SYNAPTIC DIVERSITY IN ANNS CAN FACILITATE FASTER LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Various advancements in artificial neural networks (ANNs) are inspired by biological concepts, e.g., the artificial neuron, an efficient model of biological nerve cells demonstrating learning capabilities on large amounts of data. More recent inspirations with promising results are advanced regularization techniques, e.g., synaptic scaling, and backpropagation alternatives, e.g., Targetprop. While neurosciences continuously discover and better understand the mechanisms of biological neural networks (BNNs), new opportunities for a transfer of these concepts towards ANNs arise. However, only a few concepts are readily applicable, and improvements for ANNs are far from being guaranteed. In this paper, we focus on the inhomogeneous and dynamically changing structures of BNNs in contrast to mostly homogeneous and fixed topologies of ANNs. More specifically, we transfer concepts of synaptic diversity, namely spontaneous spine remodeling, diversity in synaptic plasticity, and multi-synaptic connectivity to ANNs. We observe ANNs enhanced by these concepts to learn faster, predict with higher accuracy, and be more resilient to gradient inversion attacks. Our proposed methods are easily applicable to existing ANN topologies and are therefore supposed to stimulate an adaptation of and further research into these mechanisms.

1 INTRODUCTION

Biologically motivated methods have been frequently proposed in machine learning and artificial intelligence and many important theories and methods are inspired by biology, e.g., artificial neurons like the McCulloch-Pitts cell (McCulloch & Pitts, 1943), backpropagation (Rumelhart et al., 1986; LeCun et al., 2012), or the visual cortex inspired CNNs (Hubel & Wiesel, 1968; Lecun et al., 1998; Krizhevsky et al., 2012). Follow-up studies again refer to and discuss the biological plausibility of these methods, e.g., Targetprop being more plausible than backpropagation while yielding similar performance (Lee et al., 2015; Meulemans et al., 2020; Bartunov et al., 2018) or biologically inspired random backward connections not existing in today's artificial networks (Lillicrap et al., 2016). Other studies propose entirely new methods based on whole new models of artificial neurons (Pogodin & Latham, 2020). Among these studies, especially those proposing new kinds of neural connections (Hasani et al., 2019), focusing on binocular data processing (Nguyen et al., 2016), and studying spiking neural networks recently gained research focus (Woźniak et al., 2020). A promising candidate for future advances are plasticity rules like Hebbian plasticity (Hebbs, 1949; Fox & Stryker, 2017). Bredenbberg et al. used synaptic plasticity and a recurrent framework to realize unsupervised learning (cp. (Bredenberg et al., 2020)). Limbacher et al. designed a network which is capable of harnessing synaptic plasticity to learn and recall input-output associations (Limbacher & Legenstein, 2020). Hofmann and Mäder (Hofmann & Mäder, 2021) propose synaptic scaling, a regularization based on a plasticity rule (Tetzlaff et al., 2011). Blier et al. found that training with random learning rates spanning over magnitudes improved robustness to hyper-parameter variations (Blier et al., 2020). Few of the recent advances, however, are directly compatible with state-of-the-art artificial network architectures and the limited scalability of biologically plausible learning has also been discussed as a limitation (Bartunov et al., 2018).

From a neuroscience perspective, biological neural networks are highly complex structures consisting of a wide array of neuron and synapse types. Decades of experimental neuroscience research revealed

that synapses are diverse and dynamic. Their number, molecular composition and morphology are constantly subject to changes (Choquet & Triller, 2013; Fauth & Tetzlaff, 2016; Berry & Nedivi, 2017). Such modulations result in synapse and neuron specific synaptic plasticity, multi-synaptic connectivity between pairs of neurons and spontaneous remodeling of synaptic connections. In contrast, ANNs do not feature this synaptic diversity yet.

In this paper, we aim to study synaptic diversity added to common ANN architectures. More specifically, we study the three biologically inspired phenomena: diversity in synaptic plasticity, spontaneous spine remodelling, and multi-synaptic connectivity. For each, we propose a computationally light-weight realization aiming for applicability across state-of-the-art architectures. In an experimental setup with three of these architectures and three benchmark datasets, we evaluate each method separately and in combination in terms of learning speed, model performance, and robustness against gradient inversion.

We will publish the code for our proposed methods and a converter function for patching arbitrary PyTorch models with our proposed methods on acceptance of the manuscript.

2 Background on Synaptic Diversity of BNNs

Diversity in synaptic plasticity. Learning is associated with synaptic plasticity, which is the ability of synapses to change their conductivity in response to neuronal activity. The most established forms of synaptic plasticity are Long-Term Potentiation (LTP) and Depression (LTD). However, the amount of potentiation and depression expressed at synapses depends on various factors such as the type and frequency of neuronal firing patterns but also on the brain region, neuron and synapse type (Abbott & Nelson, 2000; Buchanan, 2010). In addition, the location of synapses on the dendritic tree influences synaptic plasticity. Back-propagating action potentials are attenuated as they travel along the dendritic tree, thereby, synapses distant from the soma are less susceptible to potentiation than proximal synapses (Froemke, 2010; Bono & Clopath, 2017). The ability of synapses to undergo activity-induced changes also depends on the synapse's and neuron's history, size and age. Such history dependent modulation of synaptic plasticity is termed metaplasticity (Abraham, 2008). While there exist numerous biological mechanisms that mediate metaplasticity, metaplasticity manifests itself in that it inhibits or facilitates the induction and persistence of potentiation and depression of synaptic strength. Recent studies have shown that the stimulus-response and population coupling of individual neurons are variable. Whereas some neurons exhibit stable responses to a specific stimulus over days, others are highly dynamic (Ranson, 2017; Okun et al., 2015; Sweeney & Clopath, 2020). It has been hypothesized that this variability originates from different, neuron-specific learning rates, which has functional implications: whereas neurons with high learning rates can flexibly learn new stimulus associations, less plastic neurons act as a stable, perturbation resistant "backbone" of stimulus representations (Sweeney & Clopath, 2020). Thus, there exist numerous mechanisms that influence synaptic plasticity and, thereby, the rate at which synapses undergo changes. In contrast, in ANNs, the learning rate is usually fixed across the network. We integrate diverse learning rates into our model by applying a constant random factor to the gradient of each synapse in the network, a method we call fuzzy learning rates. Blier et al. observe that variation in learning rates of artificial neurons benefits hyper-parameter robustness (Blier et al., 2020). We hypothesize that diversity in synaptic plasticity, realized as fuzzy learning rates, may have a stabilizing and regularizing effect on the learning.

