Enhancing Learning in Sparse Neural Networks: A Hebbian Learning Approach

Alexander de Ranitz

Supervised by Ardion Beldad & Elena Mocanu

University of Twente.

alexanderderanitz@gmail.com

github.com/alexander-de-ranitz/HebbSET

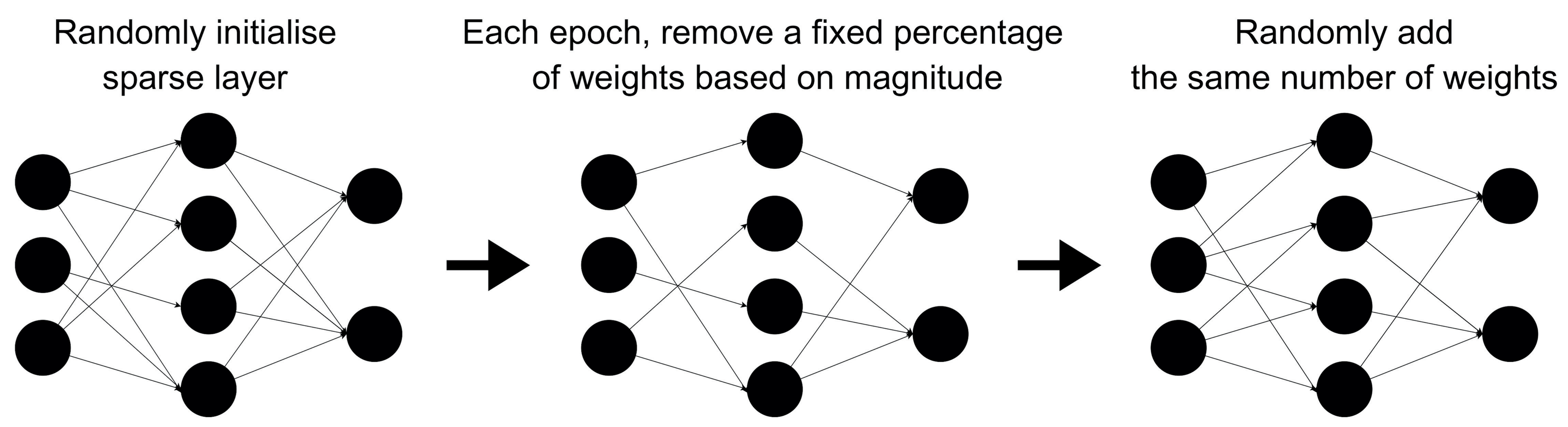
1 Introduction

Although artificial neural networks (ANNs) are loosely based on the human brain, there are many differences between ANNs and biological neural networks. One such difference is the way they learn. A famous concept in neuroscience is Hebb's postulate, often summarised as: ***"Neurons that fire together, wire together"*** [1]. While this concept is fundamental to our understanding of human learning, modern ANNs that use gradient descent methods do not (explicitly) incorporate this idea. Here, it is investigated whether Hebbian learning can be used to enhance learning in truly sparse neural networks trained using Sparse Evolutionary Training (SET) [2]. The resulting novel algorithm is named **HebbSET**.

2 Sparse Neural Networks

Artificial neural networks are typically extremely large, requiring considerable time and energy to train. However, research has shown that many of the connections in ANNs can be removed without reducing performance [2][3]. These ***sparse neural networks*** offer a promising method of improving learning efficiency and reducing the computational load of ANNs [4].

Sparse Evolutionary Training



Fully-Connected ANNs

- Number of weights scales quadratically with neurons
- Possibly many redundant or near-zero weights
- Static topology

Sparse Evolutionary Training

- Number of weights scales linearly with neurons
- Massively reduced number of weights (>95%)
- Network topology is optimised during training

3 Hebbian Learning

In his 1949 Book titled *The Organisation of Behavior*, Donald Hebb introduced a theory of how synaptic strengths change in the brain. This rule of thumb can be summarised as follows:

> *"Neurons that fire together, wire together"*
>
> *Donald Hebb*

To apply this idea to help train artificial neural networks, it needs to be extended to include both weight increases and decreases, and it must be appropriately quantified. This yields the following equation:

$$\Delta w_{ji}^{l} = a_i^{l-1} \cdot (a_j^l - \overline{a_j^l})$$

Weight update; Activation of the presynaptic neuron; Activation of the postsynaptic neuron; Average activation of the postsynaptic neuron

The resulting behaviour of this rule can be summarised as:

1. If presynaptic **neuron A is active** and postsynaptic **neuron B is active**, the weight between A and B should **increase**.
2. If presynaptic **neuron A is active** but postsynaptic **neuron B is inactive**, the weight between A and B should **decrease**.
3. If presynaptic **neuron A is inactive**, the weight between A and B **does not change**.

