Highlights

A Survey on the intersections between Continual Learning, Neuroscience, and Neuromorphic Continual Learning

S Falciglia

- Spike neural networks (SNNs), due to their closeness to biological neural networks, are less prone to fundamental problems that afflict standard ANNs.
- We categorise representative continual learning methods at the level of synaptic plasticity and at the level of regional cooperation, according to the neurological basis studied in neuroscience.

A Survey on the intersections between Continual Learning, Neuroscience, and Neuromorphic Continual Learning

S Falciglia[a,1]

[a]Department of Computer, Control and Management Engineering, Via Ariosto 25, Rome, 00185, Italy,

ARTICLE INFO

Keywords:
lifelong learning
continual learning
neuromorphic continual learning
catastrophic forgetting
synaptic plasticity
regional collaboration

ABSTRACT

One final brain-inspired capability we want our intelligent agent to gain is lifelong learning (LLL), also known as continual learning (CL). This capability refers to the ability of living beings to acquire information, knowledge, and experience incrementally throughout their lifetime. In contrast to biological neural networks, artificial neural networks (ANNs) are severely affected by the problem of catastrophic forgetting in sequential learning; that is, they forget past experiences abruptly when learning a new task. On the other hand, due to their similarity to biological neural networks, neuromorphic networks such as spike neural networks (SNNs) possess properties that could result in the development of faster and more energy-efficient systems. These systems may be less susceptible than standard ANNs to the fundamental problems that plague them. The purpose of this work is to provide a detailed overview of lifelong learning mechanisms by analyzing the actual brain mechanisms and then correlating them with the options for implementing architectures and training for ANNs and SNNs. Our findings reveal that there is a trade-off between the computational cost and efficiency of continual learning scenarios, and this trade-off results from their design.

1. Introduction

Continuous Learning in the context of AI is a field that intends to provide AI systems with the same adaptability that humans demonstrate throughout their childhood, adolescence, and adulthood [1]. Although people seem to excel at continuous learning, Deep Learning systems need to overcome several major challenges such as data distribution shifts, task interference, and computational efficiency. In real-world scenarios, neural networks are trained in a controlled environment with a pre-determined and limited set of samples that contain interleaved classes. On-chip learning allows for the periodic training of a network with changing data in deployed systems. However, when the networks are trained on a new class, they overwrite the configuration of their parameters that enabled them to perform the first task, and they modify these parameters to achieve the best configuration for performing the second of the two subsequent tasks. This phenomenon is known as catastrophic forgetting [2].

Modern strategies implemented for deep learning systems are now grounded in biological neural networks, as studies in neuroscience have shown. This opens up possibilities for massive deployment of low-power neuromorphic computing systems, which are inspired by biological neural networks, leading to a reduced carbon footprint of our neural network architectures [3]. The algorithms and applications designed to operate within neuromorphic networks, achieving the aim of continual learning, are now categorized as Neuromorphic Continual Learning (NCL) paradigms [4], with a growing interest in this field.

This work aims to provide an original perspective on Continual Learning, focusing on the parallelism between state-of-the-art strategies used for deep learning systems and NCL systems' state-of-the-art paradigms, as well as their biological counterparts' mechanisms studied in neuroscience. The rest of the paper is organized as follows. Section 2 presents a brief overview of the neurological basis of continual learning. In Section 3 we suggest a neuroscience perspective on the elaborated taxonomy of representative continual learning methods, presented by [5]. In Section 4 we focus on the state-of-the art methods for artificial neural networks. Finally, Section 5 draws the conclusions of this study.

falciglia.2015426@studenti.uniroma1.it (S. Falciglia)
ORCID(s): 0009-0009-4871-2553 (S. Falciglia)

2. Continual Learning in the brain

As suggested by [6], a deeper understanding of the brain could be crucial for the advancement of intelligent systems. Here we present a brief overview of the neurological foundations of continuous learning mechanisms, from synaptic plasticity to regional coordination. The fact that biological learning happens continuously [7, 8] makes its underlying mechanisms a valuable reference for AI models. State-of-the-art neuroimaging techniques (e.g., two-photon imaging) now enable real-time observation of dendritic spine shape and function during learning, at the level of single synapses [9]. This methodology can be applied to investigate neocortical plasticity during continuous learning [10].