Spontaneous spine remodeling. In biological neural networks, most excitatory synapses form onto dendritic spines adjoined by axonal terminals. Dendritic spines grow, stabilize, and get pruned in an activity-dependent manner. Activity-dependent spine formation and pruning can be Hebbian, acting on a timescale of hours or homeostatic, acting on a timescale of days (Fauth & Tetzlaff, 2016). In addition to these activity-driven changes, dendritic spines are also subject to activity-independent, spontaneous remodeling and degradation (Ziv & Brenner, 2018; Fauth et al., 2015). Experimental studies found that the survival probability of dendritic spines is independently determined by their size and age (Loewenstein et al., 2015; Trachtenberg et al., 2002). Thereby, some spines turn over within days (Loewenstein et al., 2015; Holtmaat et al., 2005; Fauth et al., 2017) whereas others are stable for months (Yang et al., 2009; Holtmaat et al., 2005; Trachtenberg et al., 2002). In addition, there exists a positive correlation between spine size and synaptic strength (Matsuzaki et al., 2004). Thus, as synapses experience potentiation of synaptic strength, their survival probability increases. Consistent

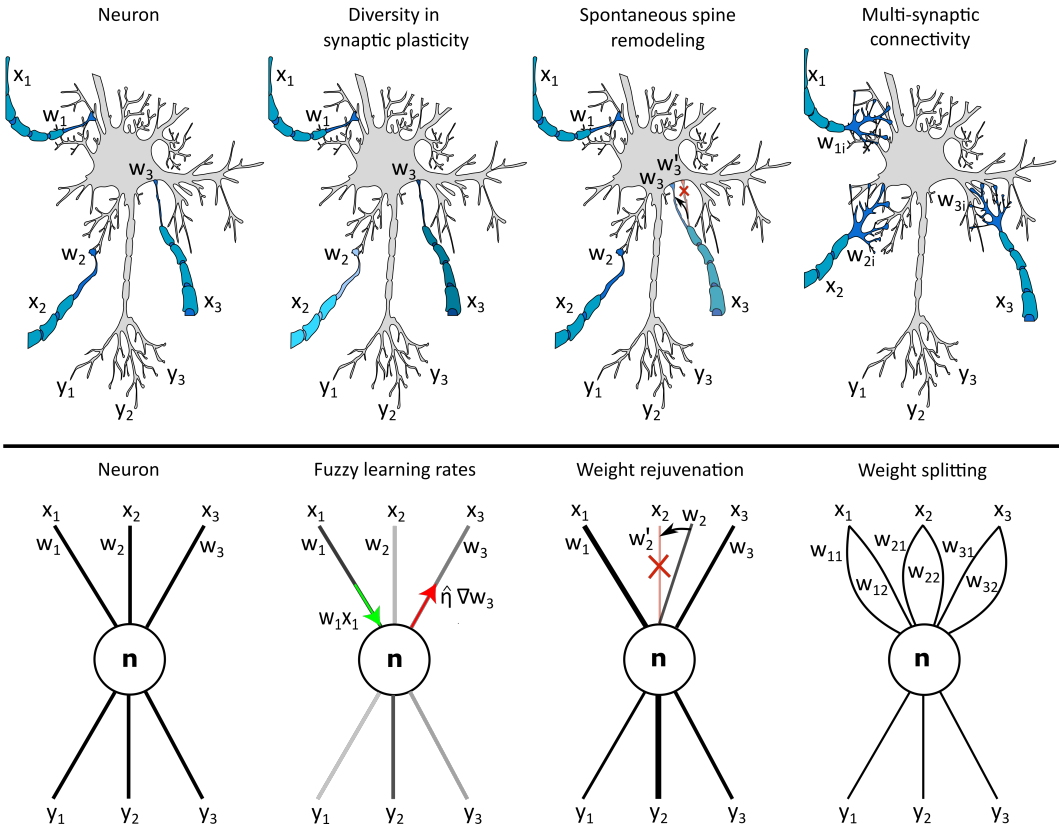


Figure 1: Top: illustration of a nerve cell highlighting each studied synaptic diversity mechanism. Here, x_i denotes the input from the pre-synaptic neuron, w_i the synaptic weight, and y_i the output of the postsynaptic neuron. Thereby, the leftmost illustration refers to a nerve cell as interpreted by most ANN models today. The second illustration shows a neuron where individual synapses show different learning speeds, highlighted by varying color brightness. The third illustration shows small synapses that are subject to random re-initialization, representing spine remodeling and pruning. The weight of the pruned synapse is indicated by an apostrophe. The fourth illustration shows a neuron with multiple synapses between two neurons. Bottom: a more formal representation of each neuron. Depiction of the biological neuron is adapted from (Mariana Ruiz Villarreal, 2007) (released into public domain).

with this notion, in mice, the fraction of persistent spines increases during development (Holtmaat et al., 2005). Finally, whereas spine maturation is often associated with long-term potentiation, several studies found that spines mature and thereby form functional synapses even in the absence of synaptic activity (Harms & Craig, 2005; Sigler et al., 2017; Sando et al., 2017). Thus, dendritic spines are highly dynamic. By contrast, in ANNs, connections between neurons are often represented as stable units that do not experience spontaneous remodeling and pruning. We formalized a model inspired by the above findings, in which synapses are spontaneously reinitialized dependent on their current weight. We call this method weight rejuvenation. Methods such a DropConnect randomly drop synaptic weights resulting in noisy activation and yielding improved generalization. Furthermore, trained ANNs are often characterized by relatively few large weights and a majority of small weights (Hofmann & Mäder, 2021). Therefore, iterative rejuvenation of small weights is not expected to be harmful to the current training progress but may help explore new training directions.

Multi-synaptic connectivity. The ongoing formation and remodeling of spines and their respective synapses result in multi-synaptic connections between pairs of neurons, averaging around 3-5 (Fauth & Tetzlaff, 2016; Markram et al., 1997; Feldmeyer et al., 2006). By contrast, in ANNs, connections are usually modeled by a single synapse. Multi-synaptic connections may have various functional implications. It has been shown theoretically, for example, that the collective dynamics of multiple

synapses can store information for long duration despite synaptic turnover (Fauth et al., 2017). In the current work, we abstract the concept of multi-synaptic connections such that it can be easily applied to existing ANN architectures. We show that such multi-synaptic ANNs are more resilient to gradient inversion attacks. Connecting multiple synapses to a single input is expected to allow for new activation patterns. For example, an input connected to a neuron via a positive and a negative weight will change activation statistics. Multi-synaptic connectivity also distributes the gradient across several synapses in back-propagation, which may harden an ANN against gradient inversion attacks.

