IMPROVING THE SPARSE STRUCTURE LEARNING OF SPIKING NEURAL NETWORKS FROM THE VIEW OF COMPRESSION EFFICIENCY

Jiangrong Shen [1,2,3,4], Qi Xu[5], Gang Pan[3], Badong Chen[2,4]

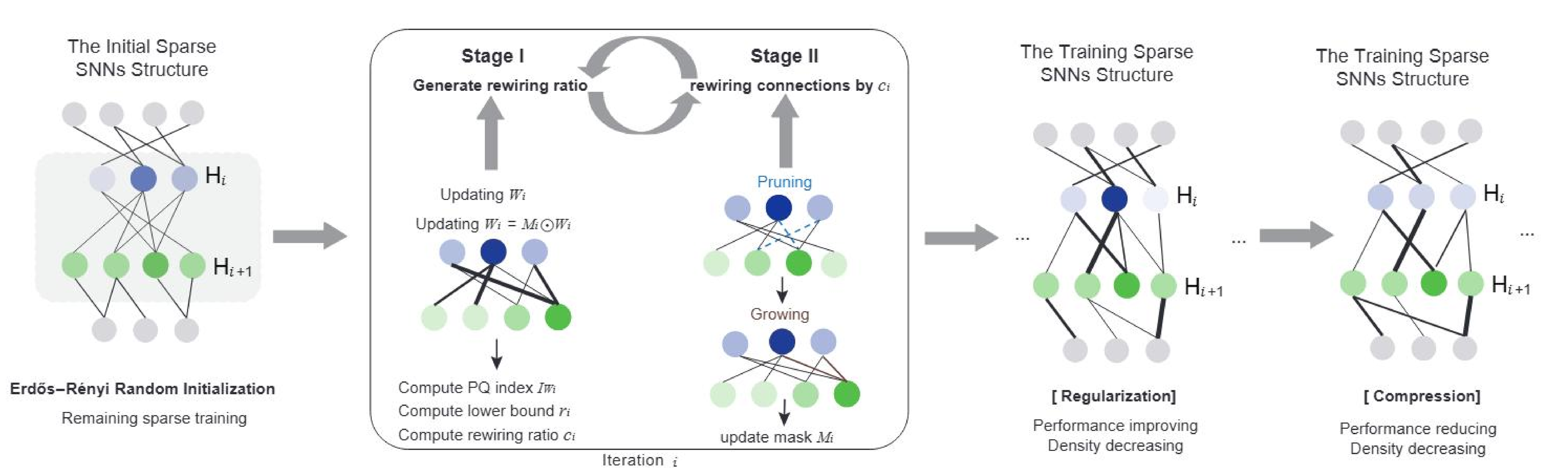
1 Faculty of Electronic and Information Engineering,Xi'an Jiaotong University
2 Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
3 State Key Lab of Brain-Machine Intelligence, Zhejiang University, China
4 National Key Lab of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University
5 School of Computer Science, Dalian University of Technology

Motivation

Motivation:

- Spiking neural networks (SNNs) have exhibited more complex spatiotemporal dynamics in comparison to ANNs, and have the potential to implement the next generation of machine intelligence with low power consumption by combining with neuromorphic hardware. → In this paper, we explore how to implement adaptive pruning of **SNNs from the View of Compression Efficiency.**
- In biological systems, rewiring process forms flexible network structure and promotes synaptic sparsity contributing to the brain's low power consumption. Therefore, emulating the brain's structural synaptic plasticity through a dynamic structure learning approach could be key to developing more flexible deep SNN models. → Therefore, in this paper, we explore **optimizing the pruning ratio from a neural network compression perspective** as explored in machine learning, where data compression theory helps quantify the compressibility of a sub-network during each connection updating iteration, thereby avoiding under or over-prunin