2.1. Hebbian synaptic plasticity

Biological neural networks exhibit evidence of specialized paradigms that depend on the modulation of synaptic plasticity in response to dynamic input. These paradigms prevent task interference during the process of learning new tasks. These mechanisms consist of a persistent regulation of synaptic plasticity, or meta-plasticity [11]. This regulation is achieved by reducing the liability of a proportion of strengthened synapses, resulting in an enlargement of dendritic

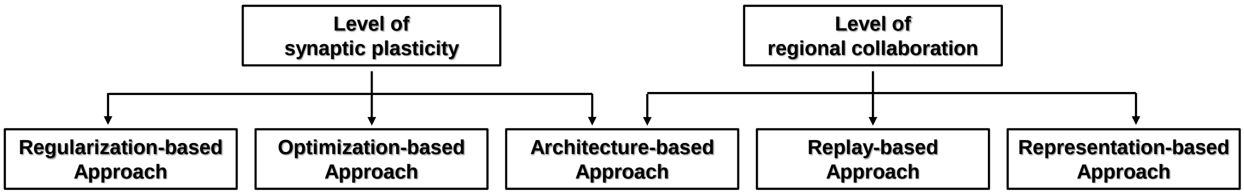

Figure 1: State-of-the-art classification of representative continual learning methods, based on the work of []. A preliminary more general classification based on neuroscience studies has been added.

spines that persists even after the learning of other tasks [12]. Furthermore, this process is followed by a formation of new memory through an expansion or pruning of functional connections [13]. Simultaneously, the activity of some neurons is reduced through inhibitory synapses [14], providing flexibility for the formation of new memories. Empirical findings suggest the concept of controlled forgetting [15], where the aim is to partially forget or weaken the activity corresponding to early knowledge by identifying redundant weights. This is consistent with theoretical and neuromorphic models that suggest memories can be shielded from interference by synapses that transition through a cascade of states having different levels of plasticity [16].

2.2. Regional coordination

Lifelong memory is attributed to the cooperation between cortical and subcortical regions of the brain, according to neuroscience studies. According to one of the most reliable studies, the Complementary Learning System (CLS) theory [17], the advantages of biological learning and memory can be attributed to the complementary functions of the hippocampus and neocortex. The hippocampus can rapidly adapt to incoming data by extracting neural representations of temporally compressed previous experiences []. On the other hand, cortical neurons extract structure from input representations in an unsupervised fashion [18], with the purpose of reconstructing encoding for older stimuli. Modularity is seen as central, beginning with the biological brain, where some regions possess a modular architecture similar to that of the mixture of experts [19]; in neuromorphic systems, this is recreated and corresponds to the use of modular networks in continual learning [20].

3. General categorization

In this section, we consider the elaborated taxonomy of representative continual learning methods presented by [5], and we further categorise them according to the neurological basis presented in the previous section. Indeed, we first distinguish between the level of synaptic plasticity and the level of regional collaboration. The former comprises Regularization-based approaches and Optimization-based approaches, while the latter comprises Replay-based approaches and Representation-based approaches. Both levels share another category of continual learning methods, named Architecture-based approaches (see Fig. 1).

NCL paradigms find a better categorisation by looking at the first division into branches carried out on a neurological basis, since these mechanisms are addressed to the so-called Spiking Neural Networks (SNNs) [21], a class of neural networks closer to biological neural networks than to ANNs. All the other methods of classical continual learning find their natural classification looking at the second division into classes proposed directly by [5]. These methods addressed to ANNs maintain a surprising parallelism with paradigms of lifelong learning in the brain.

3.1. Level of synaptic plasticity

The neuroscientific insight behind this class consists of the dichotomy Hebbian plasticity - stability balance due to the dual requirement of adapting to new data while retaining important information. Specifically, we classify here methods that are characterized by adding explicit regularization terms to balance the old and new tasks, which usually requires storing a frozen copy of the old model for reference (regularization-based), strategies that explicitly design and manipulate the optimization programs (optimization-based), and strategies that construct task-specific parameters with an isolated parameter subspace dedicated to each task in the whole network (architecture-based).