3 METHOD

3.1 FUZZY LEARNING RATES (**I**)

We propose fuzzy learning rates as a method (**I**) for realizing diversity in synaptic plasticity in artificial neural networks. Fuzzy learning rates refers to varying synaptic learning rates $\hat{\eta}_{n,i}$. One $\hat{\eta}_{n,i}$ is applied to each synapse with corresponding weight $w_{n,i}$ that belongs to a layers' neuron. Neurons are enumerated with $n = 0, 1, 2, ...$ and $i = 0, 1, 2, ...$, where n denotes the post- and i the pre-synaptic neuron. We denote the neural transmission function without bias as $\phi_n = g(\sum_{i \in I} w_{n,i} x_i)$, where x_i refers to the input from neuron i and g refers to an arbitrary non-linearity, e.g., a rectified linear unit function (ReLU) (Hahnloser et al., 2000). Accordingly, a learning rate for every individual synapse is realized as a constant random factor applied to its gradient. Therefore, a typical gradient descent step changes from $w_{n,i,t+1} = w_{n,i,t} - \eta \nabla \phi_{n,i}$ to

$$w_{n,i,t+1} = w_{n,i,t} - \eta \nabla \phi_{n,i} \odot \hat{\eta}_{n,i}. \tag{1}$$

A factor $\hat{\eta}_{n,i}$ is randomly drawn from a uniform distribution per weight upon initialization of the network

$$\hat{\eta}_{n,i} \leftarrow \mathcal{U}(1 - \frac{\tau}{2}, 1 + \frac{\tau}{2}), \tag{2}$$

where $\mathcal{U}$ is the uniform distribution and τ is a scaling range.

The method's run time is independent of the size of the input sample and is executed once for all weights. However, it increases the number of operations needed to forward the network linearly with the model's size.

3.2 WEIGHT REJUVENATION (**II**)

We propose weight rejuvenation as a method (**II**) to realize random re-initialization of synaptic connections, inspired by spontaneous spine remodeling of biological neural networks. Weight rejuvenation means that a weight $w_{n,i}$ is reset to a random value with a certain probability, mimicking spine purging and formation. More specifically, the smaller a weight becomes throughout a training process, the higher its probability of being reinitialized. We denote this probability as

$$P_{\text{re}} = 1 - \frac{1}{\sigma_{\text{re}} \sqrt{2\pi}} \int_{-\infty}^{w_{n,i}} e^{-\frac{1}{2} \frac{t-\mu}{\sigma_{\text{re}}}^2} dt, \tag{3}$$

where σ_{re} is the rejuvenation variance calculated with respect to the maximum value of a layers' synaptic weights

$$\sigma_{\text{re}} = w_{\max} / d_{\text{re}}, \tag{4}$$

where d_{re} is the rejuvenation distance factor. For example, a rejuvenation distance factor of 1 means that the maximum synaptic weight of a layer $w_{\max}$ is reinitialized with a probability of $\sim 16\%$. Figure 2 shows that the number of rejuvenated synaptic weights is large in the beginning and decreases during training. After an initial phase, the number stagnates on a certain level, introducing further noise to the synaptic weights. The time consumption is independent of the size of the input data and is calculated once per training step across all weights. However, weight rejuvenation increases the overall number of a network's operations linearly with the model's size.

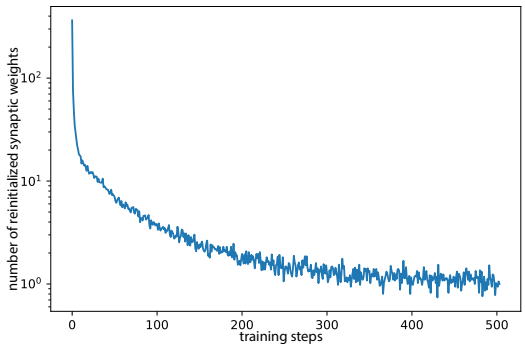


Figure 2: Course of the number of rejuvenated synaptic weights during 500 training steps of a shallow MLP.

3.3 Weight Splitting (**III**)

We propose weight splitting as method (**III**) for incorporating multi-synaptic connectivity into artificial neural networks, inspired by the observation that biological neurons can have several connections among one another. We realize this behavior by replicating $|\Gamma|$ times a neuron's linear units and aggregating their results before activation. These multiple linear units allow for varying weights per synapse. Thereby, Γ denotes the set of indices of the replicated neurons and its cardinality may be interpreted as the number of connections between a pair of neurons. Accordingly, a layers' transmission function is denoted as

$$\phi_n = g\left(\sum_{\gamma\in\Gamma}\sum_{i\in I} w_{n+\gamma\lfloor\frac{N}{|\Gamma|}\rfloor,i}\, x_i\right) \forall\, 0 < n < \frac{N}{|\Gamma|}, \tag{5}$$

where $\lfloor\frac{N}{|\Gamma|}\rfloor$ denotes the distance of the accumulated synapses indices. Split weights share the same gradient update. Therefore, their absolute value will sustain a constant difference during training (cp. Fig. 5 in the Appendix). However, the non-linearity alters the activation so that the learning behavior changes. Our method increases the overall number of a network's operations linearly with the model's size but independently on the input sample size.

4 Experimentation

We evaluate our proposed methods with experiments on accuracy and learning progress and the robustness against gradient inversion.

4.1 Setup

To evaluate the effects of the proposed biological-inspired methods, we study them in three experiments. First, we aim to examine whether the proposed methods yield comparable or even improved accuracy when integrated into typical model architectures and training benchmark problems. Second, we aim to examine whether the proposed methods help neural networks learn with comparable or even faster speed. Finally, latest research hypothesizes that gradient inversion attacks (Geiping et al., 2020) are most effective if single neurons have high gradients during distributed learning of classification tasks (Pan et al., 2020). Therefore, we argue that weight splitting might mitigate gradient inversion attacks and aim to evaluate this question.

To study how the proposed methods affect artificial neural architectures' ability, we run three-fold cross-validated experiments on the MNIST (LeCun et al., 1998), CIFAR-10 and CIFAR-100 (Krizhevsky, 2009) benchmarks. We normalize the data samples using mean and standard deviation calculated on the training splits. Moreover, we examine three network architectures: a shallow learning MLP, a modified version of AlexNet (Krizhevsky et al., 2012) and a ResNet20/32/56 (He et al., 2016). We use a cross-entropy loss function across all the classification experiments. The MLP

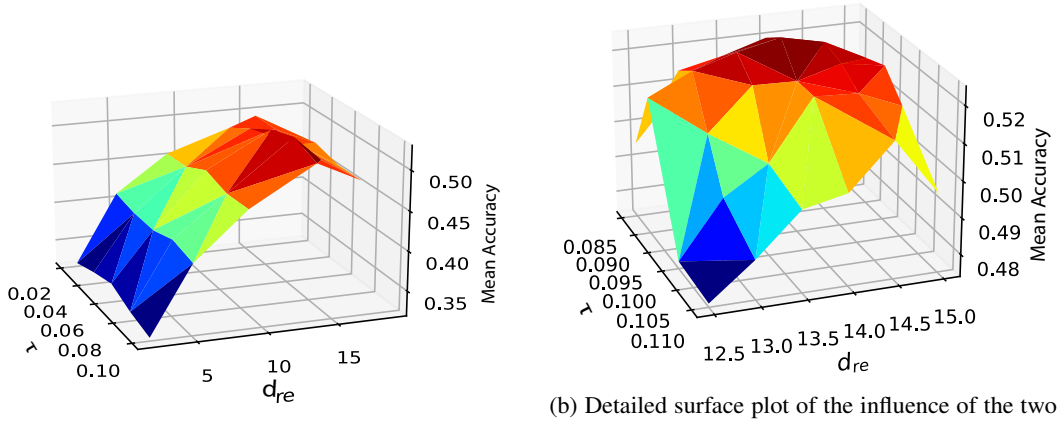
consists of one hidden layer with 1,000 neurons for the MNIST training and two hidden layers with 3,000 neurons each for the CIFAR-10 and CIFAR-100 trainings. The models with weight splitting share the same number of trainable parameters as the models without weight splitting. Accordingly, we duplicated the activations of each layer such that the number of activations of each layer persists.

