Efficient On-Device Machine Learning with a Biologically-Plausible Forward-Only Algorithm

Baichuan Huang, Amir Aminifar

Department of Electrical and Information Technology, Lund University, Sweden

baichuan.huang@eit.lth.se

Motivation

The training of the state-of-the-art Deep Neural Networks (DNNs) consumes massive amounts of energy, while the human brain learns new tasks with remarkable efficiency. Currently, the training of DNNs **relies almost exclusively on Backpropagation (BP)** [1]. However, BP faces criticism due to **its biologically implausible nature** (such as **weight transport** [2], **non-locality** [3], **update locking** [4], **and frozen activities** [5]), underscoring the significant disparity in performance and energy efficiency between DNNs and the human brain. The biologically plausible forward-only algorithms, without resorting to the BP, is explorated to bridge the existing performance–efficiency gap between the DNNs and the cortex.

Method

In this paper, we propose an efficient on-device learning algorithm, based on the biologically-plausible forward-only algorithm, called Bio-FO. Bio-FO targets the previously mentioned biological implausibility issues, which are only partially solved by the state-of-the-art forward-only algorithms (shown in Fig. 1). Moreover, Bio-FO can be flexibly extended to common networks and relatively large-scale datasets.

Our proposed training scheme performs locally, without the need for non-local information/global error. For each layer's activations, an auxiliary classifier with a fixed random matrix is employed to project the activations to the output logits. Therefore, the weights between each two connected layers are updated as soon as the input to the layer (i.e., the activation of the previous layer) is available. For each layer's weights, a sparsity mask is introduced to allow extensions to common networks such as the fully connected network, the locally connected network, and the convolutional neural network. Our proposed sparse local-training scheme supports parallel and asynchronous updates, incurring resource efficiency.

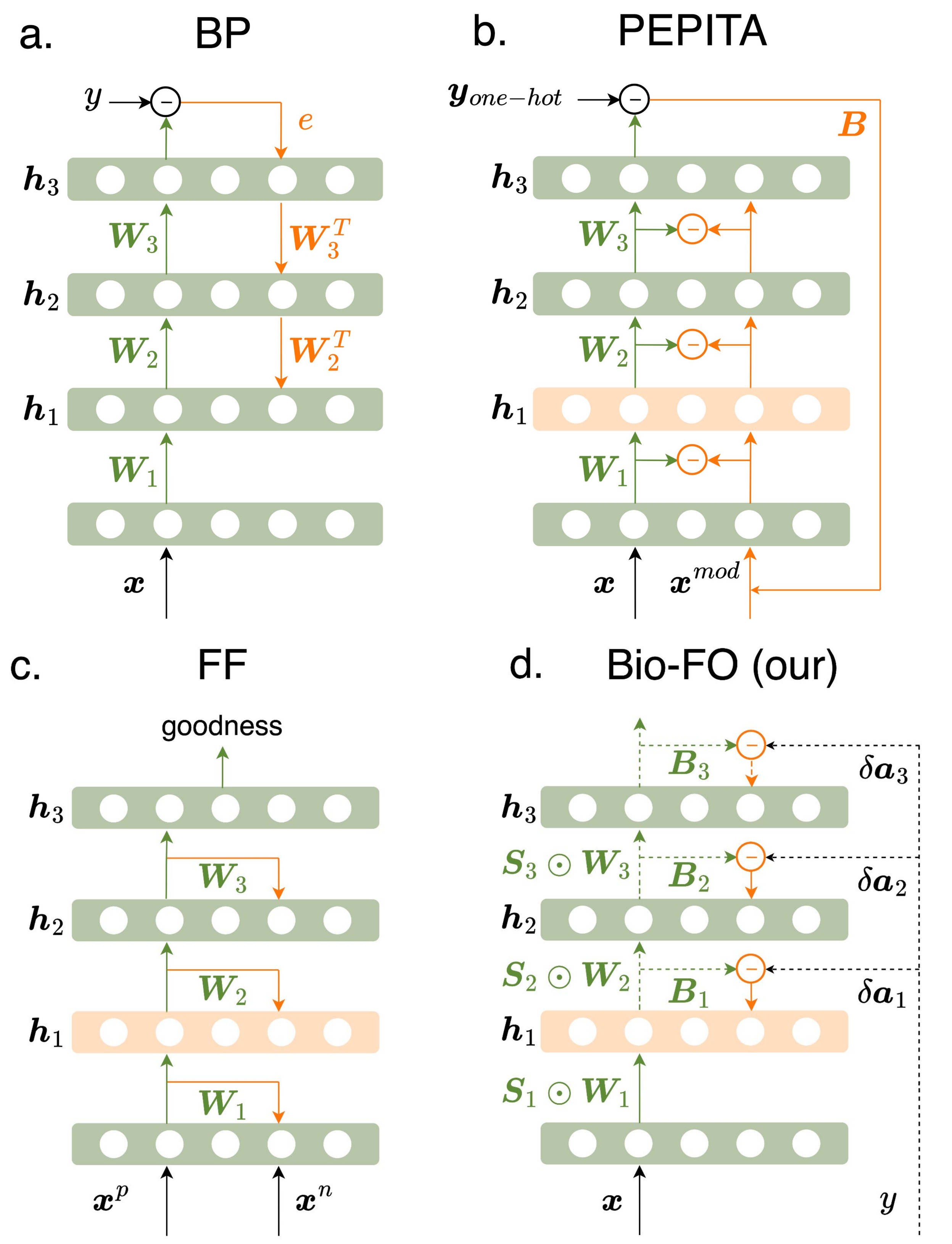

Fig. 1: An overview of various training algorithms.

