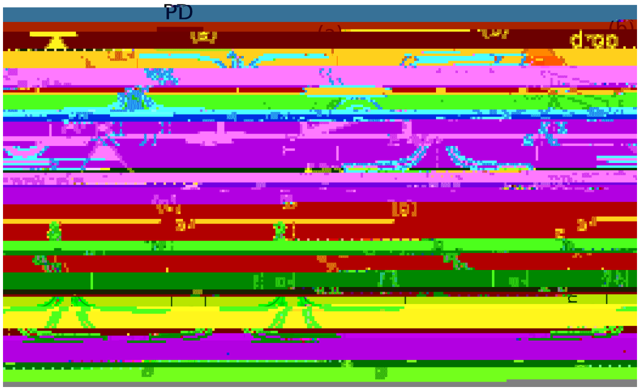
Digital Electronics and Analog Photonics for Convolutional Neural Networks (DEAP-CNNs)

Viraj Bangari, Bicky A. Marquez, *Member, IEEE*, Heidi Miller, Alexander N. Tait, *Member, IEEE*, Mitchell A. Nahmias, Thomas Ferreira de Lima, *Student Member, IEEE*, Hsuan-Tung Peng, Paul R. Prucnal, *Life Fellow, IEEE*, and Bhavin J. Shastri, *Senior Member, IEEE*

***Abstract*—Convolutional Neural Networks (CNNs) are powerful and highly ubiquitous tools for extracting features from large datasets for applications such as computer vision and natural language processing. However, a convolution is a computationally expensive operation in digital electronics. In contrast, neuromorphic photonic systems, which have experienced a recent surge of interest over the last few years, propose higher bandwidth and energy efficiencies for neural network training and inference. Neuromorphic photonics exploits the advantages of optical electronics, including the ease of analog processing, and busing multiple signals on a single waveguide at the speed of light. Here, we propose a Digital Electronic and Analog Photonic (DEAP) CNN [illegible]**

TABLE I
SUMMARY OF CONVOLUTIONAL PARAMETERS


PD
drop

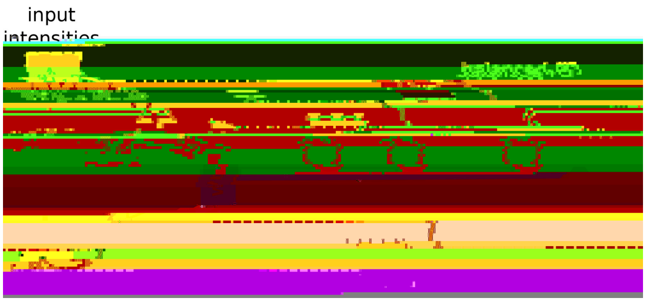


Fig. 4. An electro-optic architecture that performs dot products. A_i ($= 1, \ldots,$) are input elements encoded in intensities, multiplexed by a WDM and linked to the weight banks via a silicon waveguide. F_i are filter values that modulate the MRRs in the photonic weight bank. Drop and through output ports are connected to a balanced PD, where the matrix multiplication is performed, followed by an amplifier TIA.

we can set a particular weight using:

$$F^{*} = 2T\ (\phi\ ) - 1$$

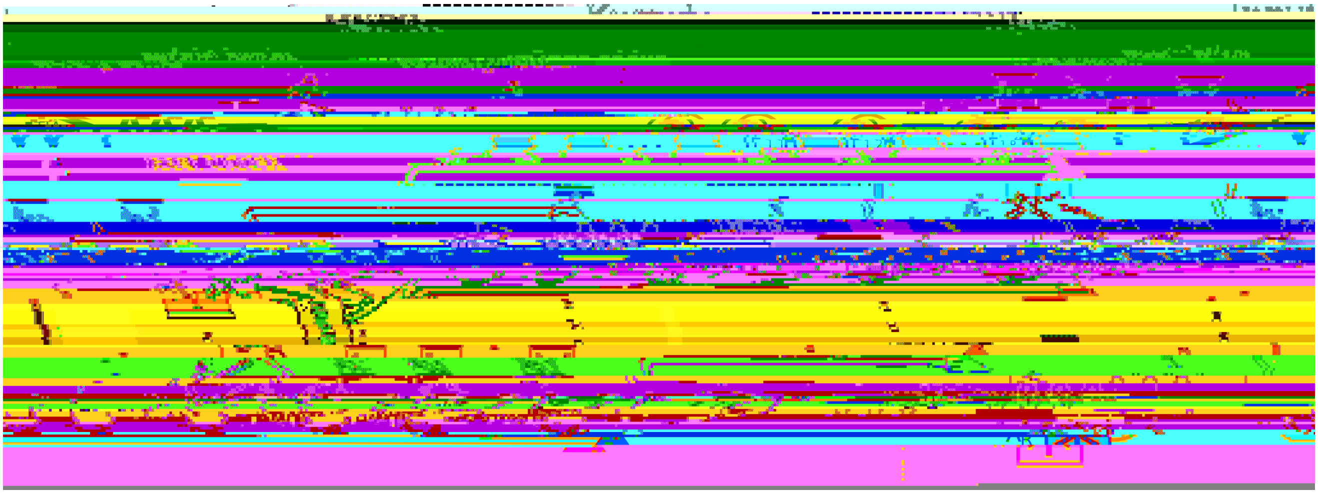


Fig. 5. Photonic architecture for producing a single convolved pixel. Input images are encoded in intensities $A_{l,h}$, where the pixel inputs $A_{m,n,k}$ with $\in [\ ,\ +\ _m]$, $\in [\ ,\ +\ _m]$, $\in [1, D_m]$ are represented as $A_{l,h}$, $= 1, \ldots, D$ and $= 1, \ldots,\ ^2$. Considering the boundary parameters, we set $D = D_m$ and $=\ _m$. Likewise, the filter values $F_{m,n,k}$ are represented as are represented as $F_{l,h}$ under the same conditions. We use an array of