All experiments are performed with the same general hyper-parameters. We use SGD as optimizer with a learning rate of $\eta = 0.01$. No other augmentation or regularization is used, except for the inherent methods per architecture, i.e., residual connections and batch normalization of the ResNet architecture. We train a network for 100 epochs and retrospectively determine the epoch where accuracy was not increasing for five following consecutive epochs (early stopping). We report this epoch as a measure for learning speed and report the model's test accuracy at this epoch. All experiments together resulted in 1,200 h of training time on GPUS of Type Nvidia 2080 Ti. The learning speed is determined in an early stopping scheme.

Our experiments that evaluate the robustness against gradient inversion attacks are based on a method described by (Geiping et al., 2020) [1]. This method is used to extract batch samples based on gradients and yields a measure for privacy in federated learning of sensible data, like health or finance data. In our setting, we present the mean reconstruction loss for five fixed consecutive batches of eight samples from CIFAR-100 reconstructed from untrained networks and networks trained for 100 epochs.

4.2 Hyper-Parameters of the Proposed Methods

Our proposed methods require adaptable hyper-parameters. To obtain mostly unbiased results, we train for ten epochs and a batch size of 1,000 using nevergrad (Rapin & Teytaud, 2018) with a budget of 100. All experiments except for the gradient reconstruction are performed on a random 2:1 split of the MNIST dataset. We determine the parameters $\tau = 0.09$, $d_{\text{re}} = 14$,and $\Gamma = 2$ to yield the highest accuracy in this setting. How accuracy changes when the parameters τ and d_{re} are changed is shown in Figure 3a. Beyond d_{re} of 15 the effect diminishes, the mean accuracy is no longer affected and stays at a level of 0.45 after 10 epochs training. All Γ values grater 2 yield inferior results. The range in which the parameters can be chosen to still have a positive effect ranges from $0.02 < \tau < 0.1$ and $7 < d_{\text{re}} < 15$. We show additional observations on the stability towards slight learning rate variations in the appendix 9.2. All hyper-parameters are fixed for all experiments.



(a) Surface plot of the influence of the two parameters τ and d_{re} on the accuracy in a hyper parameter tuning.

(b) Detailed surface plot of the influence of the two parameters τ and d_{re} on the accuracy in a hyper parameter tuning.

Figure 3: Plots showing results of the hyper parameter tuning experiment.

5 Results

We present our results for the experiments on accuracy across two tables (cp. Tables 1 and 2), while Table 3 shows the results for our gradient inversion experiment.

[1] Implementation found at: https://github.com/JonasGeiping/invertinggradients

5.1 ACCURACY AND LEARNING PROGRESS

Overall the accuracy observed on the different combinations fluctuates on a higher than baseline level and is never significantly worse than the baseline. The results are presented in Table 1. Especially in the MLP setting on the CIFAR-10 dataset, the results are not significantly different from the baseline. The highest accuracy is achieved when fuzzy learning rates, weight rejuvenation, and weight splitting are combined, but we also observe experiments without a significant effect on the accuracy. Especially using our methods in the MLP/CIFAR-10 setting shows no large effect with the highest accuracy of 56.13% compared to the baseline of 55.00%. Another observation is that the CIFAR-10 and CIFAR-100 settings were unstable for the AlexNet and ResNet56 experiments with observed accuracy of 19.4% and 1.3%. In these cases, stable results are only observed when weight splitting is used, which resulted in an accuracy of 33.71%; $32,41\%$ higher than the baseline. A similar observation regarding training stability is made in the settings AlexNet/CIFAR-10 and ResNet56/CIFAR-10. However, the results using weight rejuvenation show lower accuracy than the combinations without weight rejuvenation. Nevertheless, accuracy improvements of 21.45% and 14.42% are observed in the AlexNet/MNIST and ResNet56/CIFAR-10 settings. Intermediate results show minor improvements around 2% to 5%

We present our results on the learning speed experiments in Table 2. The largest number of epochs to achieve the highest accuracy are observed for the baseline models, whereas we observe the lowest number of epochs with a combination of weight rejuvenation and weight splitting. We observe best learning speeds when only applying fuzzy learning rates for two method-dataset combinations, while we observe mediocre results beyond the baseline for the rest of the combinations. Overall low or lowest numbers of epochs are observed in settings where all methods are combined. However, we observed worse numbers (85 and 85), largely deviating from the best results, for the AlexNet experiments on the MNIST and CIFAR-100 datasets with epochs compared to 68 and 56 epochs. For the other experiments, this combination shows no large difference to the best observed numbers.

5.2 ROBUSTNESS AGAINST GRADIENT INVERSION

The results for the gradient inversion experiment are displayed in Table 3. We calculated the optimal reconstruction error based on the gradients of a batch. The reconstruction attacks are successful if they can achieve low reconstruction errors. Overall, the reconstruction errors are small, so we decided to report them in percent. Especially for the untrained and unmodified shallow MLP and the AlexNet, the attacks are highly successful. With errors of 3.47% and 0.22% the reconstructed images show fine details like displayed in Figure 4. We observe the largest reconstruction errors with 135.35% and 155.43% for the modified (all methods) untrained and trained ResNet32 architecture. We also observed that the weight splitting alone improves the MSE in most cases, even in the untrained cases. Fuzzy learning rates alone do not improve the MSE without training in nearly all cases but lead to improvements combined with weight splitting.

6 DISCUSSION

We observed overall improving effects of a combination of all methods on the accuracy and learning speed in many experiments (dataset-architecture combinations) but also observed nearly no effect in other experiments. In particular, the MLP/MNIST combination often does not benefit from the proposed methods. Considering the accuracy results and the learning speed results, we can see that the combination of weight splitting and weight rejuvenation yields the highest accuracy together with the fastest learning speeds. Furthermore, the fuzzy learning rates are often beneficial regarding accuracy but tend to slow down learning in the more shallow architectures. Finally, this observation is confirmed by the gradient inversion experiments where the fuzzy learning rates improve the MSE in all cases except the smallest untrained architecture. Interestingly, the observation that the instabilities in our experiments are reduced and therefore the resilience against inappropriately chosen parameters is increased has resemblance with the question from neuroscience of how our brain can function reliably in the presence of various noise sources (Faisal et al., 2008).

