The role of membrane time constant in the training of spiking neural networks

Improving accuracy by per-neuron learning

Adam Pazderka [1] | Supervisors: N. Tömen[2], A. Micheli[3] | Examiner: E. A. Markatou[4]

[1]A.Pazderka@student.tudelft.nl [2]N.Tomen@tudelft.nl [3]A.Micheli@tudelft.nl [4]E.A.Markatou@tudelft.nl

1. Background

- **Spiking neural networks (SNNs)** aim to utilize mechanisms from biological neurons to bridge the computational and efficiency gaps between the human brain and machine learning systems.
- The term "spiking" stems from the **1-bit mode of communication** between neurons. Instead of having an activation function outputting a real number, spiking neurons can either be at rest or emit a spike. This implies another fundamental feature; SNNs are inherently **temporal**.
- Deep learning commonly features the **Leaky-Integrate-and-Fire (LIF) neuron** model due to its favorable trade-off between biological plausibility and computational efficiency [1, p. 1023].
- The LIF neuron integrates incoming weighted spikes into its exponentially decaying internal state called **membrane potential** and emits a spike if it exceeds a set threshold.
- The decay rate of the membrane potential is indirectly expressed by the **membrane time constant** τ, which is set as a hyperparameter in the LIF model.
- Previous work introduced a **numerically stable way to indirectly optimize** τ using a newly introduced parameter in its **Parametric LIF (PLIF) neuron** model. It used **one** τ **parameter per layer**, which resulted in improved accuracy [2].
- Using one τ per layer is arbitrary and does not stem from any cited research [2, Sec. 3.3].
- Analogy between the role of τ and that of the "forget gate" in the highly successful LSTMs suggests that exploring having τ **per neuron** is worthwhile.
- Previous work achieved a peak **test accuracy of 96.53%** on the DVS128 Gesture dataset. The same experiment is replicated in this work using the **"Baseline" model**. Furthermore, the same DVS128 Gesture dataset is used for all experiments to compare the results.

2. Research Question

What is the effect of having a learnable membrane time constant per neuron on the accuracy of a spiking neural network?

3. Methodology

New log-normal initialization and variance-based regularization methods for τ are introduced. The role of the regularization term is to enable an incremental approach over the baseline model and to prevent potential overfitting. It is formulated as $\lambda * Var(l)$ per layer l.

Four models are trained:

Baseline	uses a single learnable τ per layer, initialized as $\tau_0 = 2$. The purpose is to verify the implementation before experimenting with the new models.
No Regularization	introduces having a τ per neuron. It is expected to suffer from overfitting due to the increased number of parameters.
$\lambda = 0.01$ and $\lambda = 0.1$	The "$\lambda = 0.1$" model serves as a "control" model. It is expected to behave similarly to the baseline. The "$\lambda = 0.01$" is expected to show whether letting individual neurons optimize their τ increases final accuracy while avoiding potential overfitting.

