Robust architecture-agnostic and noise resilient training of photonic deep learning models

Supplementary Material

M. Kirtas[1], N. Passalis[1], G. Mourgias-Alexandris[2], G. Dabos[2], N. Pleros[2] and A. Tefas[1]

[1]*Computational Intelligence and Deep Learning Group*

[2]*Wireless and Photonic Systems and Networks Group*

Dept. of Informatics, *Aristotle University of Thessaloniki*, Thessaloniki, Greece

{eakirtas, passalis, mourgias, ntamposg, npleros, tefas}@csd.auth.gr

This document contains supplementary material for the paper "Robust architecture-agnostic and noise resilient training of photonic deep learning models". First, we provide additional information regarding the physical characteristics of the experimental setup used to evaluate the proposed method. Next, we report additional experiments, both regarding the optical fiber communication setup, as well as using an additional convolutional neural network-based setup.

I. Physical Characteristics of Experimental Setup

In the paper, we evaluate the proposed method in two different photonic setups. The first setup, named *Photonic Recurrent Networks for Financial Time Series*, employs a reccurent photonic network that is based on the photonic neuron, presented in the *Background* section of the paper. Detailed description of the photonic configuration is presented in the respective section in the paper. In Table I we also report some additional information on the physical characteristics of the photonic neuron. The information reported in this table refers to the physical characteristics of Mach-Zehnder Modulator (MZM), Photodetector (PD), Analog-to-Digital (ADC) and Digital-to-Analog (DAC) conversion, and lasers.

In the second evaluation experiment, named *Optical Fiber Communication*, we employed the proposed method in an IM/DD system which is trained in an end-to-end fashion as a single feed-forward ANN. The system is composed of a neural transmitter, a channel and a neural receiver. The channel suffers from inter-symbol interference as a result of the fiber dispersion. In Table II, we report the physical characteristics of the channel that combined with the details reported in the paper, describing in detail the actual photonic hardware configuration that was taken into account to evaluate the proposed method. The physical characteristics reported in this table refer to the fiber optics setup, the MZM, and the applied Low Pass Filter (LPF).

TABLE I
PHYSICAL CHARACTERISTICS OF THE PHOTONIC NEURON

Device	**IQ Modulator**
Laser P_{out}	10 dBm
MZM - ER	5.9 dB
MZM - Data Rate	10 Gbaud
PD - Responsivity	0.5 A/W
DAC/ADC resolution	8 bits
Total loss budget	7dB
MZM Bias Voltage	0V
MZM Voltage Swing	5V

TABLE II
PHYSICAL CHARACTERISTICS OF THE FIBER CHANNEL

Oversampling	4
Bits per symbol	6
Sampling rate	336Gsa/s
LPF bandwidth	32GHz - Gaussian Filter, $\sigma = 0.7$
Fiber dispersion	18 ps/nm/km
Fiber attenuation	0.18 dB/km
MZM	$A(t) = sin(t)$
Fiber dispersion	$D(z, \omega) = \exp j(\beta_2/2)\omega^2 z$
Fiber length	30 to 70 km

II. OPTICAL FIBER COMMUNICATION EXPERIMENT

In Figure 1 we report results for the photonic sigmoid based architecture using the Xavier initialization, as well as using the Xavier initialization combined with the proposed method for fiber lengths that are different from those used in the training phase. As a result, the network is trained with a different amount of noise compared to the noise experienced during inference. Therefore, these experiments investigate the effect of noise when the models are evaluated in different conditions compared to those used during the training. It should be noted that even with small changes in length (e.g., $\pm$ 1km) the system performance has significantly deteriorated. The evaluation results also indicate that the proposed method leads to models that are more robust to noise, especially when noise with similar characteristics is encountered during the inference, i.e., when the system is evaluated on fiber lengths close to the trained one ($< \pm 3$km).

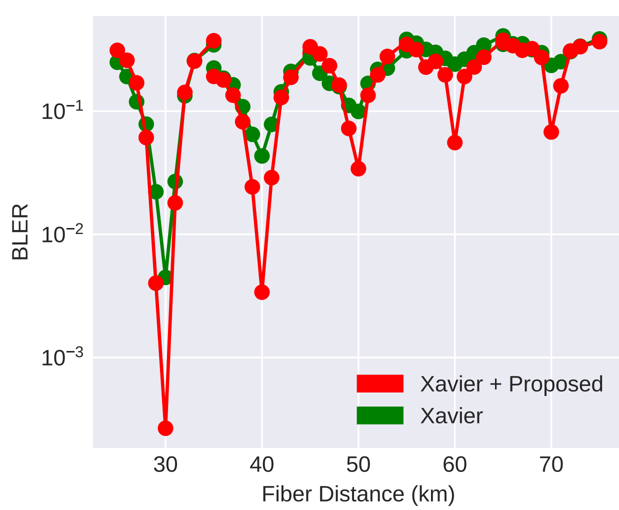

Fig. 1. Comparing the Xavier initialization method and the proposed one for different transmission distances