Our experimental setup is based on three architectures and datasets; although we tested on networks incorporating convolutions, batch normalization, and residual connections, we can not conclude that our methods have a general effect, especially for a wide variety of hyperparameter combinations as

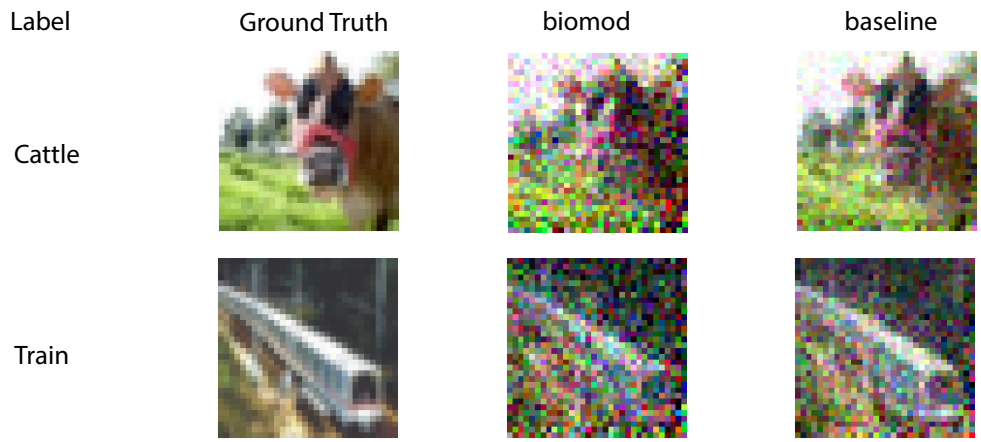


Figure 4: Images with the lowest reconstruction error of all batches reconstructed from the AlexNet architecture trained on 100 epochs. Left side with our methods combined; on the right side without our methods.

Table 1: Results of the experiment on accuracy. The table shows the observed mean accuracy in percent and standard deviations for 3-fold cross-validation runs for a given combination of parameters for the datasets MNIST (M10), CIFAR-10 and CIFAR-100 (C10 and C100). The highest accuracy for a dataset and architecture combination is highlighted in bold. The numbers **I**, **II** and **III** denote the methods of fuzzy learning rates, weight rejuvenation, and weight splitting, respectively.

Methods			MLP			AlexNet			ResNet56		
			acc [%]↑ ± std.			acc [%]↑ ± std.			acc [%]↑ ± std.		
I	**II**	**III**	M10	C10	C100	M10	C10	C100	M10	C10	C100
			95.70	55.00	23.62	76.83	19.40	1.30	96.28	50.54	37.39
			±0.21	±0.73	±0.56	±5.56	±0.75	±0.20	±0.11	±1.52	±0.38
		√	96.81	55.73	**28.33**	97.99	63.01	28.29	98.08	61.52	36.79
		√	±0.23	±0.90	±0.50	±0.05	±0.76	±0.52	±0.13	±1.16	±0.19
	√		96.63	**56.13**	26.91	97.72	57.64	2.54	96.51	50.92	33.88
	√		±0.19	±0.81	±0.64	±0.09	±0.50	±0.20	±0.34	±1.09	±0.24
	√	√	97.18	54.80	27.12	98.11	63.01	**33.71**	96.38	50.74	43.13
	√	√	±0.14	±0.96	±0.39	±0.06	±0.75	±0.83	±0.51	±1.68	±1.21
√			95.72	55.51	23.43	76.47	21.03	0.97	**98.24**	60.92	41.12
√			±0.30	±0.72	±0.53	±4.18	±0.66	±0.11	±0.14	±0.26	±0.47
√		√	96.79	55.49	28.29	98.15	62.93	28.22	98.14	62.04	35.14
√		√	±0.18	±0.40	±0.49	±0.02	±0.95	±0.93	±0.11	±0.52	±0.74
√	√		96.74	56.01	26.74	97.87	57.36	2.11	96.61	51.28	32.45
√	√		±0.14	±0.61	±0.45	±0.10	±0.43	±0.39	±0.55	±1.01	±1.41
√	√	√	**97.25**	54.42	27.07	**98.28**	**63.42**	33.14	96.20	**64.96**	**44.37**
√	√	√	±0.14	±0.60	±0.57	±0.14	±0.89	±0.16	±0.56	±1.47	±1.26

learning rates or optimizers. We observed that the combination of all our methods shows superior learning speed and classification accuracy in our experiments, but we did not observe a general effect. This is most likely due to poor parameterization leading to inferior baseline results and poor performance of some experimental settings. Especially the learning rate and optimizer may not have been optimal for all neural architectures and dataset combinations. However, we hypothesize that our methods can be highly beneficial if data privacy is needed in a field of limited computing time and resources.

Table 2: Results of the experiment on learning speed. The table shows the observed mean epoch when the maximum accuracy occurred, and its standard deviations for 3-fold cross-validation runs for a given combination of parameters for the datasets MNIST (M10), CIFAR-10, and CIFAR-100 (C10 and C100). The lowest numbers of trained epochs for dataset and architecture combinations are highlighted in bold. The numbers **I**, **II** and **III** denote the methods of fuzzy learning rates, weight rejuvenation, and weight splitting, respectively.

Methods			MLP ep(max acc)↓ ± std.			AlexNet ep(max acc)↓ ± std.			ResNet56 ep(max acc)↓ ± std.		
I	**II**	**III**	M10	C10	C100	M10	C10	C100	M10	C10	C100
			99 ±0.8	99 ±0.8	99 ±1.9	99 ±0.9	93 ±5.7	93 ±1.3	99 ±0.8	98 ±1.6	99 ±0.2
		√	100 ±0.5	67 ±9.5	96 ±1.7	79 ±9.5	76 ±7.1	89 ±2.8	86 ±4.9	90 ±5.5	78 ±7.6
	√		98 ±1.7	96 ±1.7	99 ±0.9	94 ±5.3	99 ±0.9	79 ±13.0	96 ±1.6	97 ±3.1	93 ±2.0
	√	√	88 ±3.7	**46** ±3.3	**62** ±5.0	**68** ±14.3	59 ±9.0	77 ±2.9	87 ±2.1	86 ±0.5	91 ±0.5
√			**81** ±4.1	91 ±7.4	88 ±11.9	82 ±7.9	84 ±8.9	**56** ±14.3	88 ±1.4	87 ±1.6	93 ±1.7
√		√	89 ±1.41	74 ±7.78	93 ±4.97	88 ±4.11	90 ±6.94	87 ±2.36	88 ±0.47	100 ±0.47	100 ±0.0
√	√		87 ±7.3	85 ±9.5	83 ±13.6	83 ±5.4	90 ±0.5	87 ±4.2	84 ±4.5	88 ±0.8	92 ±1.7
√	√	√	86 ±1.7	55 ±7.6	62 ±6.9	85 ±10.7	**58** ±3.3	85 ±11.5	**82** ±7.9	**85** ±4.5	**74** ±4.2

Table 3: Results of the experiment on gradient inversion. The table shows the observed reconstruction error (mean square error) for different combinations of methods for untrained models and models trained 100 epochs on CIFAR-100. Highest reconstruction errors for the dataset and architecture combinations are highlighted in bold. The numbers **I**, **II** and **III** denote the methods of fuzzy learning rates, weight rejuvenation and weight splitting respectively.