When the target of the regularization is the variation of network parameters, we are dealing with weight regularization. A typical implementation is to add a quadratic penalty in loss function that penalizes the variation of each parameter depending on its contribution or "importance" to performing the old tasks. The importance can be calculated by the Fisher information matrix (FIM), such as elastic weight consolidation (EWC) [22], or by different measures. The Synaptic Intelligence (SI) [23] online assigns a unique memory to each parameter in the neural network and estimates the importance of each parameter by its contribution to the total loss variance. Memory Aware Synapses (MAS) [24] uses a memory matrix to store important synaptic weights learned during training on previous task, updating the memory matrix to retain knowledge of old tasks and protect it from being overwritten. Riemannian Walk (RWalk) [25] combines the regularization terms of SI and EWC to integrate their advantages, considering the model's weights as points on a Riemannian manifold and performing random walks on the manifold to find new weight configurations that are compatible with both the current task and the previous task during the learning process. Instead of dealing with weight regularization through constraints on the old model, an implementation that provides a better stability-plasticity trade-off is the expansion-renormalization process, where the new task solution is obtained separately and renormalised with the old model. Active forgetting of negative transfer in continual learning (AFEC) [26] actively identify and forget information that is likely to cause negative transfer when learning new tasks, which happens whenever there exist a high similarity between the current task and the previous tasks, proceeding with the average of the old and new task solutions through a linear connector [27].

An innovative idea to manipulate the optimization programs is meta-learning, or learning-to-learn for continual learning, of sequentially arrived tasks within the inner loop, which attempts to obtain a data-driven inductive bias for different scenarios, rather than designing it manually. Attentive Independent Mechanisms (AIM) [28] introduces an attention mechanism that selectively focuses on different parts of the network's hidden representations for different tasks, using task-specific independent mechanisms (a mixture of experts) that are learned and updated for each new task.

A third strategy is to construct a well-designed architecture that assigns dedicated parameters to each task. Uncertainty-based Continual Learning (UCL) [29], Continual Learning via Neural Pruning (CLNP) [30] and Neuron-inspired Stability-plasticity Adaptation (NISPA) [31] explicitly identify the important neurons or parameters for the current task and then release the unimportant parts to subsequent tasks, which can be achieved by uncertainty estimation, iterative pruning or activation value. Sparsity restrictions on parameter usage are often needed to prevent network capacity saturation.

3.2. Level of regional collaboration

The neuroscientific insight behind this class consists of the Complementary Learning System (CLS) Theory, explained in Section 2.2. Specifically, we classify here methods for approximating and recovering old data distributions (replay-based), strategies that create and exploit the strengths of representations for continual learning (representation -based), and strategies constructing task-specific parameters through modular networks, without pre-defined task-sharing or task-specific components (architecture-based).

A first implementation of mechanisms that rely on the cooperation of multiple networks to address catastrophic forgetting is to store previously seen data or tasks in a buffer for replay during the learning process to prevent the model from forgetting old knowledge while learning new tasks. We refer to this mechanism as experience replay [5]. It is clear that an adequate use of the memory buffer is required to recover the past information. Gradient Episodic Memory (GEM) [32] introduces a constraint on the gradient during learning so that gradient updates for the current task do not negatively affect performance on the stored experience. Averaged GEM (AGEM) [33] introduces an additional gradient averaging step that not only constrains the gradient of the current task, but also computes an average gradient across multiple tasks in the memory buffer. It then imposes an orthogonality constraint between the gradient of the current task and the averaged gradient, to prevent the gradient updates for the current task from being aligned with the gradients of previous tasks, thus avoiding interference.

Other mechanisms focus on preserving the model's internal knowledge representations while adapting to new data during learning, using self-supervised learning and pre-training. Directly inspired by hippocampus-neocortex cooperation, DualNet [34] is a framework that includes a fast learning system for supervised learning of pattern-separated representations from specific tasks, and a slow learning system for unsupervised representation learning of task-agnostic general representations via a self-supervised learning technique. The two fast and slow learning systems are complementary and smoothly integrated into a holistic continual learning framework.

