

RESEARCHERS

M. Choi, A. Majumdar (arka@uw.edu), J. Xiang and E. Shlizerman, University of Washington, USA

REFERENCES

1. A. Wirth-Singh et al. Adv. Photonics Nexus **4**, 026009 (2025).
2. M. Choi et al. Nat. Commun. **16**, 5623 (2025).

# Hybrid Optical-Digital Neural Network for Low-Power Computer Vision

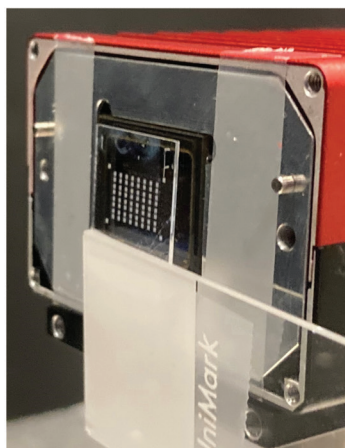
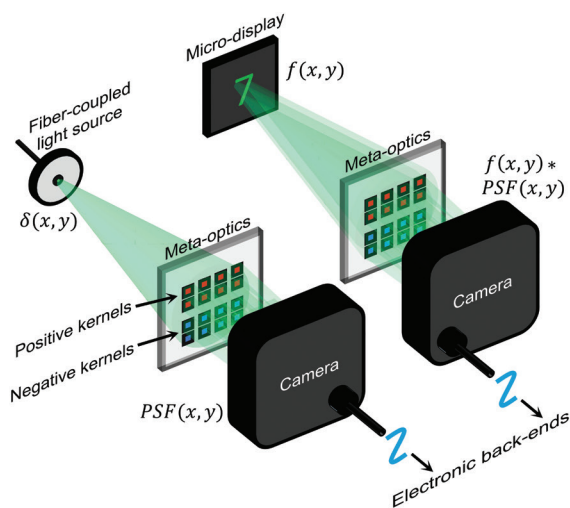
The rapid advancement of deep neural networks and the Internet of Things has enabled significant progress in computer vision tasks such as image classification, segmentation and object detection. As a result, many modern devices are now equipped with multiple cameras to capture and process visual information. However, this capability often comes at the cost of high energy consumption and increased latency. To mitigate these challenges, some approaches reduce image resolution or offload high-resolution processing to cloud servers. Yet cloud computing is not always feasible due to latency constraints, limited network infrastructure or growing privacy concerns. These limitations have driven the rise of low-power computer vision, which aims to empower mobile and edge devices with limited computing resources to perform complex vision tasks efficiently.

Low-power computer vision has been pursued from both algorithmic and hardware perspectives. On the algorithmic side, techniques such as quantization, pruning and knowledge distillation are used to reduce the number of parameters and operations in neural networks. On the hardware side, innovation focuses on improving memory access efficiency and implementing near- or in-sensor computing. Our approach synergizes both fronts—algorithmic compression and optical

hardware acceleration—to achieve ultra-low-power computer vision with significantly reduced energy consumption and latency.

In our work, we employ AlexNet, a foundational model for image classification, and compress its five convolutional layers—interleaved with nonlinear activations and pooling—into a single effective convolutional layer using knowledge distillation. We then replace this compressed convolutional layer with an inversely designed metasurface, which optically implements the learned convolutional kernels via engineered point spread functions (PSFs). This configuration performs convolution operations directly in the optical domain at the moment of image capture. The metasurface functions as an array of meta-optical channels spatially distributed across a single layer, effectively replacing a conventional lens without increasing system complexity. Our first proof of concept demonstrated image classification on the Modified National Institute of Science and Technology handwritten digit dataset.<sup>1</sup>

Building on this foundation, we extended the PSF-engineered metasurface concept to multi-color datasets such as CIFAR-10 and a subset of ImageNet. In this case, a single meta-optical channel performs multichannel convolutions through wavelength multiplexing.<sup>2</sup> This multicolor metasurface enables direct optical convolution at the hardware level, presenting a transformative approach to accelerating computer vision tasks. By executing computationally intensive functions in the optical domain, our method drastically reduces dependency on electronic computation. As a result, it achieves more than two orders of magnitude reduction in system-level energy consumption and over four orders of magnitude decrease in latency, representing a significant leap toward energy-efficient, real-time machine vision for applications such as autonomous navigation and edge computing. **OPN**



Left: Schematics of the experimental setup for point spread functions (PSFs) measurement and image convolution. Right: Photograph of the meta-optics and color camera used in the experiments.