Methods			MLP MSE[%]↑		AlexNet MSE[%]↑		ResNet20 MSE[%]↑		ResNet32 MSE[%]↑	
I	**II**	**III**	0 ep	100 ep	0 ep	100 ep	0 ep	100 ep	0 ep	100 ep
			3.47	3.62	0.22	0.20	62.20	73.65	70.78	79.61
		√	30.21	31.26	28.30	30.33	88.76	89.19	94.31	94.71
	√		3.06	4.31	0.26	0.21	59.07	62.21	70.58	81.14
	√	√	41.42	45.04	28.38	39.06	75.93	85.49	81.34	84.37
√			**55.53**	60.05	42.11	46.96	49.93	50.05	75.47	78.71
√		√	42.70	58.73	94.47	96.85	81.68	86.74	97.31	102.43
√	√		47.74	55.30	35.43	53.60	59.07	62.21	94.52	107.36
√	√	√	49.65	**64.84**	**101.79**	**112.85**	**124.21**	**132.17**	**135.35**	**155.43**

7 FUTURE WORK

The proposed methods are inspired by nature and showed a beneficial effect on classification tasks and the prevention of gradient inversion attacks. We did not provide any theoretical foundation for the different effects. Future work can focus on a solid theoretical foundation for the observed effects. Here, one should investigate the influence of the proposed methods on the weights and activations, different layer types, and optimizers. Expertise in computational neurosciences and machine learning and a more extensive experimental study, including various model architectures, datasets, and tasks, is needed to build such a theoretical foundation. Another important future work will be to understand better why these methods are beneficial in our specific settings.

8 REPRODUCIBILITY STATEMENT

Our reproduction package contains all code needed to reproduce our experiments, including the hyperparameters and network implementations. We also included a readme with instructions on installing the python package and running experiments updating arbitrary state-of-the-art architectures. We will provide the package to the reviewers only through a comment with restricted visibility. It will be published on acceptance of the manuscript.

REFERENCES

L. F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3 (S11):1178–1183, nov 2000. doi: 10.1038/81453.

Wickliffe C. Abraham. Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews Neuroscience*, 9(5):387–387, may 2008. doi: 10.1038/nrn2356.

Sergey Bartunov, Adam Santoro, Blake Richards, Luke Marris, Geoffrey E Hinton, and Timothy Lillicrap. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/63c3ddcc7b23daa1e42dc41f9a44a873-Paper.pdf.

Kalen P. Berry and Elly Nedivi. Spine dynamics: Are they all the same? *Neuron*, 96(1):43–55, sep 2017. doi: 10.1016/j.neuron.2017.08.008.

Léonard Blier, Pierre Wolinski, and Yann Ollivier. Learning with random learning rates. In Ulf Brefeld, Elisa Fromont, Andreas Hotho, Arno Knobbe, Marloes Maathuis, and Céline Robardet (eds.), *Machine Learning and Knowledge Discovery in Databases*, pp. 449–464, Cham, 2020. Springer International Publishing. ISBN 978-3-030-46147-8.

Jacopo Bono and Claudia Clopath. Modeling somatic and dendritic spike mediated plasticity at the single neuron and network level. *Nature Communications*, 8(1), sep 2017. doi: 10.1038/s41467-017-00740-z.

Colin Bredenberg, Eero Simoncelli, and Cristina Savin. Learning efficient task-dependent representations with synaptic plasticity. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 15714–15724. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/b599e8250e4481aaa405a715419c8179-Paper.pdf.

Katherine A. Buchanan. The activity requirements for spike timing-dependent plasticity in the hippocampus. *Frontiers in Synaptic Neuroscience*, 2, 2010. doi: 10.3389/fnsyn.2010.00011.

Daniel Choquet and Antoine Triller. The dynamic synapse. *Neuron*, 80(3):691–703, oct 2013. doi: 10.1016/j.neuron.2013.10.013.

A. Aldo Faisal, Luc P. J. Selen, and Daniel M. Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, apr 2008. doi: 10.1038/nrn2258.

Michael Fauth and Christian Tetzlaff. Opposing effects of neuronal activity on structural plasticity. *Frontiers in Neuroanatomy*, 10, jun 2016. doi: 10.3389/fnana.2016.00075.

Michael Fauth, Florentin Wörgötter, and Christian Tetzlaff. The formation of multi-synaptic connections by the interaction of synaptic and structural plasticity and their functional consequences. *PLOS Computational Biology*, 11(1):e1004031, jan 2015. doi: 10.1371/journal.pcbi.1004031.

Michael Fauth, Florentin Wörgötter, and Christian Tetzlaff. Long-term information storage by the interaction of synaptic and structural plasticity. In *The Rewiring Brain*, pp. 343–360. Elsevier, 2017. doi: 10.1016/b978-0-12-803784-3.00016-0.

Dirk Feldmeyer, Joachim Lübke, and Bert Sakmann. Efficacy and connectivity of intracolumnar pairs of layer 2/3 pyramidal cells in the barrel cortex of juvenile rats. *The Journal of Physiology*, 575(2): 583–602, aug 2006. doi: 10.1113/jphysiol.2006.105106.

Kevin Fox and Michael Stryker. Integrating hebbian and homeostatic plasticity: Introduction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372 (1715), 2017. doi: 10.1098/rstb.2016.0413.

Froemke. Dendritic synapse location and neocortical spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 2010. doi: 10.3389/fnsyn.2010.00029.

Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients - how easy is it to break privacy in federated learning? In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 16937–16947. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/c4ede56bbd98819ae6112b20ac6bf145-Paper.pdf.

Richard H. R. Hahnloser, Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas, and H. Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951, 2000. doi: 10.1038/35016072. URL https://doi.org/10.1038/35016072.

Kimberly J. Harms and Ann Marie Craig. Synapse composition and organization following chronic activity blockade in cultured hippocampal neurons. *The Journal of Comparative Neurology*, 490 (1):72–84, 2005. doi: 10.1002/cne.20635.

Hosein Hasani, Mahdieh Soleymani, and Hamid Aghajan. Surround modulation: A bio-inspired connectivity structure for convolutional neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/c535e3a7f97daf1c4b1eb03cc8e31623-Paper.pdf.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition, CVPR*, pp. 770–778, 2016.

D. G. Hebbs. The organization of behavior. *Wiely and Sons, New York, NY, USA*, 1949.

Martin Hofmann and Patrick Mäder. Synaptic scaling–an artificial neural network regularization inspired by nature. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021. doi: 10.1109/TNNLS.2021.3050422.

