Noise-Resistant Optical Implementation of Analogue Neural Networks

Diego Argüello Ron, Morteza Kamalian-Kopae, and Sergei Turitsyn
Aston Institute of Photonic Technologies, Aston University, B4 7ET, Birmingham, United Kingdom

Analogue artificial neural networks are widely considered as promising computational models that more closely imitate the information processing capabilities of the human brain compared to digital neural networks. The significant computation power and the much reduced power consumption per operation make the analogue implementation of neural networks very attractive. There is an active research on artificial neural networks (ANNs) implementation using both analogue photonic and electronic hardware [1–4]. However, compared to digital realisations the conventional analogue systems are more sensitive to the noise that is inevitably present in practical implementations [2, 3]. Noise properties in ANNs have been studied both in the electronic and photonic domains. However, photonic ANNs are much less investigated compared to the electronic implementations, for which some training techniques have been proposed to enhance ANNs robustness against noise [1, 4].

In this work we examine photonic ANNs and demonstrate that some known methods proposed to train noise-resilient electronic ANNs are not optimal for photonic implementations. More specifically, we reveal that the noise injection combined with the gradient clipping training approach which is shown to be effective in electronic-based hardware, is not efficient against the noise in the photonic domain. We also examine the performance of feed-forward neural networks (FNNs) and also explore how the noise propagation through recurrent neural networks (RNNs) can be adopted to reduce detrimental impact of noise.

Finally, we propose and optimised noise injection training method for analogue photonic ANNs. Both a Multilayer Perceptron (MLP) and a RNN have been trained on MINIST dataset using two approaches to noise injection, using analogue photonic noise features described in [5]. In the first one, the weight W_l of layer l is drawn from a Gaussian distribution $\mathbf{N}(\mathbf{W}_l; \mathbf{W}_0^l, \sigma_{N,l}^2 \mathbf{I})$. The noise is reference to the range of representable weights, $\sigma_{N,l}^2 = \eta(W_{max}^l - W_{min}^l)$. The values W_{max}^l and W_{min}^l are obtained based on the statistics of the weights. The hyperparameter η is a coefficient characterising the noise and the range of the weight values is reduced by clipping them to the interval $[W_{min}^l; W_{max}^l]$. The second noise injection training method that we study, defines the weight W^l of layer l from a Gaussian distribution $\mathbf{N}(\mathbf{W}_l; 0, \sigma^2 \mathbf{I})$.

In Table. 1 the performance of the two networks trained using both methods are illustrated. For these results an optimisation process has been carried out to find different optimal values of the parameters η and σ^2. Comparison to the performance of the model without noise injection during training, used as a baseline, makes it clear that the feed-forward nature of the MLP improves its resilience to noise, and that noise injection always upgrades this resilience. For the case of the MLP, it is shown that the best obtained accuracy is 89.3 % for $\eta = 8 \times 10^{-4}$, in the case of the weight clipping approach, and 89.5 % for $\sigma^2 = 10^{-5}$ when using the alternative approach. It is seen that the performance achieved by these two methods is quite similar. In the case of the RNN we demonstrate that the second approach achieves a better performance than the weight clipping method, as the achieved accuracies are 85.6% and 83.5%, respectively.

Table 1: Performance of a RNN on MINIST trained using different methods while suffering from analog non-idealities caused by pothonic noise.

Method	MLP best accuracy values (%)	RNN best accuracy values (%)
Gaussian Noise Injection with Weight Clipping	89.3 ($\eta = 8 \times 10^{-4}$)	83.5 ($\eta = 1 \times 10^{-3}$)
Gaussian Noise Injection	89.5 ($\sigma^2 = 1 \times 10^{-5}$)	85.6 ($\sigma^2 = 2 \times 10^{-3}$)
Standard Training	89	45.5

In conclusion, we demonstrate that noise injection is a promising technique that makes photonic analogue ANNs more resilient to noise. Nevertheless, further optimisation is required in order to make it an efficient method for different types of network architectures and hardware implementations.

References

[1] N. Semenova, X. Porte, L. Andreoli, M. Jacquot, L. Larger, and D. Brunner, "Fundamental aspects of noise inanalog-hardware neural networks,"Chaos **29** (10), (2019).

[2] B. J. Shastri, A. N. Tait, T. F. de Lima, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Principles of neuro-morphic photonics,"arXiv preprint arXiv:1801.00016, 2017.

[3] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics,"Light Sci. Appl. **8** (1), (2019).

[4] C. Zhou, P. Kadambi, M. Mattina, and P. N. Whatmough, "Noisy machines: Understanding noisyneural networks and enhancing robustness to analog hardware errors using distillation,"arXiv preprintarXiv:2001.04974, 2020.

[5] K.-C. Jim, C. L. Giles, and B. G. Horne, "An analysis of noise in recurrent neural networks: convergence and generalization,"IEEE Trans. Neural Networks **7** (6), (1996)