Cannistraci-Hebb Training of Convolutional Neural Networks

Hanming Li[1,2] LIHANMING2025@GMAIL.COM
Yusong Wang[1] YUSONG@U.NUS.EDU
Yingtao Zhang[1,2] ZHANGYINGTAO1024@GMAIL.COM
Carlo Vittorio Cannistraci[1,2,3] * KALOKAGATHOS.AGON@GMAIL.COM
[1] *Center for Complex Network Intelligence (CCNI), Tsinghua Laboratory of Brain and Intelligence (THBI), Department of Psychological and Cognitive Sciences*
[2] *Department of Computer Science,* [3] *School of Biomedical Engineering*
Tsinghua University, China

Editors: List of editors' names

Abstract

Dynamic sparse training enables neural networks to evolve their topology during training, which reduces computational overhead while maintaining performance. Cannistraci-Hebb Training (CHT), a brain-inspired method based on epitopological learning principles, has demonstrated significant advantages in building ultra-sparse fully connected networks. However, its application to convolutional neural networks (CNNs) faces challenges due to two fundamental constraints inherent in CNNs: receptive field locality and weight-sharing dependency. These constraints prevent the independent link manipulation that is essential for existing CHT frameworks.

We propose CHT-Conv, extending CHT to convolutional layers while adhering to the inherent constraints of convolutional layers. Experiments on CIFAR-10 and CIFAR-100 using VGG16 architectures show CHT-Conv achieves competitive or superior performance compared to SET baseline at 50% and 70% sparsity levels. This work represents the first successful extension of epitopological learning principles to convolutional architectures, opening new possibilities for brain-inspired sparse training in modern deep learning.

Keywords: Dynamic Sparse Training, Epitopological Learning, Convolutional Neural Networks, Brain-Inspired Computing

1. Introduction

Dynamic Sparse Training (DST) has emerged as a promising paradigm for training artificial neural networks with sparse connectivity, aiming to reduce computational and memory overhead while maintaining performance. Unlike static pruning methods, DST allows network topology to evolve during training, drawing inspiration from synaptic turnover in biological neural networks (Mocanu et al., 2018; Evci et al., 2020).

Among recent DST advances, Cannistraci-Hebb Training (CHT) has demonstrated remarkable capabilities in training ultra-sparse networks (Zhang et al., 2024). CHT implements epitopological learning—a brain-inspired framework that evolves network topology

* Corresponding author

© 2025 H. Li, Y. Wang, Y. Zhang & C.V. Cannistraci.

through structural reorganization rather than solely weight updates. By using the CH3-L3 link predictor for gradient-free regrowth, CHT enables artificial neural networks with merely 1% connectivity to surpass fully connected counterparts in multiple visual tasks.

However, existing CHT methods are primarily designed for fully connected layers, leaving convolutional neural networks (CNNs) largely unexplored. CNNs present unique challenges for topological evolution due to two fundamental constraints: receptive field locality, which limits the spatial range of possible connections, and weight-sharing dependency, which requires grouped manipulation of connections rather than independent link control.

This paper addresses these challenges by extending CHT to convolutional layers through CHT-Conv, using a three-step evolution method. During the evolution, we expanded the mask repeatedly to construct a global topology, which is subsequently used for link prediction. The average value across different sliding windows is then computed to determine which links should be regrowed.

We evaluate CHT-Conv on CIFAR-10 and CIFAR-100 datasets using VGG16 architectures. Results demonstrate that CHT-Conv achieves competitive or superior performance compared to traditional dynamic sparse training methods like SET in the same sparsity level. This work represents a significant step toward applying brain-inspired topological learning principles to modern deep learning architectures.

2. Related Work

2.1. Dynamic Sparse Training

Dynamic Sparse Training (DST) has emerged as a promising paradigm for training artificial neural networks (ANNs) with sparse connectivity, aiming to reduce computational and memory overhead while maintaining or even enhancing model performance. Unlike static sparse methods such as pruning-at-initialization (Frankle and Carbin, 2018), DST allows the network topology to evolve during training, drawing inspiration from synaptic turnover in biological neural networks (Holtmaat and Svoboda, 2009).