Anthony J.G.D. Holtmaat, Joshua T. Trachtenberg, Linda Wilbrecht, Gordon M. Shepherd, Xiaoqun Zhang, Graham W. Knott, and Karel Svoboda. Transient and persistent dendritic spines in the neocortex in vivo. *Neuron*, 45(2):279–291, jan 2005. doi: 10.1016/j.neuron.2005.01.003.

D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968. ISSN 0022-3751. doi: 10.1113/jphysiol.1968.sp008455.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, University of Tronto, 2009.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc, 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database, 1998. URL http://yann.lecun.com/exdb/mnist.

Yann A. LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pp. 9–48. Springer, 2012.

Dong-Hyun Lee, Saizheng Zhang, Asja Fischer, and Yoshua Bengio. Difference target propagation. In Annalisa Appice, Pedro Pereira Rodrigues, Vítor Santos Costa, Carlos Soares, João Gama, and Alípio Jorge (eds.), *Machine Learning and Knowledge Discovery in Databases*, pp. 498–515, Cham, 2015. Springer International Publishing. ISBN 978-3-319-23528-8.

Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7(1):13276, 2016. doi: 10.1038/ncomms13276. URL https://doi.org/10.1038/ncomms13276.

Thomas Limbacher and Robert Legenstein. H-mem: Harnessing synaptic plasticity with hebbian memory networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 21627–21637. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/f6876a9f998f6472cc26708e27444456-Paper.pdf.

Y. Loewenstein, U. Yanover, and S. Rumpel. Predicting the dynamics of network connectivity in the neocortex. *Journal of Neuroscience*, 35(36):12535–12544, sep 2015. doi: 10.1523/jneurosci.2917-14.2015.

Mariana Ruiz Villarreal. Complete neuron cell diagram en, 2007. URL https://commons.wikimedia.org/wiki/File:Complete_neuron_cell_diagram_en.svg. [Online; accessed Mai 28, 2021].

H Markram, J Lübke, M Frotscher, A Roth, and B Sakmann. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of Physiology*, 500(2):409–440, apr 1997. doi: 10.1113/jphysiol.1997.sp022031.

Masanori Matsuzaki, Naoki Honkura, Graham C. R. Ellis-Davies, and Haruo Kasai. Structural basis of long-term potentiation in single dendritic spines. *Nature*, 429(6993):761–766, jun 2004. doi: 10.1038/nature02617.

Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943. ISSN 0007-4985. doi: 10.1007/BF02478259.

Alexander Meulemans, Francesco Carzaniga, Johan Suykens, João Sacramento, and Benjamin F. Grewe. A theoretical framework for target propagation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 20024–20036. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/e7a425c6ece20cbc9056f98699b53c6f-Paper.pdf.

Anh Tuan Nguyen, Jian Xu, and Zhi Yang. A bio-inspired redundant sensing architecture. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/a9078e8653368c9c291ae2f8b74012e7-Paper.pdf.

Michael Okun, Nicholas A. Steinmetz, Lee Cossell, M. Florencia Iacaruso, Ho Ko, Péter Barthó, Tirin Moore, Sonja B. Hofer, Thomas D. Mrsic-Flogel, Matteo Carandini, and Kenneth D. Harris. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515, apr 2015. doi: 10.1038/nature14273.

Xudong Pan, Mi Zhang, Yifan Yan, Jiaming Zhu, and Min Yang. Theory-oriented deep leakage from gradients via linear equation solver, 2020.

Roman Pogodin and Peter Latham. Kernelized information bottleneck leads to biologically plausible 3-factor hebbian learning in deep networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7296–7307. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/517f24c02e620d5a4dac1db388664a63-Paper.pdf.

Adam Ranson. Stability and plasticity of contextual modulation in the mouse visual cortex. *Cell Reports*, 18(4):840–848, jan 2017. doi: 10.1016/j.celrep.2016.12.080.

J. Rapin and O. Teytaud. Nevergrad - A gradient-free optimization platform. https://GitHub.com/FacebookResearch/Nevergrad, 2018.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

Richard Sando, Eric Bushong, Yongchuan Zhu, Min Huang, Camille Considine, Sebastien Phan, Suyeon Ju, Marco Uytiepo, Mark Ellisman, and Anton Maximov. Assembly of excitatory synapses in the absence of glutamatergic neurotransmission. *Neuron*, 94(2):312–321.e3, apr 2017. doi: 10.1016/j.neuron.2017.03.047.

Albrecht Sigler, Won Chan Oh, Cordelia Imig, Bekir Altas, Hiroshi Kawabe, Benjamin H. Cooper, Hyung-Bae Kwon, Jeong-Seop Rhee, and Nils Brose. Formation and maintenance of functional spines in the absence of presynaptic glutamate release. *Neuron*, 94(2):304–311.e4, apr 2017. doi: 10.1016/j.neuron.2017.03.029.

Yann Sweeney and Claudia Clopath. Population coupling predicts the plasticity of stimulus responses in cortical circuits. *eLife*, 9, apr 2020. doi: 10.7554/elife.56053.

Christian Tetzlaff, Christoph Kolodziejski, Marc Timme, and Florentin Wörgötter. Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. *Frontiers in computational neuroscience*, 5:47, 2011. doi: 10.3389/fncom.2011.00047.

Joshua T. Trachtenberg, Brian E. Chen, Graham W. Knott, Guoping Feng, Joshua R. Sanes, Egbert Welker, and Karel Svoboda. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917):788–794, dec 2002. doi: 10.1038/nature01273.

Stanisław Woźniak, Angeliki Pantazi, Thomas Bohnstingl, and Evangelos Eleftheriou. Deep learning incorporating biologically inspired neural dynamics and in-memory computing. *Nature Machine Intelligence*, 2(6):325–336, 2020. doi: 10.1038/s42256-020-0187-0. URL https://doi.org/10.1038/s42256-020-0187-0.

Guang Yang, Feng Pan, and Wen-Biao Gan. Stably maintained dendritic spines are associated with lifelong memories. *Nature*, 462(7275):920–924, nov 2009. doi: 10.1038/nature08577.

Noam E. Ziv and Naama Brenner. Synaptic tenacity or lack thereof: Spontaneous remodeling of synapses. *Trends in Neurosciences*, 41(2):89–99, feb 2018. doi: 10.1016/j.tins.2017.12.003.

