al. 2024b) are evaluated on SpikFormer (Zhou et al. 2023). Task 0 and Task k describe the total number of parameter updates during the 0^{th} and k^{th} training tasks, respectively.

compare performance to an ANN baseline (Liu et al. 2025), which features a non-transformer backbone (Gu and Dao 2024) on (Yan et al. 2024). We observe **8.45%** improvement over the baselines (Table 2). Hence, CATFormer achieves better overall performance than non-spiking models across a larger number of tasks.

Performance on Spiking Datasets Neuromorphic datasets pose challenges due to their spatiotemporal dynamics and event-driven inputs, making them ideal testbeds for our dynamic threshold hypothesis. Table 2 demonstrates CATFormer's efficacy on SHD (with only timesteps of 16) and CIFAR10-DVS benchmarks, achieving **84.48%/87.85%** and **83.21%/87.14%** for 5/10 and 2/5 task splits, respectively. Notably, unlike our other computer vision evaluation datasets, namely CIFAR-10/100, CIFAR10-DVS, SHD is an audio-based classification dataset. We observe that, even without any data rehearsal strategy, CATFormer's **83.21%** performance on CIFAR10-DVS (2 tasks) closely matches the rehearsal augmented performance of ALADE-SNNs (Ni et al. 2025), 83.5%. CATFormer thereby establishes a new standard for data rehearsal-free neuromorphic continual learning.

The consistent performance gains across neuromorphic datasets validate that our context-adaptive threshold mechanism naturally aligns with the temporal processing characteristics inherent to spiking neural networks, enabling more effective utilisation of the temporal dimension for task differentiation. This neuromorphic compatibility represents a significant advancement, as these datasets remain underexplored in data rehearsal-free continual learning scenarios.

Dataset	Task	Model	Accuracy
CIFAR10	5	SA-SNN+EWC	80.39 ± 1.84
		CATFormer	**89.29 ± 2.53**
CIFAR10 -DVS	2	DSD-SNN	80.90 ± 1.20
		CATFormer	**83.21 ± 2.33**
	5	DSD-SNN	76.57 ± 0.96
		CATFormer	**87.14 ± 2.78**
SHD (T=16)	5	DSD-SNN	82.56 ± 1.15
		CATFormer	**84.48 ± 1.62**
	10	DSD-SNN	80.47 ± 1.03
		CATFormer	**87.85 ± 1.20**
Tiny Im-Net	**100**	S6MOD*	40.11 ± 0.26
		CATFormer	**48.56 ± 0.81**

Table 2: Task-wise accuracy (%) of CATFormer with respect to the state-of-the-art SNNs on static and neuromorphic datasets. Methods marked with * denote online continual learning approaches.

Ablation Studies

We conducted targeted ablation experiments on Split CIFAR-10 (5 tasks, 2 classes per task) to isolate the impact of each core component in CATFormer. The ablated variants and their average accuracy (in %) after all tasks are described in Table 3. The **Fixed Threshold** variant, where all neurons use their initial firing threshold for all tasks, leads to pronounced catastrophic forgetting, with accuracy dropping to 42.87%. In this configuration, performance on the first task is reasonable, but subsequent tasks trigger a consistent 15–18% degradation after each increment, mirroring classical forgetting patterns in SNNs without adaptive mechanisms. Hence, we observe that, when studying catastrophic forgetting in brain-inspired SNNs, it is important to pay close attention to the role of spiking thresholds, unlike previous studies (Ni et al. 2025; Han et al. 2023; Shen et al. 2024). Biologically, such dynamic threshold behavior can potentially be mediated by neuromodulation (Tsuda et al. 2026; Oh and Disterhoft 2015; Liu et al. 2022; Xu et al. 2005).

Ablation Variant	Accuracy	Acc. on task 0
Fixed Threshold	42.87±1.26	72.59±1.86
SpikIdentityFormer	59.38±0.98	70.62±1.75
Random Identity Former	53.17±2.13	62.43±0.99
FFN Frozen	63.24±1.78	72.17±1.59
CATFormer	**89.29±2.53**	**93.87±0.45**

Table 3: Ablation study of CATFormer evaluated on CIFAR-10 (5 tasks).