III. Convolutional Networks for Image Classification

We also conducted additional experiments to evaluate the effect of using the proposed method on larger convolutional architectures in order to examine the scalability of the proposed method. Therefore, we evaluated the proposed method on two additional image datasets: FashionMNIST [1] and CIFAR10 [2]. The convolutional neural network used is composed of 4 convolutional layers followed by 2 fully connected layers. More precisely, two convolutional layers with 32 and 64 filters of size 3×3 followed by a 2×2 average pooling layer are used. Next, another two convolutional layers are employed with 128 and 256 filters of size 3×3 and a final 2×2 average pooling layer. Then, the extracted feature map is flattened and fed to a hidden layer with 512 neurons, before forwarding it to the final classification layer. This led to a network with about 2.5 million parameters for the FashionMNIST dataset and 3.7 million parameters for the CIFAR10 dataset. Optimization was carried out for 50 epochs with a learning rate of $\eta = 10^{-4}$ using Adam optimizer, while the networks were further fine-tuned for another 50 epochs with a reduced learning rate of $\eta = 10^{-5}$. The proposed initialization method ran for $N_{init} = 5$ epochs with the initial learning rate set to $\eta = 0.1$. Four different initialization approaches were evaluated: a) Xavier initialization, b) Xavier initialization along with proposed method, c) He initialization and d) He initialization along with proposed method. Three different activation functions were used for the conducted experiments: a) (regular) sigmoid, b) photonic sigmoid, and c) photonic sinusoidal. Additionally, we evaluated the methods on different noise levels that were applied on weights and inputs of every layer. All models were trained with AGWN noise drawn from $\mathcal{N}(0, \sigma^2)$, where σ ranges from 0 to 0.2.

The training accuracy after the training process is reported for the different initialization schemes and noise levels on the FashionMNIST and CIFAR10 datasets in Tables III and IV, respectively. On the FashionMNIST dataset we observe that the proposed method results in a stable performance for all different activation functions, in contrast to traditional initialization methods that are significantly deteriorated on the cases of sigmoid and photonic sigmoid. This is more clear in the CIFAR10 dataset where the performance of traditional initialization methods collapses in those cases. Indeed, we observed that in most of the cases the training process halted and it could not progress even after 100 epochs of training. Additionally, even in cases for which the traditional methods achieve acceptable performance, e.g., photonic sinusoidal activation, the training accuracy again collapses when higher noise levels are applied. On the other hand, the proposed method outperforms all the traditional initialization schemes, regardless of the used initialization scheme and/or level of noise. These experimental results confirm the scalability of the proposed method, enabling us to employ it in deep, easily saturated, and noisy architectures. Note that even though these networks are beyond the current capacity of existing neuromorphic hardware, the architecture employed holds the credentials of being integrated into a single photonic chip by utilizing components and circuits already offered by CMOS-compatible Photonic Integrated Circuit (PIC) technology, allowing one to implement significantly larger networks [3].

References

[1] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms,” 2017.

[2] A. Krizhevsky, “Learning multiple layers of features from tiny images,” Tech. Rep., 2009.

TABLE III

FashionMNIST - Classification Error (%) for Different Activation Functions and Noise Levels

Activation Function	Initialization	Noise Level 0.0	0.05	0.1	0.2
Sigmoid	Xavier	90.0	58.46	89.65	90.08
	+ Proposed	**8.74**	**9.11**	**11.14**	**15.29**
	He	13.67	14.09	15.42	18.66
	+ Proposed	**8.81**	**10.26**	**12.37**	**16.18**
Photonic Sigmoid	Xavier	17.88	23.31	29.36	89.57
	+ Proposed	**8.63**	**12.4**	**19.21**	**25.16**
	He	19.17	23.45	28.29	29.23
	+ Proposed	**9.06**	**12.81**	**21.53**	**26.73**
Photonic Sinusoidal	Xavier	8.04	9.4	11.85	28.83
	+ Proposed	**7.65**	**8.82**	**10.48**	**16.75**
	He	9.16	10.22	14.51	24.38
	+ Proposed	**8.62**	**9.54**	**12.0**	**16.03**

TABLE IV

CIFAR10 - Classification Error (%) for Different Activation Functions and Noise Levels

Activation Function	Initialization	Noise Level 0.0	0.05	0.1	0.2
Sigmoid	Xavier	90.0	89.97	89.53	90.03
	+ Proposed	**20.54**	**25.93**	**25.75**	**35.64**
	He	90.0	90.63	89.98	90.43
	+ Proposed	**23.81**	**25.33**	**32.21**	**42.45**
Photonic Sigmoid	Xavier	90.0	89.77	89.9	90.33
	+ Proposed	**20.52**	**29.4**	**37.33**	**56.01**
	He	74.17	89.54	90.14	90.0
	+ Proposed	**25.18**	**28.23**	**39.94**	**89.82**
Photonic Sinusoidal	Xavier	18.62	21.06	28.2	74.23
	+ Proposed	**16.36**	**19.47**	**24.97**	**41.55**
	He	52.28	43.92	56.3	90.3
	+ Proposed	**21.67**	**21.74**	**28.47**	**48.85**

[3] C. Huang, S. Fujisawa, T. F. de Lima, A. N. Tait, E. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H. G. Batshon, H.-T. Peng, B. J. Shastri, Y. Inada, T. Wang, and P. R. Prucnal, "Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems," in *2020 Optical Fiber Communications Conf. and Exhibition (OFC)*, 2020, pp. 1–3.