9 APPENDIX

9.1 WEIGHT SPLITTING

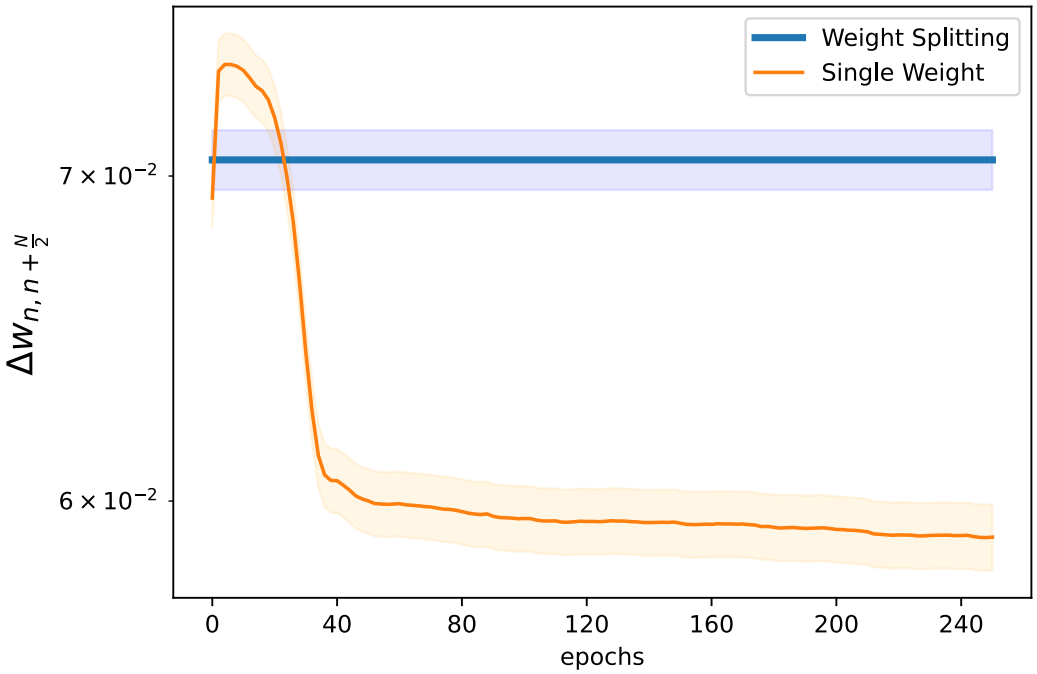
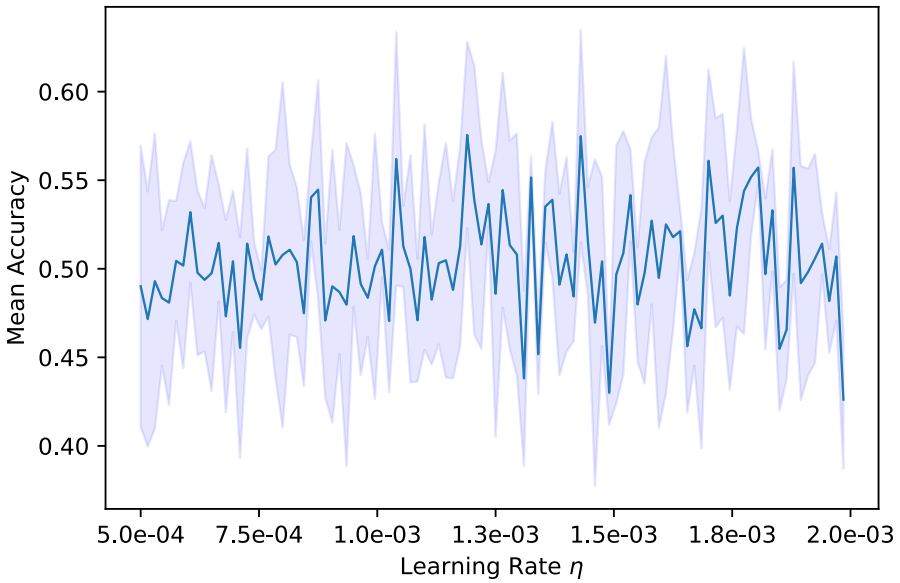
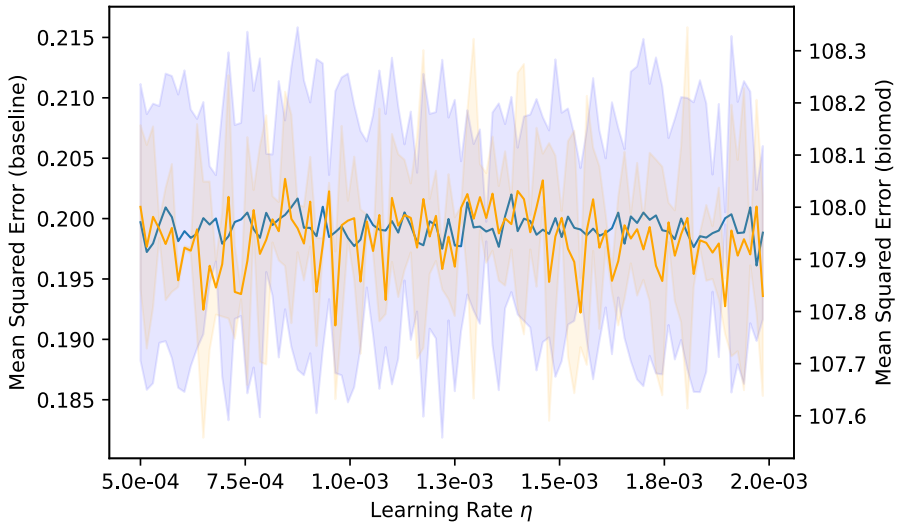


Figure 5: We conducted an experiment on ResNet20 to evaluate how weight splitting is influencing a neuron's synaptic weights with the aim to receive an observation on the balance of the synapses that connect two neurons. The blue curve denotes the case with enabled weight splitting; two neurons are connected by two corresponding synapses in this case and their absolute weight difference is calculated and averaged over the whole network. In case of no weight splitting, denoted by the orange curve, there are no corresponding synapses, but we still calculated the difference to another synapse for comparison. We observe that weight splitting leads to a constant difference in the weights and that in the case without weight splitting the differences shrink. These observation could cause the observed difference in convergence behavior and generalization properties that we observed in our experiments. We also observe that while the absolute values shrink, the variances of the differences are constant throughout the training; this means that all weights are affected in the same way, no matter if they are larger or smaller at the beginning of the training.

9.2 LEARNING RATE VARIATION



(a) Accuracy in the learning rate variation experiment for different learning rates η. The AlexNet is trained on a 1:2 split of MNIST for 10 epochs. We observe that the mean and variance over the whole range of η changes randomly, with a slightly lower mean value towards 0.05 and a marginally lower variance towards 0.2. Therefore, we conclude that a learning rate of 0.1 as a typical standard is a stable choice for our experimental setup.



(b) Mean Square Reconstruction Error in learning rate variation experiment for different learning rates η. The AlexNet is trained on a 1:2 split of CIFAR-100 for 10 epochs and the gradient reconstruction is conducted on the saved weights. **blue** to the left denotes the baseline and **orange** to the right the modified model. We observe a overal larger variation in the baseline than in the modified model but the relative variation of the variances is observed larger for the modified model. While the mean value of the modified model is observed marginally larger for learning rates towards 0.05 and lower towards 0.2, the mean value for the baseline model is observed constant over the whole range. Considering this preliminary observations, we find no significant changes of the models behavior regarding the learning rate.

Figure 6: Plots showing results of the learning rate variation experiment.