Results

Bio-FO collectively achieves a lower error compared to DRTP, PEPITA, and FF across different number of layers, for example in Fig. 2 (a). Moreover, as shown in Fig. 2 (b-d), Bio-FO also enjoys faster convergence than PEPITA, and FF.

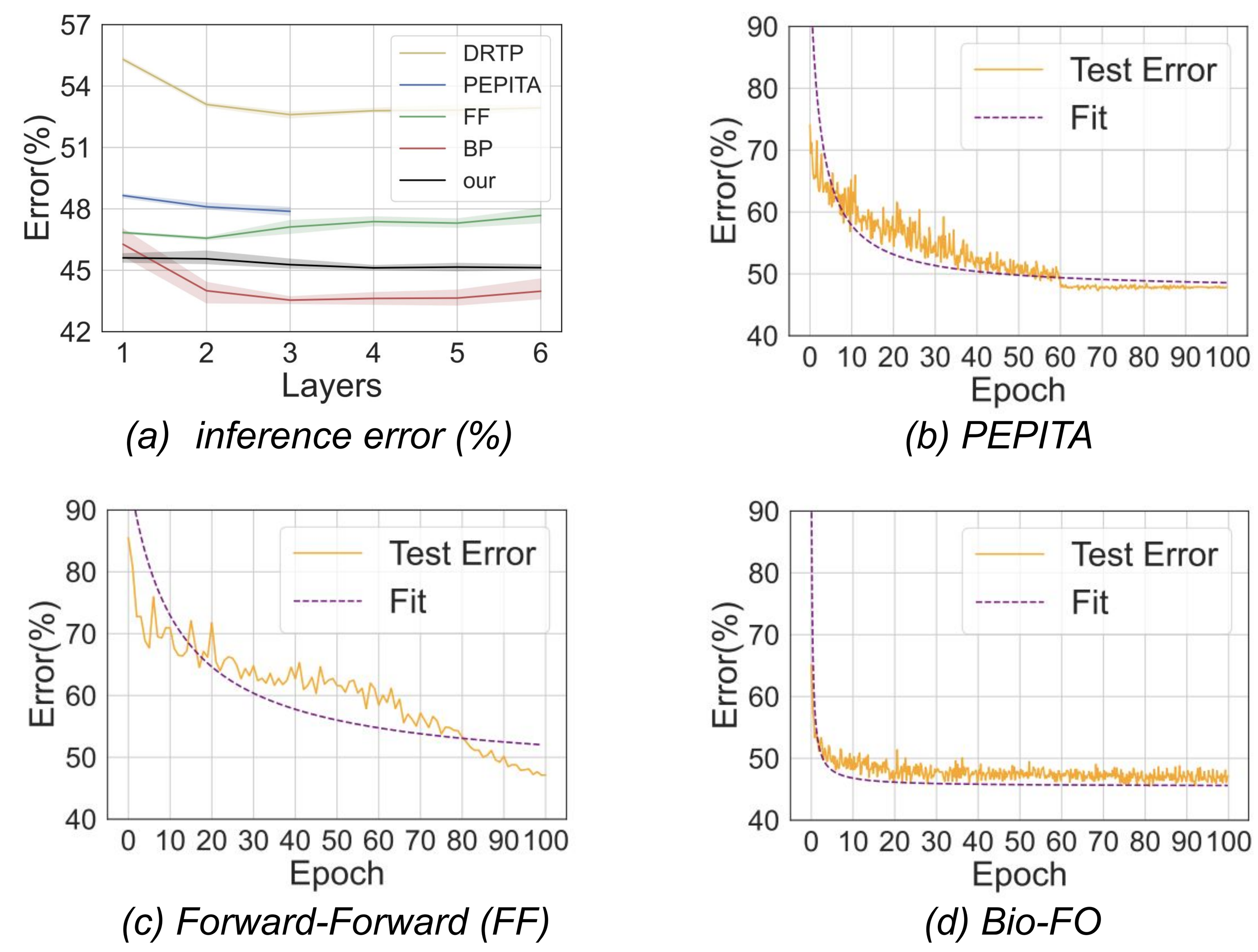

Fig. 2: An overview of results for CIFAR-10

Table 1 presents that Bio-FO outperforms the state-of-the-art forward-only algorithms of DRTP, PEPITA, and FF in terms of energy consumption for training on the NVIDIA Jetson Nano. We highlight the **best** and <u>second best</u> results.

Algorithms	MNIST	CIFAR-10	CIFAR-100	CHB-MIT	MIT-BIH
DRTP	121.6	110.8	131.9	6.4	317.7
PEPITA	**89.9**	<u>91.7</u>	<u>123.9</u>	5.9	<u>191.0</u>
FF	174.4	211.1	753.5	<u>4.8</u>	221.9
Bio-FO	<u>99.8</u>	**83.1**	**37.9**	**3.5**	**121.1**

Table 1: Energy consumption for forward-only algorithms.

References

[1] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. nature, 323(6088):533–536, 1986.

[2] Grossberg, S. Competitive learning: From interactive activation to adaptive resonance. Cognitive science, 11(1): 23–63, 1987.

[3] Whittington, J. C. and Bogacz, R. Theories of error backpropagation in the brain. Trends in cognitive sciences, 23 (3):235–250, 2019.

[4] Jaderberg, M., et al. Decoupled neural interfaces using synthetic gradients. In International conference on machine learning, pp. 1627–1635. PMLR, 2017.

[5] Liao, Q., et al. How important is weight symmetry in backpropagation? In Proceedings of the AAAI Conference on Artificial Intelligence, volume 30, 2016.

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Swedish Research Council

ELLIIT Excellence Center at Linköping – Lund in Information Technology

NAISS