Recent works on vision transformer (Yu et al. 2022, 2023) show that its performance is not significantly impacted by the choice of token mixers. In fact, they show that even after removing all vanilla attention token mixers, the ViT achieves good baseline performance. Driven by these observations, we evaluate the performance of our spiking vision transformer (CATFormer) on the CIFAR-10 (5 tasks) platform. As part of this, we design **SpikIdentityFormer** by either replacing all spike attention with identity mapping in CATFormer (**SpikIdentityFormer**) or replacing spike attention with uniformly distributed random numbers in CATFormer (**Random Identity Former**). Both the **SpikIdentityFormer** (removing all attention) and **Random Identity Former** (disrupting transformer blocks) yield similar, markedly reduced accuracy, confirming that structured feature transformation is crucial for robust continual learning. Significantly, a **Random Identity Former** performs lower (53.17%) than a **SpikIdentityFormer** (59.38%). Moreover, freezing the feed-forward network (**FFN Frozen**) provides only moderate improvement (63.24%), indicating that adaptive intermediate representations are also necessary for effective knowledge retention over time. Compared to only FFN learning in **SpikIdentityFormer**, a learnable token mixer in **FFN Frozen** (with FFN frozen) performs better by 3.86%. A similar trend is observed in first-task learning. These results demonstrate that both the transformer's structured feature extraction and, critically, its dynamic, task-specific threshold modulation are essential for effective, data-rehearsal-free continual learning, thereby justifying our design choices. Overall, CATFormer surpasses prior rehearsal-free methods on static and neuromorphic benchmarks, achieving stable or improved accuracy as the task count increases. This makes it well-suited for continual learning on embedded, memory-constrained robotic and autonomous systems that must adapt to dynamic, unpredictable environments. Unlike conventional rehearsal-based methods, which face prohibitive storage, energy, privacy, and bandwidth constraints, CATFormer almost eliminates the need for memory buffers by leveraging dynamic, task-specific neuronal plasticity. This biologically inspired and hardware-efficient design aligns with practical constraints highlighted in robotics and embodied AI research (Lesort et al. 2020; Hajizada et al. 2022), enabling robust lifelong continual learning in real-world deployment.

Discussion

CATFormer demonstrates that biologically inspired dynamic threshold adaptation enables rehearsal-free continual learning in spiking neural networks, maintaining or improving accuracy across up to 50 tasks. This is especially relevant for robotics and physical AI, where limited onboard memory makes storing replay buffers impractical—for instance, retaining 2,000 CIFAR-100 images for rehearsal consumes (Ni et al. 2025) approximately 25–30MB, a significant overhead for resource-constrained hardware. Moreover, continual network growth or pruning introduces complexity and unstable resource demands, hindering deployment on embedded or neuromorphic platforms. By leveraging intrinsic neuronal modulation through task-specific dynamic thresholds, CATFormer provides a memory and computation-efficient solution that avoids these pitfalls while sustaining robust, scalable continual learning. Future work should explore adaptive threshold learning in streaming, non-stationary environments for *lifelong learning* and investigate direct hardware implementations on Loihi 2 (Shrestha et al. 2024) to accelerate the real-world deployment of lifelong SNN agents.

Acknowledgments

This work was supported by the Faculty Startup Grant from IIT Guwahati.

References

Aljundi, R.; Babiloni, F.; Elhoseiny, M.; Rohrbach, M.; and Tuytelaars, T. 2018. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European conference on computer vision (ECCV)*, 139–154.

Beaulieu, S.; Frati, L.; Miconi, T.; Lehman, J.; Stanley, K. O.; Clune, J.; and Cheney, N. 2020. Learning to continually learn. *arXiv preprint arXiv:2002.09571*.

Buzzega, P.; Boschini, M.; Porrello, A.; Abati, D.; and Calderara, S. 2020. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33: 15920–15930.

Chaudhry, A.; Ranzato, M.; Rohrbach, M.; and Elhoseiny, M. 2018. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420*.