TABLE II
BENCHMARKING PARAMETERS FOR DEAP

W	

their errors are deterministic and predictable. On the other hand, the errors from photonics are due to stochastic shot, spectral, Johnson-Nyquist and flicker noises, as well as quantization noise in the ADC, and distortion from the RF signals applied to the modulators. However, artificially adding random noise to CNNs have been shown to reduce over-fitting [52], meaning that some degree of stochastic behaviour is tolerable in the domain of machine learning problems.

Finally, MRRs have only been shown to have up to 6-bits of precision, which is significantly smaller than the range precision supported by even half-precision (16-bit) floating point representations. In conclusion, photonics has the potential to perform convolutions at speeds faster than top-of-the-line GPUs while having a lower energy consumption. Moving forward, the greatest challenges to overcome have to do with increasing the precision of photonic components so that they are comparable to classical floating-point representations. Overall, silicon photonics has the potential to outperform conventional electronic hardware for convolutions while having the ability to scale up in the future.

Acknowledgment

The authors thank Prof. Sudip Shekhar, Mohammed Al-Qadasi, Hugh Morison and Matthew Filipovich for edits and suggestions.

References

[1] X. Li, G. Zhang, H. H. Huang, Z. Wang, and W. Zheng, “Performance analysis of GPU-based convolutional neural networks,” in *Proc. 45th Int. Conf. Parallel Process.*, Aug. 2016, pp. 67–76.

[2] M. Jaderberg, A. Vedaldi, and A. Zisserman, “Speeding up convolutional neural networks with low rank expansions,” 2014, *arXiv:1405.3866*.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[4] Y. LeCun and C. Cortes, “MNIST handwritten digit database,” 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[5] F. Duport, A. Smerieri, A. Akrout, M. Haelterman, and S. Massar, “Fully analogue photonic reservoir computer,” *Scientific Rep.*, vol. 6, 2016, Art. no. 22381. [Online]. Available: http://dx.doi.org/10.1038/srep22381

[6] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, “Parallel photonic information processing at gigabyte per second data rates using transient states,” *Nature Commun.*, vol. 4, 2013, Art. no. 1364. [Online]. Available: http://dx.doi.org/10.1038/ncomms2368

[7] K. Vandoorne *et al.*, “Experimental demonstration of reservoir computing on a silicon photonics chip,” *Nature Commun.*, vol. 5, 2014, Art. No. 3541. [Online]. Available: http://dx.doi.org/10.1038/ncomms4541

[8] L. Larger *et al.*, “Photonic information processing beyond turing: An optoelectronic implementation of reservoir computing,” *Opt. Express*, vol. 20, no. 3, pp. 3241–3249, Jan. 2012. [Online]. Available: http://dx.doi.org/10.1364/OE.20.003241

[9] P. R. Prucnal and B. J. Shastri, *Neuromorphic Photonics*. Boca Raton, FL, USA: CRC Press, 2017.

[10] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, “Broadcast and weight: An integrated network for scalable photonic spike processing,” *J. Lightw. Technol.*, vol. 32, no. 21, pp. 4029–4041, Nov. 2014.

[11] A. N. Tait *et al.*, “Microring weight banks,” *IEEE J. Sel. Topics Quantum Electron.*, vol. 22, no. 6, pp. 312–325, Nov. 2016.

[12] A. N. Tait *et al.*, “A silicon photonic modulator neuron,” *Phys. Rev. Applied*, vol. 11, no. 6, p. 064043, Jun. 2019.

[13] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, “Training of photonic neural networks through in situ backpropagation and gradient measurement,” *Optica*, vol. 5, no. 7, pp. 864–871, Jul. 2018, doi: 10.1364/OPTICA.5.000864.

[14] Y. Shen *et al.*, “Deep learning with coherent nanophotonic circuits,” *Nature Photon.*, vol. 11, pp. 441–446, Jun. 2017. [Online]. Available: http://dx.doi.org/10.1038/nphoton.2017.93

[15] T. F. De Lima, *et al.* “Machine learning with neuromorphic photonics,” *J. Lightw. Technol.*, vol. 37, no. 5, pp. 1515–1534, Mar. 2019.

[16] H. Bagherian *et al.*, “On-chip optical convolutional neural networks,” 2018, *arXiv:1808.03303*. [Online]. Available: http://arxiv.org/abs/1808.03303

[17] A. Mehrabian, Y. Al-Kabani, V. J. Sorger, and T. A. El-Ghazawi, “PCNNA: A photonic convolutional neural network accelerator,” in *Proc. 31st IEEE Int. System-on-Chip Conf. (SOCC)*, Arlington, VA, USA, 2018, pp. 169–173, doi: 10.1109/SOCC.2018.8618542.

[18] W. Liu *et al.*, “Holylight: A nanophotonic accelerator for deep learning

Paul R. Prucnal