Researchers develop energy-efficient optical neural networks

August 6 2024, by Celia Luterbacher

$$E_{out} = HT_{Ln}(\mathbf{x}) \ldots HT_{L2}(\mathbf{x})\, HT_{L1}(\mathbf{x})\, E_{il}$$

$$T_{Ln} = e^{j(\mathbf{x} \odot \mathbf{s}_{Ln} + \mathbf{b}_{Ln})}$$
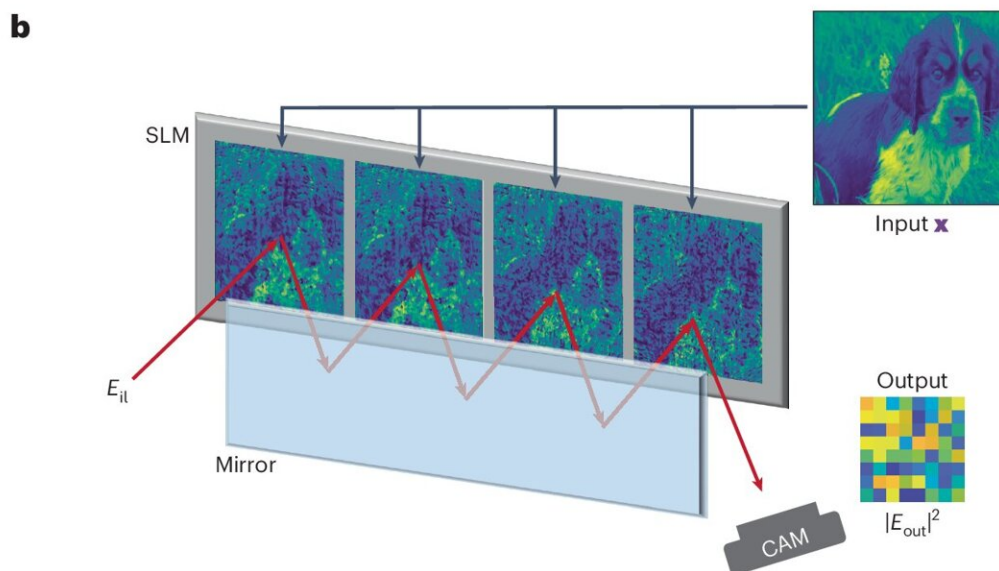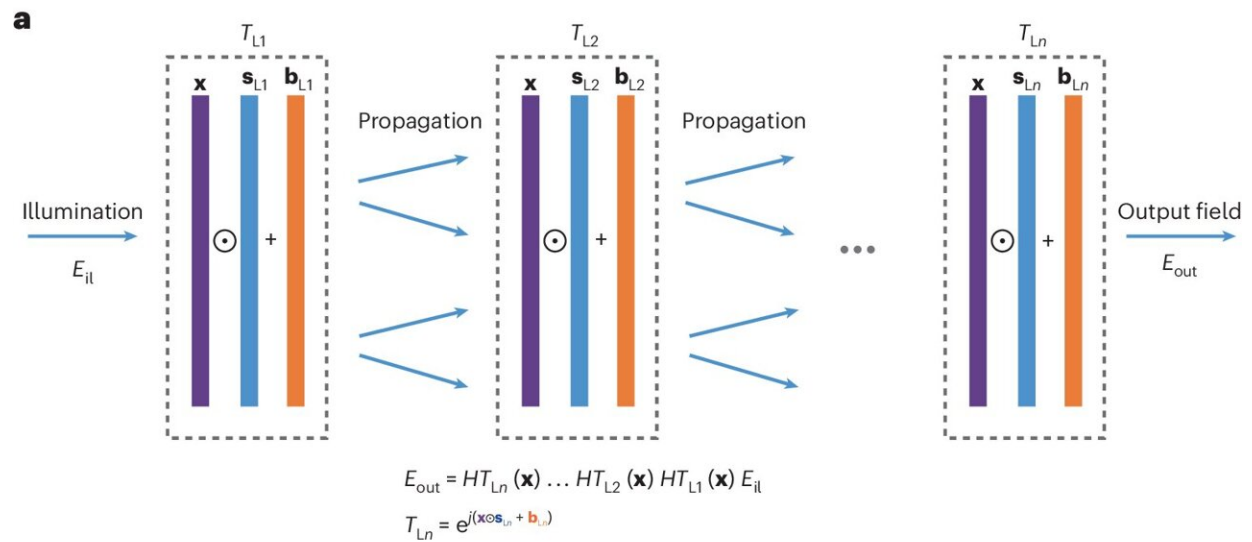
The nPOLO framework. Credit: *Nature Photonics* (2024). DOI:

EPFL researchers have published a programmable framework that overcomes a key computational bottleneck of optics-based artificial intelligence systems. In a series of image classification experiments, they used scattered light from a low-power laser to perform accurate, scalable computations using a fraction of the energy of electronics.

As digital artificial intelligence systems grow in size and impact, so does the energy required to train and deploy them—not to mention the associated carbon emissions. Recent research suggests that if current AI server production continues at its current pace, their annual energy consumption could outstrip that of a small country by 2027.

Deep neural networks, inspired by the architecture of the human brain, are especially power-hungry due to the millions or even billions of connections between multiple layers of neuron-like processors.

To counteract this mushrooming energy demand, researchers have doubled down on efforts to implement optical computing systems, which have existed experimentally since the 1980s. These systems rely on photons to process data, and although light can theoretically be used to perform computations much faster and more efficiently than electrons, a key challenge has hindered optical systems' ability to surpass the electronic state-of-the art.

