PHOTONIC NEURAL NETWORKS

Competitive photonic neural networks

Photonics offers high hopes for next-generation neural network processors. Now it has been shown that even entirely using off-the-shelf photonics allows surpassing speed and energy efficiency of cutting-edge GPUs.

Daniel Brunner, Demetri Psaltis

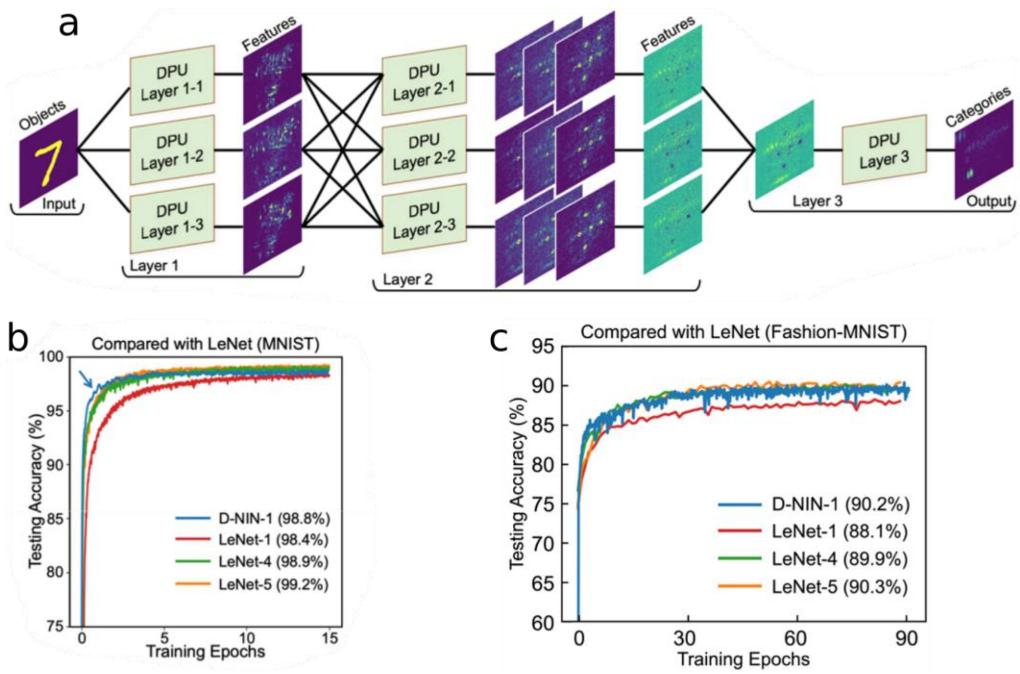

Figure 1| Optical neural network computing. a, Zhou *et al.* implemented a reconfigurable photonic neural network using off-the-shelf components that is superior to cutting edge GPUs in speed and energy efficiency. **b,c,** Crucially, the inference accuracy of their system can compete with that of previous breakthrough NN architectures emulated on classical computers.

Neural networks (NNs) are a neuro-inspired concept that realizes computation based on the collective response of nonlinear elements (the neurons). Through this NNs mimic one of the most elementary aspects of computing in biological brains. A particular information processing task is 'programmed' by adjusting the network's topology, i.e. learning, usually on the basis of statistical optimization leveraging examples. NNs can learn to tell cats from dogs – or to carry out much more serious and useful feats and represent one of computing's most exciting recent developments. However, each neuron's state needs to be calculated across the network's connections, and the associated computational-toll exceedingly brings classical processors to their knees.

Instead of emulating NNs with digital computers, substantial research-efforts are directed towards hardware whose governing laws of physics mimic a specific NN concept[1]. A major focus lies on physically realizing a NN's connections, and photonic parallelism holds promises for substantially faster and more efficient NN processors[2].

Writing in *Nature Photonics*, Zhou *et al.* demonstrate that even photonic NNs realized with non-specialized and off-the-shelf components can outperform top-notch GPUs specifically tailored for NN applications[3].Their photonic system is faster, more energy efficient, programmable and rivals the accuracy of competitive digital NN benchmark models.