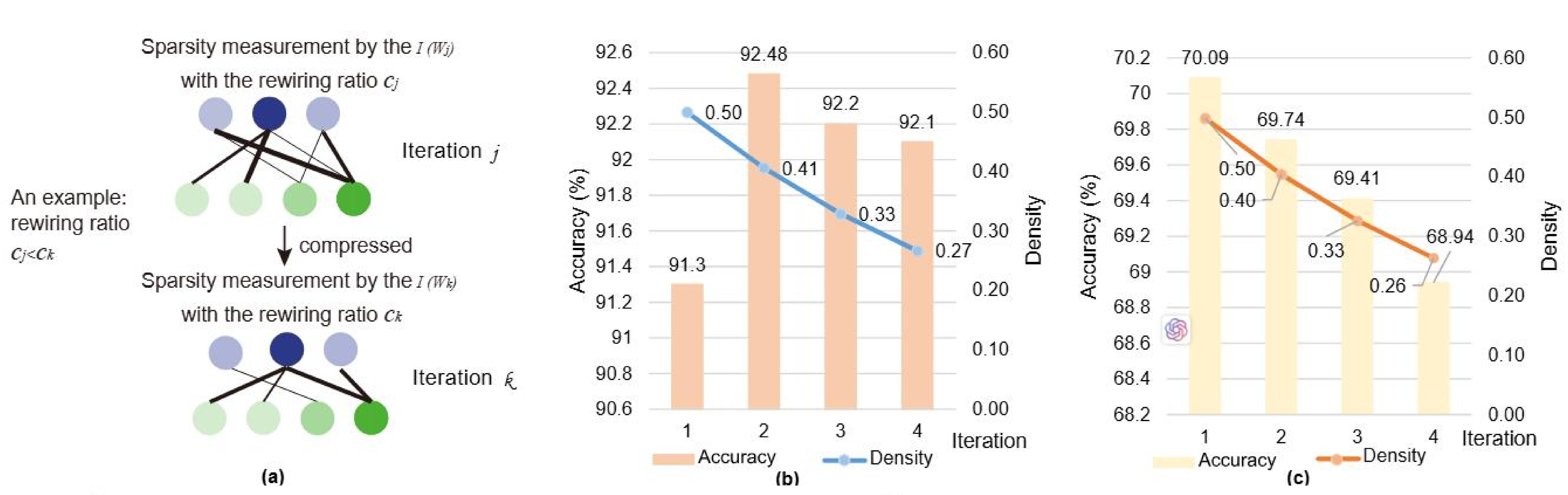
Method



◆ We introduce a pioneering two-stage dynamic structure learning framework for deep SNNs that utilizes the PQ index to gauge and dynamically adjust structure of sparse subnetworks according to compressibility. This novel approach tailors the rewiring ratio throughout the training process, providing a fine-tuned, adaptive mechanism that enhances the foundational training dynamics of deep sparse SNNs.

◆ Our methodology extends traditional sparse training approaches for SNNs by implementing a continuous, iterative learning process across two stages. In the first stage, the PQ index informs the adjustment of synaptic connection rewiring ratios. In the second stage, these ratios guide a dynamic rewiring strategy that includes both the pruning and regrowth of connections. Thus the methodology optimizes the SNNs' structural efficiency and operational effectiveness far beyond conventional static pruning techniques.

