Supplementary Material - LLS: Local Learning Rule for Deep Neural Networks Inspired by Neural Activity Synchronization

Marco P. E. Apolinario, Arani Roy , Kaushik Roy
School of Electrical and Computer Engineering, Purdue University
West Lafayette, IN, USA
{mapolina, roy173, kaushik}@purdue.edu

1. Experimental Setup

In this section, we describe the architecture of all models used in this work, the datasets and preprocessing operations, the training details including hyperparameters for each experiment, and the compute resources employed.

1.1. Model architecture

In this work, we use four models: SmallConv, SmallConvL, VGG8 [8], and MobileNetV1 [4]. These models are built using the following three basic blocks: ConvBlock, ConvDWBlock, and LinearBlock.

- ConvBlock is composed of three layers in the following order: a convolutional layer (Conv), a batch normalization layer (BN), and a Leaky ReLU (LeakyReLU).

- ConvDWBlock is composed of five layers in the following order: a depthwise convolutional layer (ConvDW), a BN layer, a Conv layer with kernel size of 1 (Conv1x1), another BN layer, and a LeakyReLU layer.

- LinearBlock is composed of three layers: a fully-connected layer (Linear), a BN layer, and a LeakyReLU.

The architecture of each of the models is described in Table 1. Note that LLS was applied at the outputs of each ConvBlock, ConvDWBlock, and LinearBlock, after the output dimensions were reduced to a size of 2048 (or lower depending on the output dimensions) using an Adaptive Average Pooling (AdaptiveAvgPool) layer.

1.2. Datasets

In this section, we provide a brief description of the datasets used in this work: MNIST [7], FashionMNIST [9], CIFAR10 [5], CIFAR100 [5], IMAGENETTE [3], TinyIMAGENET [6], and Visual Wake Words (VWW) [1].

MNIST: This dataset consists of 70000 grayscale images of handwritten digits (0-9), each of size 28x28 pixels. It is divided into 60000 training images and 10,000 test images.

FashionMNIST: This dataset consists of 70000 grayscale images of fashion items, such a clothing and accessories, each of size 28x28 pixels. Similar to MNIST, it is divided into 60,000 training images and 10000 test images.

CIFAR10: This dataset consists of 60000 color images in 10 different classes, with each class containing 6000 images. The images are 32x32 pixels in size and the dataset is split into 50000 training images and 10000 test images.

CIFAR100: It is similar to CIFAR-10 but contains 100 classes with 600 images per class. The images are each of size 32x32 pixels. The dataset is divided into 50000 training images and 10,000 test images. Each class has 500 training images and 100 test images. Additionally, CIFAR-100 includes labels for twenty super-classes, each grouping together five similar classes, providing a hierarchical structure for more detailed analysis.

IMAGENETTE This dataset is a subset of the larger ImageNet dataset, containing 10 easily classified classes such as tench, English springer, cassette player, chain saw, church, French horn, garbage truck, gas pump, golf ball, and parachute. It consists of 13000 images each with a resolution of 160x160 pixels.

TinyIMAGENET This dataset is a scaled-down version of the ImageNet dataset, containing 200 classes with 500 training images, 50 validation images, and 50 test images per class. The images are resized to 64x64 pixels.

Visual Wake Words (VWW): This dataset is designed for tiny, low-power computer vision models. It contains images labeled with the presence or absence of a person. The

Table 1. Model architectures. For the ConvBlock and ConvDWBlock A,B,C means A means the kernel size, B the number of output channels and C the stride. For Linear Block, A means the number of output neurons.