4. Results

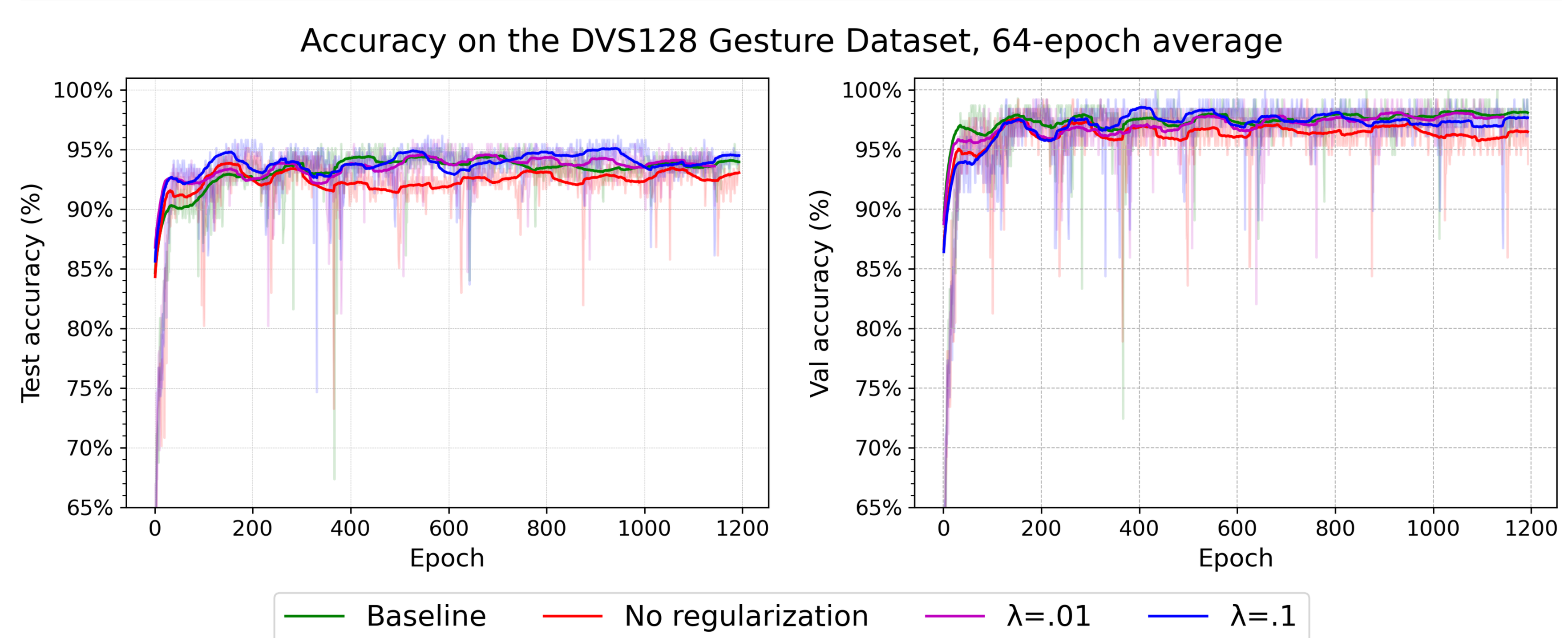

Figure 1. **Test and validation accuracy during training on the DVS128 Gesture dataset visualized as a 64-epoch moving average.** The x-axis denotes the epoch number. The y-axis denotes accuracy.