2.2. Epitopological Learning and Cannistraci-Hebb Training

A significant innovation in DST is the introduction of epitopological learning (EL), a brain-inspired framework rooted in network science and complex systems theory (Daminelli et al., 2015; Cannistraci, 2018). EL is derived from a dual interpretation of Hebb's axiom — "neurons that fire together wire together" — and emphasizes epitopological plasticity: the ability of a network to learn through structural reorganization rather than solely through weight updates (Cannistraci et al., 2013). EL can be addressed as a link prediction task using a series of topology-based link predictors such as CH3-L3 (Muscoloni et al., 2022), with which Cannistraci-Hebb Training (CHT) (Zhang et al., 2024) implements EL and demonstrates that artificial neural networks with merely 1% connectivity can surpass fully connected counterparts in multiple visual tasks.

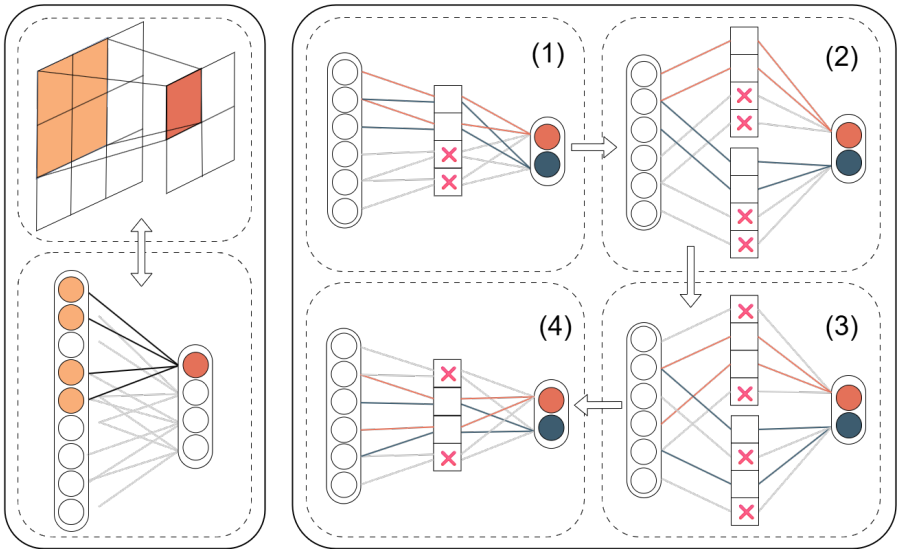

Figure 1: Left: The topology of a convolutional layer can be represented as a bipartite graph. Right: Demonstration of CHT-Conv.

3. CHT-Conv

In this section, we provide a detailed explanation of how the CHT method is extended to convolutional layers within CNN.

3.1. Topology Matrix in Convolutional Layer

To extend the CHT framework to CNNs, the first essential step is to establish an appropriate definition of topology within convolutional layers. A bipartite graph can be constructed between the inputs and outputs of a convolutional layer. As illustrated in the first column in Figure 1, for each kernel, the first partite set consists of activations from the input feature maps, while the second partite set comprises neurons corresponding to each sliding window position in the output.

However, the topological structure in convolutional neural networks is subject to two constraints: **(1) Receptive Field Locality.** Due to the limited receptive field of each sliding window, the range of activations that can connect to a given output neuron is constrained by the network architecture. **(2) Weight-Sharing Dependency.** Owing to the inherent weight-sharing mechanism in convolutional operations, individual connections cannot be independently removed or regrown.

Constraint 2 poses a challenge for implementing Epitopological Learning via CHT, where the link predictor independently computes a score for each connection in the network topology and selects the most promising ones for regrowth.

3.2. Proposed Method

3.2.1. Initialization

For each convolutional layer, we initialize a random boolean tensor mask of the same shape as each kernel to govern which connections are removed at this phase. Note that the same mask is applied to all the sliding windows, thereby preserving the shared-weight nature of convolutional operations.

3.2.2. Evolution

The evolution process in CHT-Conv follow the following steps. **Removal.** On each kernel, the links with the smallest weight magnitudes are removed. **Link Prediction.** We adopts CH link predictors to calculate the scores for all the disabled links. **Averaging.** The average of link prediction scores across different sliding windows is computed to determine which links are ultimately regrown in this round of evolution.