A third strategy is the use of modular networks, i.e. sub-networks or sub-modules that learn incremental tasks in a differentiated way. ModelZoo [35] grows an ensemble of small models, each trained during an episode of continual learning, effectively demonstrating how a continual learner could benefit from dividing its learning capacity across sets of synergistic tasks.

4. SOTA methods

In this section, we provide a more in-depth explanation of some state-of-the-art methods already introduced in Section 3. In particular, we focus on both implementation and training details that allow to reduce catastrophic forgetting in both SNNs and ANNs, recalling the same biological mechanisms for continual learning.

Given their properties, neuromorphic networks have the advantage of relying directly on brain-inspired concepts such as spike-timing-dependent plasticity (STDP). This mechanism has been exploited by [36], who emulate a complementary learning system through a hybrid supervised-unsupervised network. Their proposed network consists of three main blocks, namely a supervised convolutional network, an equalization block, and an unsupervised STDP network as the final classifier. The first block returns feature maps that are unique for each class of the dataset, but characterized by a different number of firing neurons. By defining for each feature map its own pattern density P as the number of firing neurons divided by the total number of neurons, through the second block of the network, the authors equalize P between the different feature maps so as to impose a fair competition between the different patterns in the STDP network, which operates as a winner-take-all (WTA) network (each postsynaptic neuron of the STDP network compares the incoming current from the presynaptic neurons with an internal threshold). The network was trained to learn 28x28 input patterns from the MNIST dataset, considering two different cases, full training and the case where three classes were not presented during supervised training. The experiment performed consisted of presenting two patterns consecutively for 500 ms, followed by presenting a third untrained pattern for 500 ms. The first two patterns are easily learned in the first phase, thanks to the specialization of the post-synaptic neurons, while the later presentation of the additional pattern leads to unsupervised learning in the synapses of one (out of three) post-synaptic neurons, without affecting the previously trained synapses. Although the average accuracy drops from 98% for full training to 93% for continuous learning, the network overcomes catastrophic forgetting, which is unavoidable in a fully connected network. However, the implementation uses previously known labels to train the feature identifier, which is a limiting option for continuous learning agents as they may encounter input stimuli with features not previously seen.

A continual learning setting, where all network weights are shared and no task information is available, has been studied by [37], who worked on the firing threshold mechanism of SNNs as a gate for the activity of the network. The authors developed a neuromodulated SNN (Nm-SNN) consisting of two separate networks, a convolutional SNN that learns the classification task and the neuromodulator network (NmN) responsible for adapting the spiking threshold of the last fully connected layer of the SNN. During each task of the continual learning (CL) scenario, the trainable parameters of the SNN are optimized using surrogate gradient learning (SGL) [38], a technique that allows deep SNNs to be trained in a supervised manner using backpropagation and gradient descent as in ANNs. Meanwhile, the parameters of the SNNs are frozen and optimized at the end of the CL scenario using an evolutionary strategy, where the selection criterion is the performance in the continual learning scenario. The classes used to train the SNN (convolutional block plus subsequent FC layer) are not used in the evolutionary optimization nor in the testing phase, leading to results that are generalizable to new tasks not seen during pre-training and evolutionary optimization. However, the implementation is slow to converge due to the evolutionary training.

Speaking of runtime efficiency, several algorithms allow higher performance for ANNs than for SNNs, at the cost of lower performance in mitigating catastrophic forgetting. Already in Section 3 we referred to EWC [22]. This algorithm consists of a soft quadratic constraint that slows down learning for certain weights based on how important they are to previously learned tasks, allowing a runtime that is linear in both the number of parameters and the number of training examples. Assuming that tasks have the same structure, networks trained with EWC reuse common components of the network, demonstrating that the EWC algorithm can be effectively combined with deep neural networks to support continual learning in several challenging scenarios, such as supervised learning and reinforcement learning problems. In addition, the EWC algorithm can be based on Bayesian approaches to learning. Formally, when there is a new task to learn, the network parameters are tempered by a prior, which is the posterior distribution on the parameters given data from the previous task(s). This allows fast learning rates for parameters that are poorly constrained by the previous tasks, and slow learning rates for those that are critical.