Results



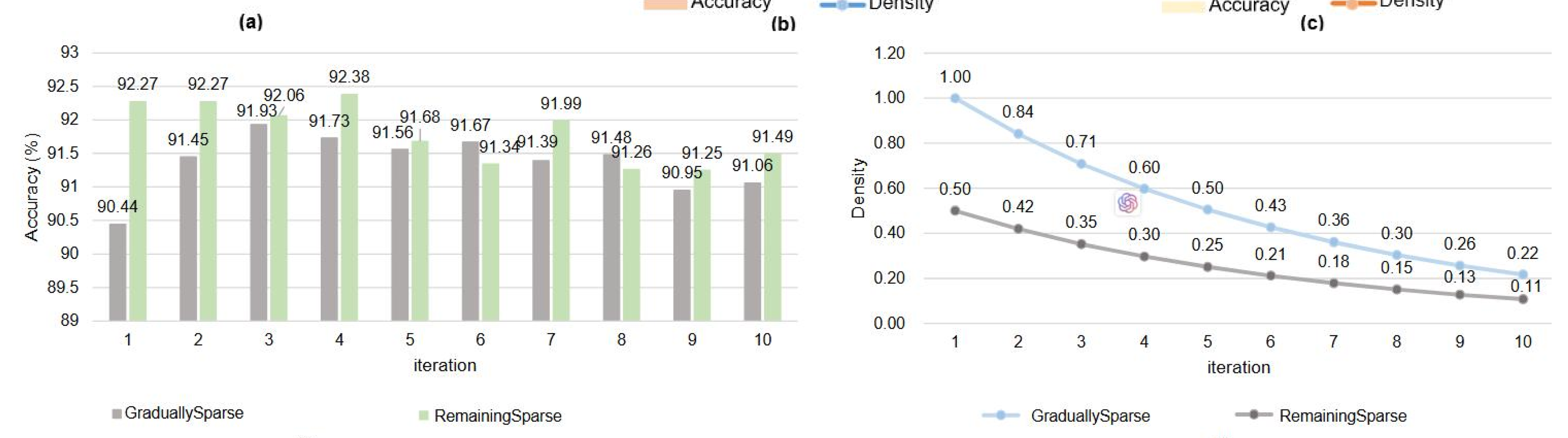


Table 1: Performance comparison of the proposed two-stage sparse structure learning approach for SNNs with other models.

Dataset	Pruning Method	Architecture	T	Top-1 Acc.(%)	Acc. Loss(%)	Conn. (%)	Param. (M)	SOPS (M)
CIFAR10	ADMM	7 Conv, 2 FC	8	90.19	-0.13	25.03	15.54	-
	Grad R	6 Conv, 2 FC	8	92.54	-0.30	36.72	10.43	-
	ESLSNN	ResNet19	2	91.09	-1.7	50	6.3	180.56
	STDS	6 Conv, 2 FC	8	92.49	-0.35	11.33	1.71	147.22
	UPR	6 Conv, 2 FC	8	92.05	-0.79	1.16	9.56	16.47
	This work	**ResNet19 Neuron-wise**	**2**	**92.48** **92.1**	**+1.18** **+0.8**	**40.58** **26.63**	**5.12** **3.36**	**158.35** **121.49**
		ResNet19 Layer-wise	**2**	**92.38** **91.99**	**+0.11** **-0.28**	**29.72** **17.91**	**3.7** **2.26**	**133.26** **110.65**
CIFAR100	ESLSNN	ResNet19	2	73.48	-0.99	50	6.32	186.25
	UPR	SEW ResNet18	4	70.45 69.41	-3.71 -4.75	3.60 2.48	-	9.60 6.79
	This work	**ResNet19 Layer-wise**	2	**70.3**	**+1.07**	**29.48**	**3.73**	**140.27**
DVS-CIFAR10	ESLSNN	VGGSNN	10	78.3	-0.28	10	0.92	129.64
	STDS	VGGSNN	10	79.8	-2.6	4.67	0.24	38.85
	UPR	VGGSNN	10	78.3 81.0	-0.5 -1.4	0.77 4.46	1.81 2.5	6.75 31.86
	This work	**VGGSNN Layer-wise**	**10**	**78.4**	**+0.08**	**30**	**2.76**	**189.02**

Conclusion

- To summarize, this study has introduced a novel two-stage dynamic structure learning method tailored for SNNs that effectively addresses the challenges of fixed pruning ratios and the limitations of static sparse training methods prevalent in current models.
- The experimental results validate that the proposed dynamic structure learning greatly improves the compression efficiency of SNNs. Additionally, it either matches or exceeds the performance benchmarks set by current models in certain circumstances.

For further information please contact with Jiangrong Shen: jrshen@zju.edu.cn