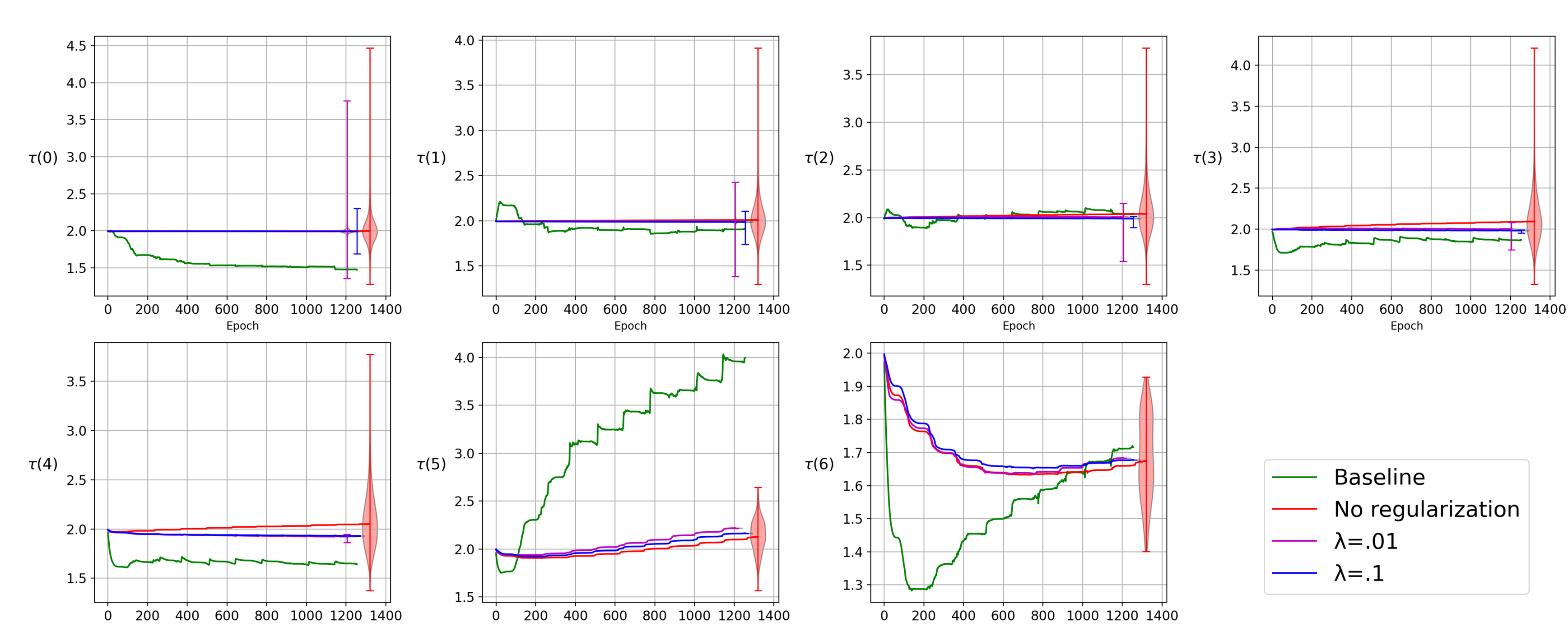

Figure 2. **Evolution of the membrane time constant** τ **in individual layers of PLIF neurons.** The x-axis denotes the epoch number, while the y-axis denotes the value of τ in layer l as $\tau(l)$. The mean value is used for models with multiple τ parameters per layer (all except the baseline), and their final distribution, including extrema, is visualized using a violin plot.

5. Conclusion

- Results suggest that **having a membrane time constant per neuron instead of per layer does not have a significant effect on final accuracy.**
- This conclusion needs to be treated with low confidence, and further experiments across more datasets, ideally with multiple initializations per model, are needed.

Model	Test Accuracy	Val Accuracy
Baseline	95.5%	100%
No Reg.	95.1%	99.2%
$\lambda = .01$	95.8%	99.2%
$\lambda = .1$	96.2%	100%

Table 1: Comparison of the maximum achieved test and validation set accuracies. Each model has been trained once for 1200-1500 epochs using the seed 2020. All the best-performing instances on the test set were encountered in the first 600 epochs.

6. Discussion and Future Work

Two new hypotheses are formed:

- There is an inverse relation between the number of τ parameters and their learning rate
- There is a general relationship between the learning capacity of a spiking layer and the number of learnable τ and input weight parameters

Four areas for future research are identified:

Layers as heterogeneous networks It was assumed that a single layer of neurons forms a population whose τ parameters should be regularized to stay together. However, there seems to be no reason why a single layer should correspond to only one population. It is possible that performance improvements can be achieved by treating spiking layers as heterogeneous networks [3, Ch. 12.2.2]. This would require an improved regularization term that incentivizes clustering but does not limit the number of clusters to one.

Initialization of τ It still remains a mystery why the initializations of $\tau_0 = 2$ and $\tau_0 = 16$ produced different results in the previous work [2, Fig. 6].

Adaptive τ **learning** The baseline model's τ in the sixth layer overshoots and corrects during the first 200 epochs, which suggests a complex training dynamic between input weights and τ parameters. This interplay requires further analysis and could enable more efficient learning by dynamically adapting the respective learning rates.

Local regularization It is possible that different layers would benefit from different regularization terms.

References

[1] Jason K. Eshraghian, Max Ward, Emre O. Neftci, Xinxin Wang, Gregor Lenz, Girish Dwivedi, Mohammed Bennamoun, Doo Seok Jeong, and Wei D. Lu. Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE*, 111(9):1016–1054, 2023.

[2] Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 2641–2651. IEEE, 2021.

[3] Wulfram Gerstner, Werner Kistler, Richard Naud, and Liam Paninski. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. 08 2014.