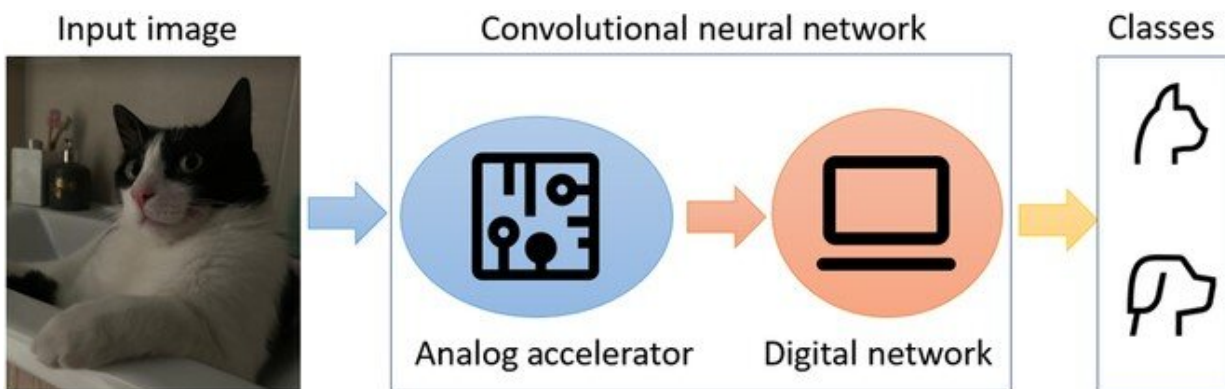# How photonics is revolutionizing convolutional neural networks

July 5 2023

The first neural layers are implemented with an analog accelerator (electronic, photonic, etc.) with low power consumption and high processing bandwidth, whereas the last layers are implemented digitally with high precision and low complexity. Credit: Aris Tsirigotis et al.

In the modern era, there has been explosive growth in the demand for computing power for cognitive image and video processing. While convolutional neural networks offer improved performance for image processing, they also come with a significant increase in power consumption and memory requirements.

According to a survey conducted by OpenAI, the period between 2012 and 2018 witnessed a staggering increase in the number of computations by a factor larger than 300,000. In contrast, during the same period,

Moore's law only accounted for a seven-fold increase in computational power.

To address this escalating need for computing power, researchers have turned to photonics as a means to enhance convolutional neural networks. A team of researchers at the University of the Aegean, the University of West Attica, and the National Technical University of Athens has provided a comprehensive overview of the rapidly evolving landscape of integrated photonic neuromorphic architectures for the implementation of convolutional neural networks.

Their review article was published in *Intelligent Computing*.

Why is photonic computing needed for convolutional neural networks?

Convolutional neural networks are a class of artificial neural network that have revolutionized various fields, particularly computer vision tasks such as image recognition, object detection, and image segmentation. Convolutional neural networks are inspired by the visual processing mechanism of the human brain and are designed to automatically learn hierarchical representations from input data.

When convolutional neural networks are made deeper and have more trainable parameters, their performance tends to improve. However, this improvement comes at a cost. The power consumption and memory requirements also increase significantly. In the field of image processing, the convolution stages of convolutional neural networks use up around 80% of the total power consumed.

To keep up with the growing demand, companies and researchers are trying to use multiple chips and take advantage of parallel processing.

However, this approach leads to a significant increase in [energy usage](#), which raises concerns both in terms of financial costs and ecological impact when scaling up the systems.

How does photonic computing work?

Photonics harnesses light's unique properties for transmitting and processing information, revolutionizing applications like data communication. Integrated photonic convolutional neural networks are a way to speed up the calculations in [convolutional neural networks](#) using special optical devices.

The trick is to transform the convolution step in a convolutional neural network into a series of matrix multiplications. In this method, the image is split into small patches, and each patch is turned into a row in a matrix. The filters or kernels used in the convolution are represented as columns in another matrix, with each column containing the weights of a kernel. The convolutional result is obtained by performing matrix multiplication between these two matrices.

In integrated photonics, a similar process is employed but with light. The image patches are converted into an optical signal, and the kernel weights are stored in a special chip. The optical signal is sent into the chip, and inside the chip, it goes through a process that performs the matrix multiplication using light. This chip is designed specifically for these calculations and is like a mini computer for light. There are different types of these chips, but they all use light to do the calculations.

These chips are called photonic integrated circuits. Some [photonic integrated circuits](#) are reprogrammable, meaning they can be changed to perform different calculations. Others are fixed and designed for specific tasks.

How does photonic computing accelerate convolutional neural networks?

One way to speed up image processing using light is the optical spectrum slicing approach. Imagine a machine that can analyze images very quickly by breaking them down into different colors and patterns. This machine does not need complicated circuits or preprocessing of the images before analyzing them. The method uses special filters that separate the image into different parts based on their colors and patterns. These filters work together to extract important features from the image.

By using this approach, the machine becomes scalable, meaning it can handle larger and more complex images. This method consumes very little power, as it only needs a small amount of energy for detecting the light and processing the signals. It also works instantly, without any delay, so it can process images in real-time.

Another photonic accelerator follows a bioisomorphic approach by using miniaturized spiking laser neurons and unsupervised bioinspired training. This deep architecture offers a noise-resilient and power-efficient solution. Laser neurons simulate the spiking behavior of biological neurons, providing robustness against noise. Unsupervised bioinspired training autonomously extracts meaningful features from data, enabling pattern recognition without explicit labels.

Photonics-based information processing offers energy efficiency. By leveraging these technologies, the accelerator achieves noise resilience and reduced power consumption. In summary, the alternative approach combines miniaturized spiking laser neurons, unsupervised bioinspired training, and photonics to create a power-efficient and noise-resilient deep learning solution.

More information: Aris Tsirigotis et al, Unconventional Integrated Photonic Accelerators for High-Throughput Convolutional Neural Networks, *Intelligent Computing* (2023). [DOI: 10.34133/icomputing.0032](#)

Provided by Intelligent Computing