ID	SmallConv	SmallConvL	VGG8	MobileNetV1
1	ConvBlock 3, 32, 1	ConvBlock 3, 96, 1	ConvBlock 3, 128, 1	ConvBlock 3, 32, 2
2	MaxPool 2, 2	MaxPool 2, 2	ConvBlock 3, 256, 1	ConvDWBlock 3, 64, 1
3	ConvBlock 3, 64, 1	ConvBlock 3, 192, 1	MaxPool 2, 2	ConvDWBlock 3, 128, 2
4	MaxPool 2, 2	MaxPool 2, 2	ConvBlock 3, 256, 1	ConvDWBlock 3, 128, 1
5	ConvBlock 3, 128, 1	ConvBlock 3, 512, 1	ConvBlock 3, 256, 1	ConvDWBlock 3, 256, 2
6	AdaptiveAvgPool (2, 2)	AdaptiveAvgPool (2, 2)	Max Pool 2, 2	ConvDWBlock 3, 256, 1
7	LinearBlock 512	LinearBlock 1024	ConvBlock 3, 512, 1	ConvDWBlock 3, 512, 2
8	-	-	ConvBlock 3, 512, 1	ConvDWBlock 3, 512, 1
9	-	-	AdaptiveAvgPool (2, 2)	ConvDWBlock 3, 512, 1
10	-	-	LinearBlock 1024	ConvDWBlock 3, 512, 1
11	-	-	-	ConvDWBlock 3, 512, 1
12	-	-	-	ConvDWBlock 3, 512, 1
13	-	-	-	ConvDWBlock 3, 1024, 2
14	-	-	-	ConvDWBlock 3, 1024, 1
15	-	-	-	AdaptiveAvgPool (2, 2)

images are resized to 128x128 pixels. The dataset is divided into 115000 training images and 8000 test images.

These datasets provide a diverse range of image classification challenges, facilitating the evaluation of models across various levels of complexity and application scenarios.

1.3. Training Details

All models reported in this work were trained with a batch size of 128 using the Schedule-Free AdamW optimizer [2] with a learning rate of 5×10^{-3}, betas of 0.9 and 0.999, weight decay of 0. For experiments with the MNIST dataset, the data augmentation applied included a random crop transformation with padding 4, followed by a normalization transformation. For FashionMNIST, a similar data augmentation was used, with the addition of a random horizontal flip. Below, we report the specific settings used for particular models.

1.3.1 Experiments with SmallConv and SmallConvL

For experiments with the SmallConv and SmallConvL models, we used light data augmentation for CIFAR10, CIFAR100, and IMAGENETTE. For CIFAR10 and CIFAR100, only a random horizontal flip was applied. For IMAGENETTE, the images were resized to 132x132 pixels and then randomly cropped to 128x128 pixels, followed by a random horizontal flip. The models were trained for 100 epochs for the experiments reported in Table 1 and Table 2.

1.3.2 Experiments with VGG8

We used more extensive data augmentation for experiments with CIFAR10, CIFAR100, IMAGENETTE, and TinyIM-AGENET. The data augmentation consisted of a random crop, followed by a random horizontal flip, then a normalization layer, and a random erasing [10] with a probability of 0.2. When VGG8 was trained on MNIST and FashionM-NIST, the model was trained for 100 epochs. For the other datasets, the model was trained for 300 epochs and dropout layers with a probability of 0.2 were used after each ConvBlock.

1.3.3 Experiments with MobileNetV1

For the experiments with the Visual Wake Words (VWW) dataset, the training images were resized and randomly cropped to a size of 128x128 pixels, followed by normalization. The model was trained for 500 epochs for the experiments reported in Table 3.

1.4. Experimental Compute Resources

All experiments were conducted on a shared internal Linux server equipped with an AMD EPYC 7502 32-Core Processor, 504 GB of RAM, and four NVIDIA A40 GPUs, each with 48 GB of GDDR6 memory. Additionally, code was implemented using Python 3.9 and PyTorch 2.2.1 with CUDA 11.8.

References

[1] Aakanksha Chowdhery, Pete Warden, Jonathon Shlens, Andrew Howard, and Rocky Rhodes. Visual Wake Words Dataset. *arXiv: 1906.05721*, 6 2019. 1

[2] Aaron Defazio, Xingyu Yang, Konstantin Mishchenko, Ashok Cutkosky, Harsh Mehta, and Ahmed Khaled. Schedule-Free Learning - A New Way to Train. https://github.com/facebookresearch/schedule_free, 2024. 2

[3] fast.ai. fastai/imagenette: A smaller subset of 10 easily classified classes from Imagenet, and a little more French, 2021. 1

[4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*, 4 2017. 1

[5] Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. 2009. 1

[6] Ya Le and Xuan S Yang. Tiny ImageNet Visual Recognition Challenge. 2015. 1

[7] Yann LeCun, Corinna Cortes, and C J Burges. MNIST handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010. 1

[8] Arild Nøkland and Lars H Eidnes. Training Neural Networks with Local Error Signals. In *Proceedings of the 36 th International Conference on Machine Learning*, 2019. 1

[9] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 1

[10] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random Erasing Data Augmentation. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, pages 13001–13008, 8 2017. 3