Cramer, B.; Stradmann, Y.; Schemmel, J.; and Zenke, F. 2020. The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7): 2744–2757.

Ding, J.; Dong, B.; Heide, F.; Ding, Y.; Zhou, Y.; Yin, B.; and Yang, X. 2022. Biologically inspired dynamic thresholds for spiking neural networks. *Advances in neural information processing systems*, 35: 6090–6103.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Engel, A. K.; Fries, P.; and Singer, W. 2001. Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10): 704–716.

Farmer, G. E.; and Thompson, L. T. 2012. Learning-dependent plasticity of hippocampal CA1 pyramidal neuron postburst afterhyperpolarizations and increased excitability after inhibitory avoidance learning depend upon basolateral amygdala inputs. *Hippocampus*, 22(8): 1703–1719.

Fernando, C.; Banarse, D.; Blundell, C.; Zwols, Y.; Ha, D.; Rusu, A. A.; Pritzel, A.; and Wierstra, D. 2017. Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*.

Grossberg, S. 2017. Acetylcholine neuromodulation in normal and abnormal learning and memory: vigilance control in waking, sleep, autism, amnesia and Alzheimer's disease. *Frontiers in neural circuits*, 11: 82.

Gu, A.; and Dao, T. 2024. Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*.

Hajizada, E.; Berggold, P.; Iacono, M.; Glover, A.; and Sandamirskaya, Y. 2022. Interactive continual learning for robots: a neuromorphic approach. In *Proceedings of the international conference on neuromorphic systems 2022*, 1–10.

Hammouamri, I.; Masquelier, T.; and Wilson, D. G. 2022. Mitigating Catastrophic Forgetting in Spiking Neural Networks through Threshold Modulation. *Transactions on Machine Learning Research*.

Han, B.; Zhao, F.; Zeng, Y.; Pan, W.; and Shen, G. 2023. Enhancing efficient continual learning with dynamic structure development of spiking neural networks. *arXiv preprint arXiv:2308.04749*.

Hasselmo, M. E. 2006. The role of acetylcholine in learning and memory. *Current opinion in neurobiology*, 16(6): 710–715.

Hasselmo, M. E.; and Barkai, E. 1995. Cholinergic modulation of activity-dependent synaptic plasticity in the piriform cortex and associative memory function in a network biophysical simulation. *Journal of Neuroscience*, 15(10): 6592–6604.

He, J.; Duan, Z.; and Zhu, F. 2025. CL-LoRA: Continual low-rank adaptation for rehearsal-free class-incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30534–30544.

Huang, C.; Resnik, A.; Celikel, T.; and Englitz, B. 2016. Adaptive spike threshold enables robust and temporally precise neuronal encoding. *PLoS computational biology*, 12(6): e1004984.

Kim, T.; Hu, S.; Kim, J.; Kwak, J. Y.; Park, J.; Lee, S.; Kim, I.; Park, J.-K.; and Jeong, Y. 2021. Spiking neural network (SNN) with memristor synapses having non-linear weight update. *Frontiers in computational neuroscience*, 15: 646125.

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.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Le, Y.; Yang, X.; et al. 2015. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7): 3.

Lesort, T.; Lomonaco, V.; Stoian, A.; Maltoni, D.; Filliat, D.; and Díaz-Rodríguez, N. 2020. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information fusion*, 58: 52–68.

Li, H.; Liu, H.; Ji, X.; Li, G.; and Shi, L. 2017. Cifar10-dvs: an event-stream dataset for object classification. *Frontiers in neuroscience*, 11: 244131.

Li, J.; Wang, S.; Qian, B.; He, Y.; Wei, X.; Wang, Q.; and Gong, Y. 2025. Dynamic integration of task-specific adapters for class incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30545–30555.

Liang, Y.-S.; and Li, W.-J. 2024. Inflora: Interference-free low-rank adaptation for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23638–23647.

Lin, E.; Luo, W.; Jia, W.; Chen, Y.; and Yang, S. 2025. Online Continual Learning via Spiking Neural Networks with Sleep Enhanced Latent Replay. *arXiv preprint arXiv:2507.02901*.