4. Experiments

4.1. Experimental Setup

To evaluate the performance of CHT-Conv, we conducted preliminary experiments on two widely-used benchmark datasets: CIFAR-10 and CIFAR-100 (Krizhevsky et al., 2009), with VGG16 (Simonyan and Zisserman, 2014) as our backbone. For the CHT-Conv method, we evaluated the performance of the CH2-L3n and CH3-L3n predictors. The baselines included the dense model and the sparse model trained under SET (Mocanu et al., 2018). To focus on examining proposed CHT-Conv, when conducting experiments using CHT-Conv or SET, we sparsified only to the convolutional parts while keeping the Multi-Layer Perceptron at the end of the network dense. The detailed experimental setup is described in Appendix A.

Table 1: Performance comparison of Dense, SET and CHT-Conv

Sparsity	**Method**	**CIFAR-10**	**CIFAR-100**
0%	Dense	92.14 ± 0.06	**72.58 ± 0.11**
50%	SET	92.08 ± 0.10	71.86 ± 0.12
	CHT-CH2	92.04 ± 0.21	72.08 ± 0.06
	CHT-CH3	**92.32 ± 0.06**	**72.14 ± 0.06**
70%	SET	91.63 ± 0.07	**71.03 ± 0.13**
	CHT-CH2	91.73 ± 0.15	70.89 ± 0.23
	CHT-CH3	**92.04 ± 0.05**	70.75 ± 0.00

4.2. Results

The accuracy of each experimental group is reported in Table 2, where the error intervals are derived from standard error. Preliminary experimental results indicate that: (1) As

sparsity progressively increases, the performance of the network declines; (2) Compared to SET, the CHT method either outperforms SET (on CIFAR-10) or performs at least comparably to SET (on CIFAR-100).

5. Conclusion

This paper successfully extends Cannistraci-Hebb Training to convolutional neural networks through CHT-Conv, addressing the fundamental challenge of applying epitopological learning principles to weight-sharing architectures. Experimental results on CIFAR datasets indicate that CHT-Conv achieves performance competitive with existing dynamic sparse training methods across multiple sparsity levels. In subsequent work, we plan to further refine the performance of CHT-Conv and conduct more extensive experimental evaluations to strengthen its robustness and general applicability.

References

Carlo Vittorio Cannistraci. Modelling self-organization in complex networks via a brain-inspired network automata theory improves link reliability in protein interactomes. *Scientific reports*, 8(1):15760, 2018.

Carlo Vittorio Cannistraci, Gregorio Alanis-Lobato, and Timothy Ravasi. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Scientific reports*, 3(1):1613, 2013.

Simone Daminelli, Josephine Maria Thomas, Claudio Durán, and Carlo Vittorio Cannistraci. Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks. *New Journal of Physics*, 17(11):113037, 2015.

Utku Evci, Trevor Gale, Jacob Menick, Pablo Samuel Castro, and Erich Elsen. Rigging the lottery: Making all tickets winners. In *International conference on machine learning*, pages 2943–2952. PMLR, 2020.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.

Anthony Holtmaat and Karel Svoboda. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience*, 10(9):647–658, 2009.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Decebal Constantin Mocanu, Elena Mocanu, Peter Stone, Phuong H Nguyen, Madeleine Gibescu, and Antonio Liotta. Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nature communications*, 9(1):2383, 2018.

Alessandro Muscoloni, Umberto Michieli, Yingtao Zhang, and Carlo Vittorio Cannistraci. Adaptive network automata modelling of complex networks. 2022.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Yingtao Zhang, Jialin Zhao, Wenjing Wu, and Alessandro Muscoloni. Epitopological learning and cannistraci-hebb network shape intelligence brain-inspired theory for ultra-sparse advantage in deep learning. In *The Twelfth International Conference on Learning Representations*, 2024.

Appendix A. Experimental Setup

Data Augmentation. For both CIFAR10 and CIFAR100, we applied random horizontal flip (p=0.5) on training set, and normalization with the mean and std on valid and test sets.

Reproducibility. For all experiments, we used the random seeds of 14, 15, and 16. We also turned on all the deterministic flags in Python, Pytorch and NumPy.

Training Hyperparameters.

Table 2: Hyperparatmeters for Training VGG16 on CIFAR10 and CIFAR100

Dataset	**Optimizer**	**Learning Rate**	**Batch Size**	**#Epochs**	**Scheduler**
CIFAR10	Adam	0.01	128	100	Linear-Linear
CIFAR100	SGD	0.1	128	240	Linear-CosAnnealing