Finally, this Hebbian term is combined with mini-batch gradient descent in order to give the complete learning rule:

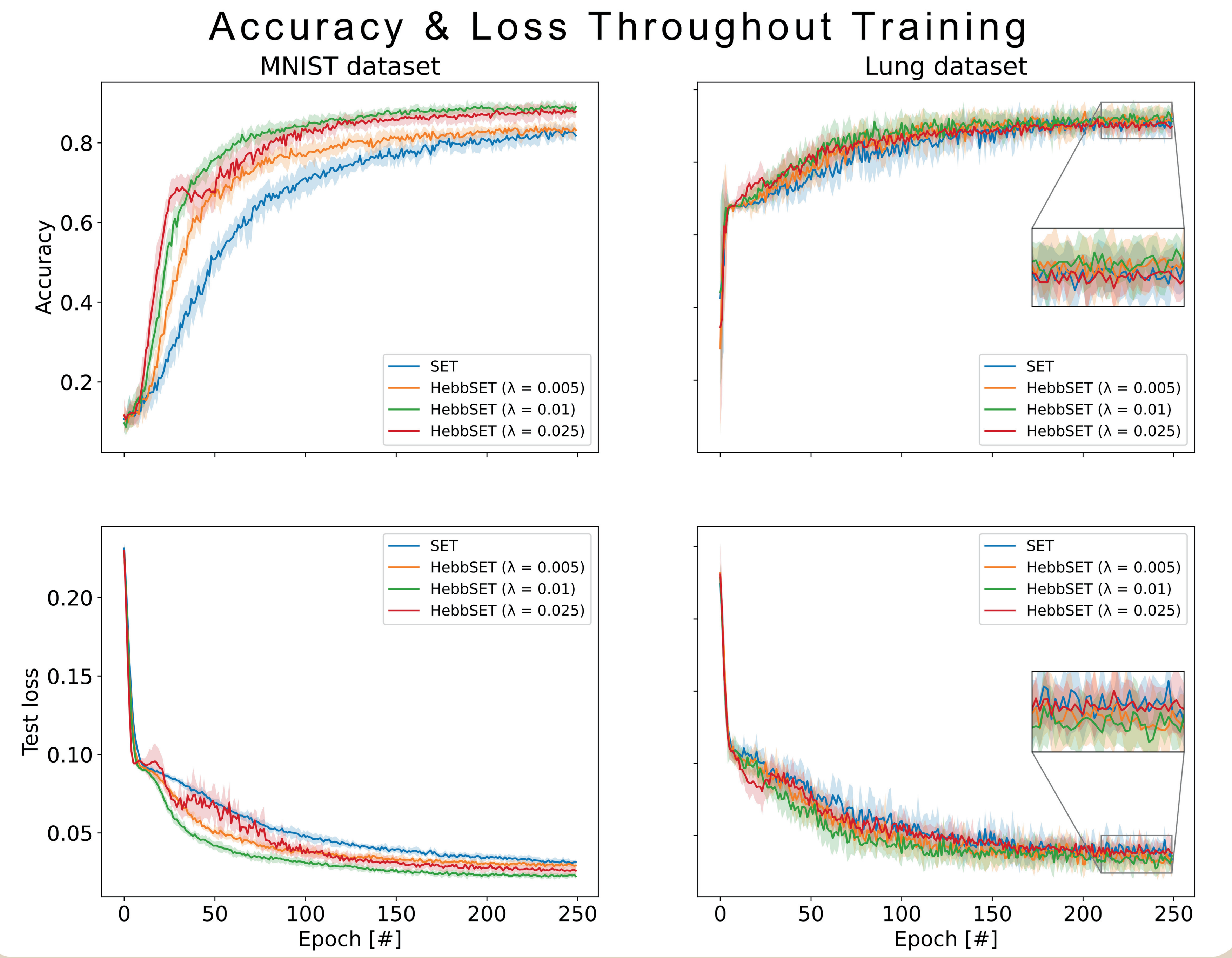
$$\Delta w_{ji}^{l} = -\alpha \cdot \frac{\partial J}{\partial w_{ji}^{l}} + \lambda(t) \cdot a_i^{l-1} \cdot (a_j^l - \overline{a_j^l})$$