Almost the same computational and memory efficiency as EWC is achieved by Averaged GEM (AGEM) [33], showing the best trade-off between average accuracy at the end of the learning experience and computational/memory cost. Compared to the original GEM algorithm, AGEM is about 100 times faster and requires 10 times less memory. Whereas GEM ensures that at each training step the loss of each individual previous task, approximated by the samples in episodic memory, does not increase, A-GEM tries to ensure that at each training step the average episodic memory loss over the previous tasks does not increase. However, the authors point out that there is still a significant performance gap between efficient continuous learning methods, including A-GEM, trained in a sequential learning setting, and the same ANN trained in a non-sequential multitasking setting, despite seeing the same data samples.

Modular networks seem to address this problem. The method, called Model Zoo [35], grows an ensemble of small models, each of which is trained during an episode of continual learning. The authors demonstrate, using statistical learning theory and experimental research, how numerous tasks can interact in non-trivial ways when a single model is trained on them. Model Zoo outperforms both existing methods and an isolated continual learner, demonstrating the usefulness of splitting the learner's capacity across multiple tasks. It is also possible to compare Model Zoo and the isolated learner using a single epoch setting. Larger forms of Isolated and Model Zoo do not do well here since training contemporary deep networks requires more than one epoch. However, Model Zoo and variants show less forgetting, which is essentially zero. This suggests that even though existing methods are designed to avoid forgetting (the single epoch setting directly supports this), e.g. A-GEM or EWC, they do forget. Model Zoo's capacity division method can help to reduce forgetting. Furthermore, as compared to previous approaches, the simplicity of Model Zoo and its modifications results in substantially lower training periods and similar inference times.

5. Conclusions

In this survey, we discuss the intersections between state-of-the-art methods for ANNs and SNNs, always in the context of the brain counterparts from which they draw inspiration. We describe both their architecture and training procedures and show how their efficiency in continuous learning scenarios is strongly related to their closeness to biological lifelong learning mechanisms. From this perspective, and encouraged by the broad AI community's interest in continuous learning, we believe that neuromorphic systems will continue to evolve due to their lower environmental impact and their potential to produce faster, more energy-efficient NNs that are less susceptible to the fundamental problems affecting standard ANNs.

CRediT authorship contribution statement

S Falciglia: Conceptualization of this study.

References

[1] Z. Chen, B. Liu, Lifelong machine learning, volume 1, Springer, 2018.

[2] M. McCloskey, N. J. Cohen, Catastrophic interference in connectionist networks: The sequential learning problem, in: Psychology of learning and motivation, volume 24, Elsevier, 1989, pp. 109–165.

[3] T. Luo, W.-F. Wong, R. S. M. Goh, A. T. Do, Z. Chen, H. Li, W. Jiang, W. Yau, Achieving green ai with energy-efficient deep learning using neuromorphic computing, Communications of the ACM 66 (2023) 52–57.

[4] R. Mishra, M. Suri, A survey and perspective on neuromorphic continual learning systems, Frontiers in Neuroscience 17 (2023) 1149410.

[5] L. Wang, X. Zhang, H. Su, J. Zhu, A comprehensive survey of continual learning: Theory, method and application, arXiv preprint arXiv:2302.00487 (2023).

[6] D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick, Neuroscience-inspired artificial intelligence, Neuron 95 (2017) 245–258.

[7] R. Hadsell, D. Rao, A. A. Rusu, R. Pascanu, Embracing change: Continual learning in deep neural networks, Trends in cognitive sciences 24 (2020) 1028–1040.

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

[9] J. Nishiyama, R. Yasuda, Biochemical computation for spine structural plasticity, Neuron 87 (2015) 63–75.

[10] J. Cichon, W.-B. Gan, Branch-specific dendritic ca2+ spikes cause persistent synaptic plasticity, Nature 520 (2015) 180–185.

[11] W. C. Abraham, Metaplasticity: tuning synapses and networks for plasticity, Nature Reviews Neuroscience 9 (2008) 387–387.

[12] G. Yang, F. Pan, W.-B. Gan, Stably maintained dendritic spines are associated with lifelong memories, Nature 462 (2009) 920–924.

[13] C. H. Bailey, E. R. Kandel, K. M. Harris, Structural components of synaptic plasticity and memory consolidation, Cold Spring Harbor perspectives in biology 7 (2015) a021758.