In the work by Zhou *et al.*, a single neuron's state is the reflection of a mirror inside a digital micro-mirror device (DMD), illuminated by a solid state laser. Diffraction of the DMD's signal by a phase-mask displayed on a liquid-crystal spatial light modulator (SLM) creates parallel and reconfigurable network connections, and the modulus-square of optical detection via a fast camera adds nonlinearity to each NN layer. Such cascaded nonlinear operations strongly amplify the dimensionality of data representation, which ultimately is what allows NNs to unearth concealed features they can then leverage for challenging computations. The interconnections they implement are highly constrained, yet they achieve competitive performance by cascading multiple layers. They calculate that their diffractive processing unit (DPU) achieves 240.1 TOPs/s with an energy efficiency of 1.578 TOPs/J. All components and devices involved in running the system are included in this budget. Remarkably, in both metrics the authors beat Nvidia's top of the line Tesla V100 tensor core GPU.

A further substantial advance is the accuracy the authors achieve. Let us use the example of MNIST digit recognition and a 3-layer deep NN implemented by the DPU through temporal multiplexing. An initial simulation via a physics model achieves 97.6% testing accuracy. However, that significantly dropped to a bit above chance when the pre-optimized network configurations were transferred to the physical DPU. The authors continued to optimize the SLM's phase mask through iterative experimental updates until they experimentally obtain 96.2% testing accuracy after 15 training epochs. A more complex DPU temporal multiplexing scheme (D-NIN-1(++)), see Fig. 1a, implements a convolutional NN that with 99% testing accuracy in MNIST outperforms the LeNet-4 architecture's 98.9% accuracy, Fig. 1b. This competitive performance is confirmed in further tests such as the fashion MNIST, Fig. 1c, as well as human action recognition.

The work of Zhou *et al.* has many implications. The DPU physically implements network connections exploiting the parallelism of photonics. It has recently been shown in random recurrent[4] as well as deep linear networks[5] with fixed topologies that the photonic approach is competitive and enables GPU-superior scaling. Zhou *et al.* quantify this advantage for a range of more general topologies, include learning and confirm that photonic NNs can compete with similar NN models run on GPUs in terms of inference accuracy in several present-day benchmark data sets. Furthermore, the concept is size-scalable: neurons are implemented in 2D-planes while connections leverage optical propagation along the third dimension, and 3D makes the physical footprint scale linear when augmenting the number of neurons[6].

We expect that the intense current interest in optics for machine learning[7–9] is only the beginning. Most significantly, the implementation of the nonlinearity in optical neural networks has been done electronically via the electronic-to-optical-to-electronic transduction. A future challenge is to implement the nonlinearity optically, avoiding the bottlenecks due to the conversion. This promises to drastically improves the performance[10,11]. In-situ learning in the optical domain is another open issue that needs to be addressed and demonstrated in hardware.

The authors declare the following competing interests:
The authors declare no competing interests.

Daniel Brunner it in the FEMTO-ST/Optics Dept., UMR CNRS 6174, Univ. Bourgogne Franche-Comté, Besançon, France. Demetri Psaltis is at the Optics Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

References

1. Marković, D., Mizrahi, A., Querlioz, D. & Grollier, J. *Nat. Rev. Phys.* **2**, 499–510 (2020).
2. Psaltis, D., Brady, D., Gu, X.-G. & Lin, S. *Nature* **343**, 325–330 (1990).

3. Zhou, T. *et al. Nat. Photonics* **15**, 367-373 (2021).
4. Rafayelyan, M., Dong, J., Tan, Y., Krzakala, F. & Gigan, S. *Phys. Rev. X* **10**, 41037 (2020).
5. Lin, X. *et al. Science (80-.).* **361**, 1004–1008 (2018).
6. Dinc, N. U., Psaltis, D. & Brunner, D. *Photoniques* **114**, 34–38 (2020).
7. Wetzstein, G. *et al. Nature* **588**, 39–47 (2020).
8. Xu, X. *et al. Nature* **589**, 44–51 (2021).
9. Feldmann, J. *et al. Nature* **589**, 52 (2021).
10. Teğin, U., Yıldırım, M., Oğuz, İ., Moser, C. & Psaltis, D. *arXiv:2012.12404* (2020).
11. Porte, X. *et al. arXiv:2012.11153* 1–16 (2020).