Learning rate; Cost derivative; Hebbian factor

4 Results

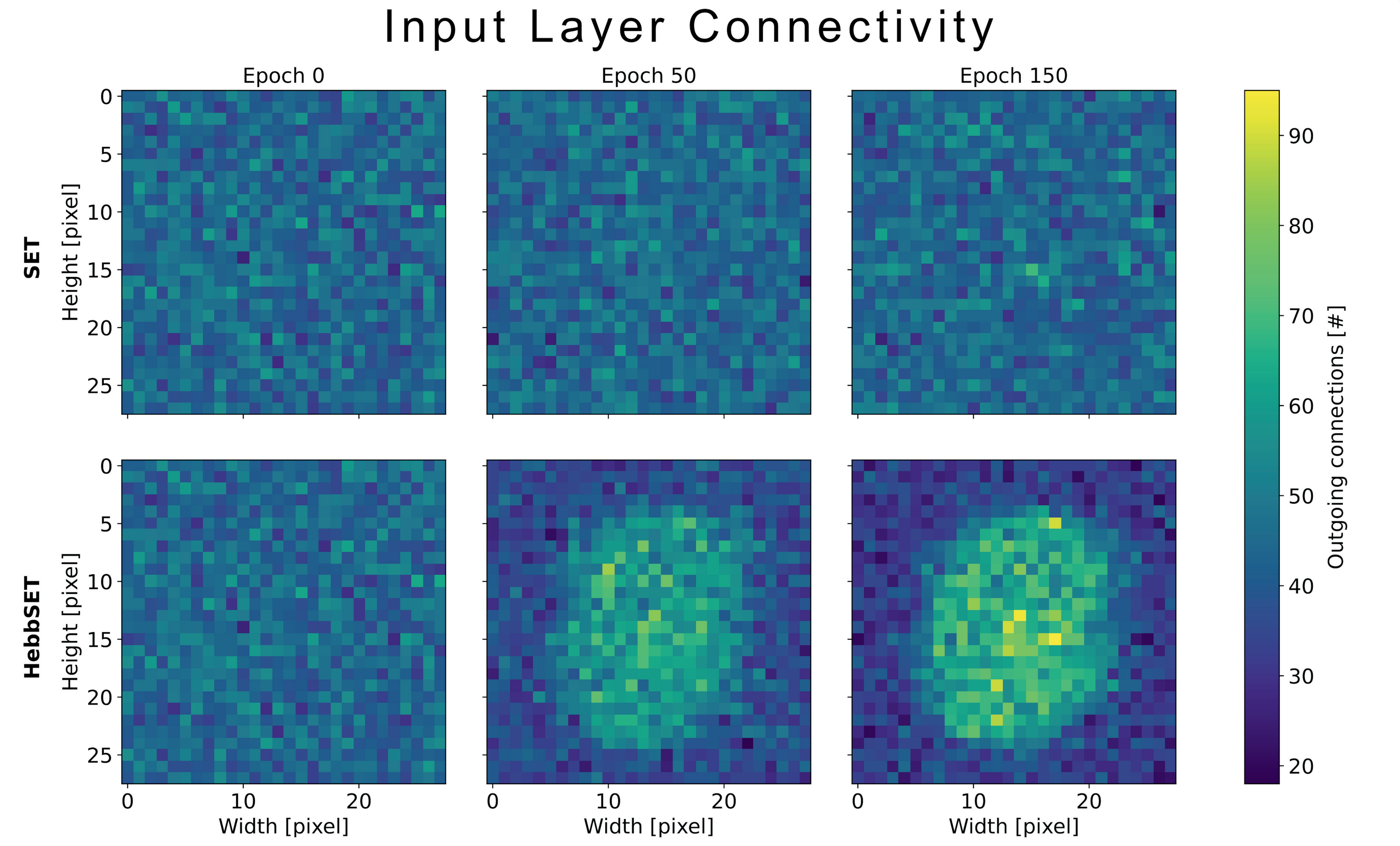
Network Performance

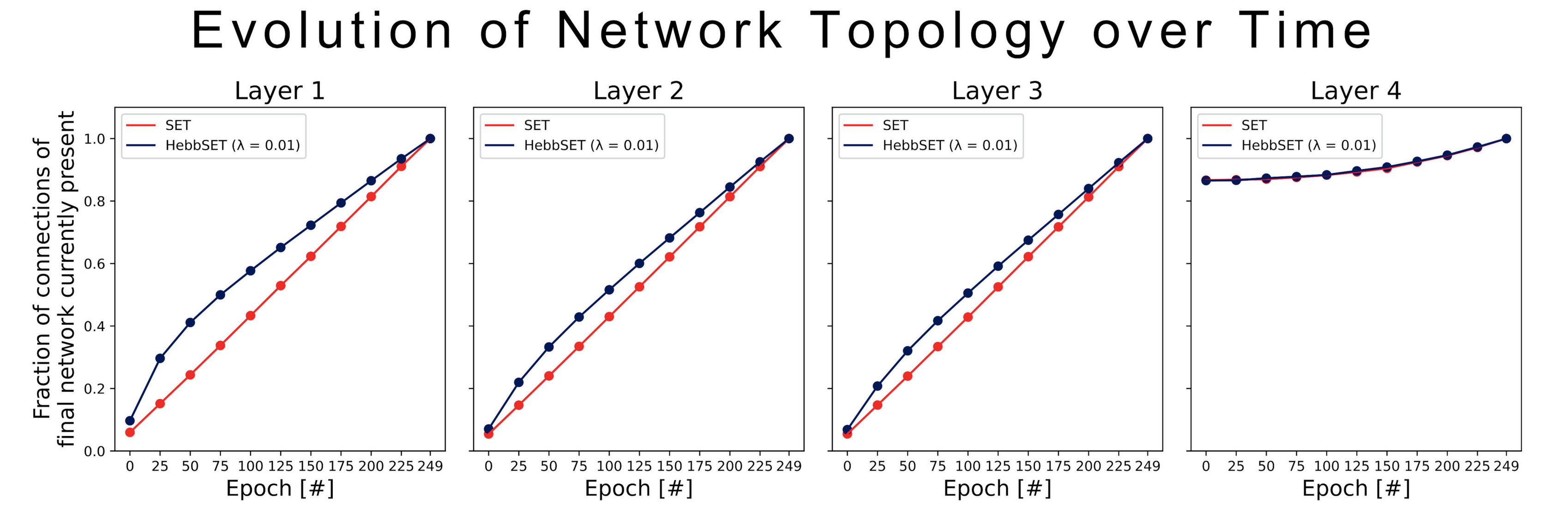
- Increased learning speed
- Improved final accuracy (especially on MNIST dataset)



Network Topology

- Improved adaptation to training data
- Faster convergence to (locally) optimal topology





5 Conclusion

- Incorporating Hebbian learning into the training of sparse neural networks allows the network to learn in a manner more similar to human learning whilst improving performance
- Although more research is necessary to better understand the effectiveness of HebbSET, our results demonstrate how useful it can be to combine findings from Neuroscience with AI

6 References

[1] Hebb, D.O.: The organisation of behaviour: A neuropsychological theory. John Wiley and Sons (1949)

[2] Mocanu, D.C., Mocanu, E., Stone, P., Nguyen, P.H., Gibescu, M., Liotta, A.: Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. Nature Communications (2018)

[3] Frankle, J., Carbin, M.: The lottery ticket hypothesis: Finding sparse, trainable neural networks. ICLR (2019)

[4] Hoefler, T., Alistarh, D., Ben-Nun, T., Dryden, N., Peste, A. Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks, JMLR, (2021)