[14] A. C. Arevian, V. Kapoor, N. N. Urban, Activity-dependent gating of lateral inhibition in the mouse olfactory bulb, Nature neuroscience 11 (2008) 80–87.

[15] H. Liu, T. Wu, X. G. Canales, M. Wu, M.-K. Choi, F. Duan, J. A. Calarco, Y. Zhang, Forgetting generates a novel state that is reactivatable, Science advances 8 (2022) eabi9071.

[16] S. Fusi, P. J. Drew, L. F. Abbott, Cascade models of synaptically stored memories, Neuron 45 (2005) 599–611.

[17] J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory., Psychological review 102 (1995) 419.

[18] M. F. Carr, S. P. Jadhav, L. M. Frank, Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval, Nature neuroscience 14 (2011) 147–153.

[19] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton, Adaptive mixtures of local experts, Neural computation 3 (1991) 79–87.

[20] D. Hu, S. Yan, Q. Lu, L. Hong, H. Hu, Y. Zhang, Z. Li, X. Wang, J. Feng, How well does self-supervised pre-training perform with streaming data?, arXiv preprint arXiv:2104.12081 (2021).

[21] S. Ghosh-Dastidar, H. Adeli, Spiking neural networks, International journal of neural systems 19 (2009) 295–308.

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

[23] F. Zenke, B. Poole, S. Ganguli, Continual learning through synaptic intelligence, in: International conference on machine learning, PMLR, 2017, pp. 3987–3995.

[24] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, T. Tuytelaars, Memory aware synapses: Learning what (not) to forget, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 139–154.

[25] A. Chaudhry, P. K. Dokania, T. Ajanthan, P. H. Torr, Riemannian walk for incremental learning: Understanding forgetting and intransigence, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 532–547.

[26] L. Wang, M. Zhang, Z. Jia, Q. Li, C. Bao, K. Ma, J. Zhu, Y. Zhong, Afec: Active forgetting of negative transfer in continual learning, Advances in Neural Information Processing Systems 34 (2021) 22379–22391.

[27] G. Lin, H. Chu, H. Lai, Towards better plasticity-stability trade-off in incremental learning: A simple linear connector, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 89–98.

[28] E. Lee, C.-H. Huang, C.-Y. Lee, Few-shot and continual learning with attentive independent mechanisms, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 9455–9464.

[29] H. Ahn, S. Cha, D. Lee, T. Moon, Uncertainty-based continual learning with adaptive regularization, Advances in neural information processing systems 32 (2019).

[30] S. Golkar, M. Kagan, K. Cho, Continual learning via neural pruning, arXiv preprint arXiv:1903.04476 (2019).

[31] M. B. Gurbuz, C. Dovrolis, Nispa: Neuro-inspired stability-plasticity adaptation for continual learning in sparse networks, arXiv preprint arXiv:2206.09117 (2022).

[32] D. Lopez-Paz, M. Ranzato, Gradient episodic memory for continual learning, Advances in neural information processing systems 30 (2017).

[33] A. Chaudhry, M. Ranzato, M. Rohrbach, M. Elhoseiny, Efficient lifelong learning with a-gem, arXiv preprint arXiv:1812.00420 (2018).

[34] Q. Pham, C. Liu, S. Hoi, Dualnet: Continual learning, fast and slow, Advances in Neural Information Processing Systems 34 (2021) 16131–16144.

[35] R. Ramesh, P. Chaudhari, Model zoo: A growing" brain" that learns continually, arXiv preprint arXiv:2106.03027 (2021).

[36] I. Muñoz-Martín, S. Bianchi, G. Pedretti, O. Melnic, S. Ambrogio, D. Ielmini, Unsupervised learning to overcome catastrophic forgetting in neural networks, IEEE Journal on Exploratory Solid-State Computational Devices and Circuits 5 (2019) 58–66.

[37] I. Hammouamri, T. Masquelier, D. Wilson, Mitigating catastrophic forgetting in spiking neural networks through threshold modulation, Transactions on Machine Learning Research (2022).

[38] E. O. Neftci, H. Mostafa, F. Zenke, Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks, IEEE Signal Processing Magazine 36 (2019) 51–63.