Liu, S.; Yang, Y.; Li, X.; Clifton, D. A.; and Ghanem, B. 2025. Enhancing online continual learning with plug-and-play state space model and class-conditional mixture of discretization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20502–20511.

Liu, Y. H.; Smith, S.; Mihalas, S.; Shea-Brown, E.; and Sümbül, U. 2022. Biologically-plausible backpropagation through arbitrary timespans via local neuromodulators. *Advances in Neural Information Processing Systems*, 35: 17528–17542.

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.

Miller, E. K.; and Cohen, J. D. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1): 167–202.

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.

Oh, M. M.; and Disterhoft, J. F. 2015. Increased excitability of both principal neurons and interneurons during associative learning. *The Neuroscientist*, 21(4): 372–384.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2001–2010.

Rusu, A. A.; Rabinowitz, N. C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; and Hadsell, R. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

Shaban, A.; Bezugam, S. S.; and Suri, M. 2021. An adaptive threshold neuron for recurrent spiking neural networks with nanodevice hardware implementation. *Nature Communications*, 12(1): 4234.

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.

Shrestha, S. B.; Timcheck, J.; Frady, P.; Campos-Macias, L.; and Davies, M. 2024. Efficient video and audio processing with loihi 2. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 13481–13485. IEEE.

Tsuda, B.; Pate, S. C.; Tye, K. M.; Siegelmann, H. T.; and Sejnowski, T. J. 2026. Neuromodulators Generate Multiple Context-Relevant Behaviors in Recurrent Neural Networks. *Neural Computation*, 1–36.

Van de Ven, G. M.; and Tolias, A. S. 2019. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*.

Wang, H.; Lu, H.; Yao, L.; and Gong, D. 2025. Self-expansion of pre-trained models with mixture of adapters for continual learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 10087–10098.

Wang, L.; Zhang, X.; Su, H.; and Zhu, J. 2024a. A comprehensive survey of continual learning: Theory, method and application. *IEEE transactions on pattern analysis and machine intelligence*, 46(8): 5362–5383.

Wang, Q.; Wang, R.; Li, Y.; Wei, D.; Wang, H.; Ma, K.; Zheng, Y.; and Meng, D. 2024b. Relational experience replay: Continual learning by adaptively tuning task-wise relationship. *IEEE Transactions on Multimedia*, 26: 9683–9698.

Wei, W.; Zhang, M.; Qu, H.; Belatreche, A.; Zhang, J.; and Chen, H. 2023. Temporal-coded spiking neural networks with dynamic firing threshold: Learning with event-driven backpropagation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10552–10562.

Xu, J.; Kang, N.; Jiang, L.; Nedergaard, M.; and Kang, J. 2005. Activity-dependent long-term potentiation of intrinsic excitability in hippocampal CA1 pyramidal neurons. *Journal of Neuroscience*, 25(7): 1750–1760.

Yan, H.; Wang, L.; Ma, K.; and Zhong, Y. 2024. Orchestrate latent expertise: Advancing online continual learning with multi-level supervision and reverse self-distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 23670–23680.

Ye, F.; and Bors, A. G. 2025. Online task-free continual learning via dynamic expansionable memory distribution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 20512–20522.

Yu, W.; Luo, M.; Zhou, P.; Si, C.; Zhou, Y.; Wang, X.; Feng, J.; and Yan, S. 2022. MetaFormer Is Actually What You Need for Vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10819–10829.

Yu, W.; Si, C.; Zhou, P.; Luo, M.; Zhou, Y.; Feng, J.; Yan, S.; and Wang, X. 2023. Metaformer baselines for vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2): 896–912.

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

Zhou, Z.; Zhu, Y.; He, C.; Wang, Y.; YAN, S.; Tian, Y.; and Yuan, L. 2023. Spikformer: When Spiking Neural Network Meets Transformer. In *The Eleventh International Conference on Learning Representations*.

Zhu, K.; Zhai, W.; Cao, Y.; Luo, J.; and Zha, Z.-J. 2022. Self-sustaining representation expansion for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9296–9305.