"Our method is up to 1,000 times more power-efficient than state-of-the-art deep digital networks, making it a promising platform for realizing optical neural networks," says Demetri Psaltis.

"In order to classify data in a neural network, each node, or 'neuron,'

must make a 'decision' to fire or not based on weighted input data. This decision leads to what is known as a nonlinear transformation of the data, meaning the output is not directly proportional to the input," says Christophe Moser, head of the Laboratory of Applied Photonics Devices in EPFL's School of Engineering.

Moser explains that while digital neural networks can easily perform nonlinear transformations with transistors, in optical systems, this step requires very powerful lasers.

Moser worked with students Mustafa Yildirim, Niyazi Ulas Dinc, and Ilker Oguz, as well as Optics Laboratory head Demetri Psaltis, to develop an energy-efficient method for performing these nonlinear computations optically.

Their new approach involves encoding data, like the pixels of an image, in the spatial modulation of a low-power laser beam. The beam reflects back on itself several times, leading to a nonlinear multiplication of the pixels.

"Our image classification experiments on three different datasets showed that our method is scalable, and up to 1,000 times more power-efficient than state-of-the-art deep digital networks, making it a promising platform for realizing optical neural networks," says Psaltis.

The research has recently been published in *Nature Photonics*.
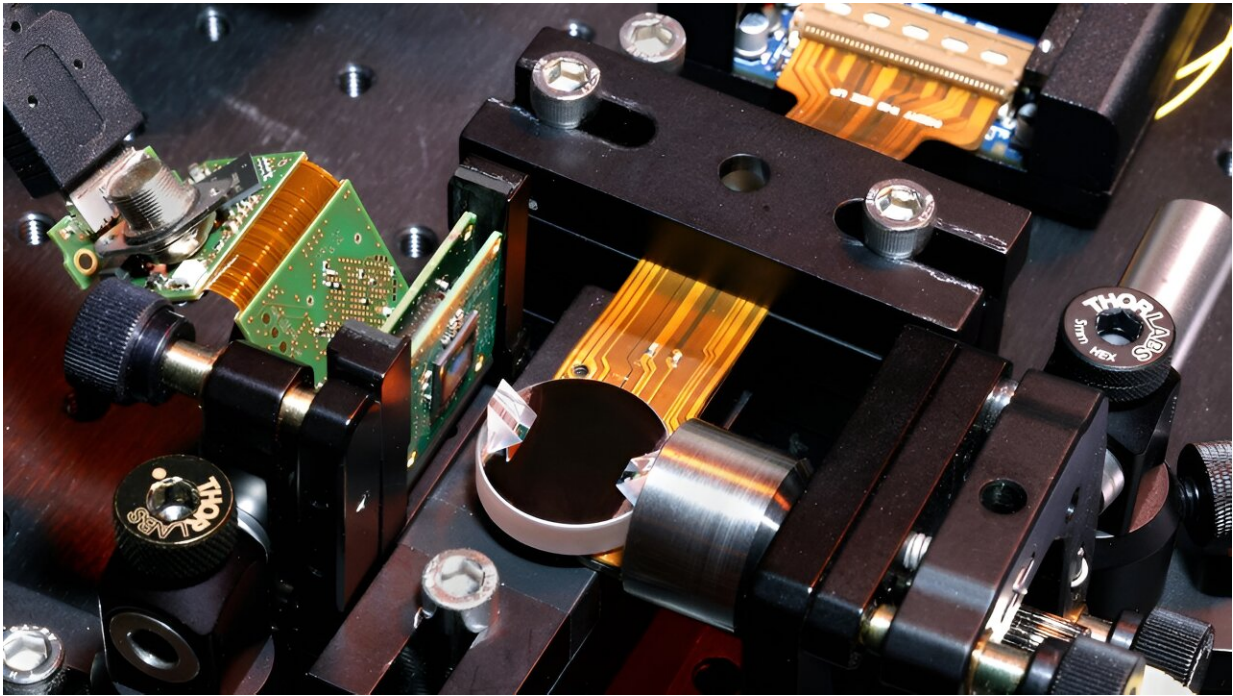
The core of the optical processor. Credit: Alain Herzog/Ecole Polytechnique Federale de Lausanne

A simple structural solution

In nature, photons do not directly interact with each other the way charged electrons do. To achieve nonlinear transformations in optical systems, scientists have therefore had to "force" photons to interact indirectly, for example by using a light intense enough to modify the optical properties of the glass or other material it passes through.

The scientists worked around this need for a high-power laser with an elegantly simple solution: they encoded the pixels of an image spatially on the surface of a low-power laser beam. By performing this encoding twice, via adjustment of the trajectory of the beam in the encoder, the pixels are multiplied by themselves, i.e., squared.

Since squaring is a non-linear transformation, this structural modification achieves the non-linearity essential to neural network calculations, at a fraction of the energy cost. This encoding can be carried out two, three or even ten times, increasing the non-linearity of the transformation and the precision of the calculation.

"We estimate that using our system, the energy required to optically compute a multiplication is eight orders of magnitude less than that required for an electronic system," Psaltis says.

Moser and Psaltis emphasize that the scalability of their low-energy approach is a major advantage, as the ultimate goal would be to use hybrid electronic–optical systems to mitigate the energy consumption of digital neural networks.

However, further engineering research is needed to achieve such scale-up. For example, because optical systems use different hardware than electronic systems, a next step that the researchers are already working on is developing a compiler to translate digital data into code that optical systems can use.

  More information: Mustafa Yildirim et al, Nonlinear processing with linear optics, *Nature Photonics* (2024). DOI: 10.1038/s41566-024-01494-z

Provided by Ecole Polytechnique Federale de Lausanne