OPEN ACCESS

Optics & Optoelectronics COMMENTARY

The dawn of physical artificial intelligence: Image generation powered by light

Haofan Huang,[1,5] Chi Man Woo,[1,5] Zhiyuan Wang,[1] Tianting Zhong,[1,*] and Puxiang Lai[1,2,3,4,*]

[1]Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR 999077, China
[2]Shenzhen Research Institute, The Hong Kong Polytechnic University, Shenzhen 518057, China
[3]Photonics Research Institute, The Hong Kong Polytechnic University, Hong Kong SAR 999077, China
[4]Jinjiang Technology and Innovation Research Institute, The Hong Kong Polytechnic University, Jinjiang 362251, China
[5]These authors contributed equally
*Correspondence: tianting.zhong@connect.polyu.hk (T.Z.); puxiang.lai@polyu.edu.hk (P.L.)

Received: September 25, 2025; Accepted: November 4, 2025; Published Online: November 5, 2025; https://doi.org/10.59717/j.xinn-inform.2025.100007

Citation: Huang H., Woo C. M., Wang Z., et al. (2025). The dawn of physical artificial intelligence: Image generation powered by light. *The Innovation Informatics* 1:100007.

The rapid advancement of large-scale artificial intelligence (AI) models has unlocked unprecedented capabilities in generative tasks. Yet, mainstream frameworks like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models typically operate on electronic computing platforms that demand substantial digital processing units and extensive storage infrastructure. As models grow in scale and complexity, the ensuring energy consumption, computational latency, and carbon footprint during inference have emerged as critical constraints,[1] posing a major bottleneck to broader deployment. In response, optical computing has risen as a promising alternative, offering inherent advantages such as ultra-low power consumption and massive parallelism. This approach is increasingly being applied to deep learning, particularly in the inference stage. By harnessing fundamental physical properties of light, including diffraction, interference, and waveguide transmission, researchers are accelerating the development of optical neural networks (ONNs) and high-speed optical inference systems, thereby paving the way toward energy-efficient and low-latency AI.

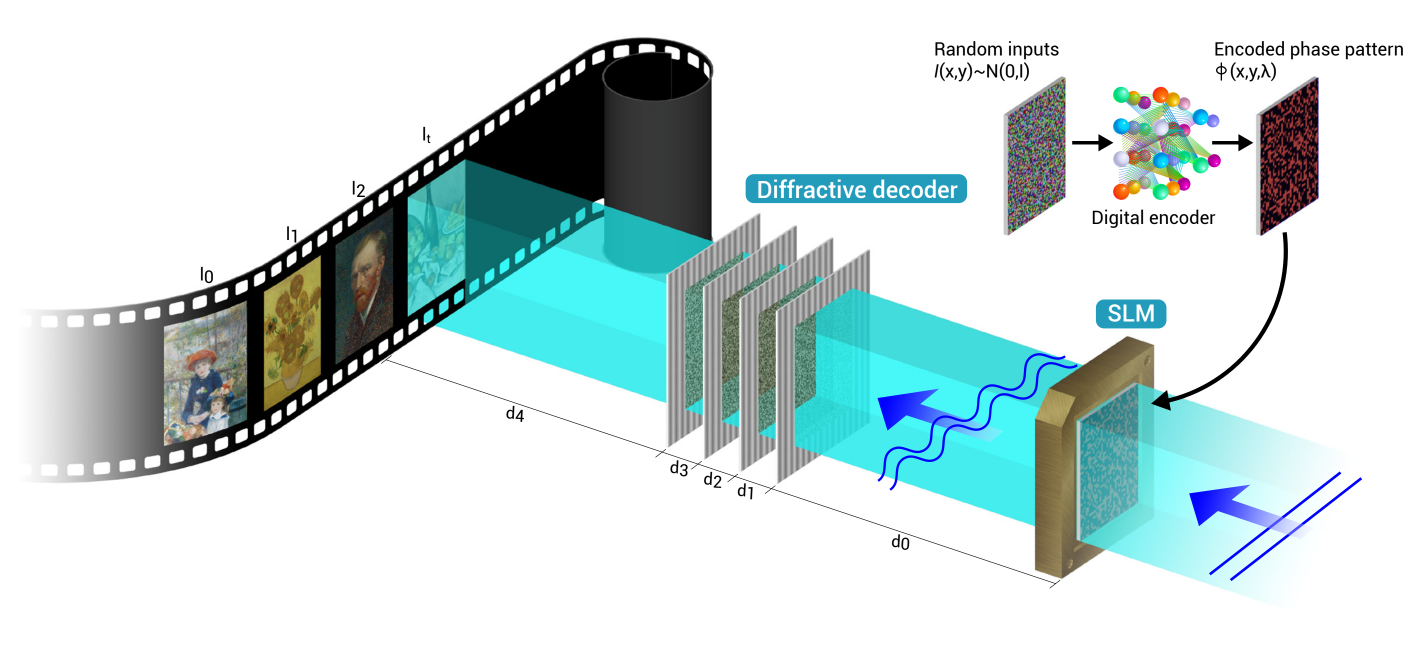

Figure 1. Schematic diagram of the optical generative model The generator produces images by modulating a laser beam. A digital encoder first converts random Gaussian noise into two-dimensional phase patterns, which can be randomly accessed by a SLM and used to represent optical generative seeds in latent space. Each optical seed, once loaded onto the SLM and illuminated by a plane wave, is processed through a reconfigurable diffractive decoder optimized for a specific data distribution to synthesize an image. The generative model is trained using datasets from the same distribution. Once the training is complete, feeding different random seeds into the model yields distinct images that share common features characteristic of the target distribution.

THE OPTICAL GENERATIVE MODEL

Recently, Chen et al. reported a groundbreaking "optical generative model" in Nature.[2] This system directly modulates light beams to synthesize high-quality images that conform to target distributions—such as handwritten digits, articles of clothing, human faces, and even Van Gogh-style paintings. The model begins with a shallow digital encoder that maps Gaussian noise into a two-dimensional phase pattern. This pattern is then encoded into a coherent light beam by a spatial light modulator (SLM), forming an "optical generative seed". Subsequently, the input light field propagates through a diffraction decoder that consists of a series of pre-trained and optimized diffraction layers. Through careful design, the decoder controls the mutual interference of photons, effectively transforming the random optical seed into a coherent image captured by the sensor array. During training, a large digital diffusion model acts as an auxiliary teacher to generate noise-image pairs. The digital encoder and optical decoder are then jointly optimized, allowing the entire system to learn to map random inputs to a target image distribution, demonstrating a powerful ability to generate multiple types of content.

SPEED, EFFICIENCY, AND A NEW COMPUTATIONAL PHILOSOPHY

The core innovation of this work lies in its partial migration of the generative inference process from the digital to the optical domain, leveraging the physical propagation and diffraction of light to achieve image generation with near-zero energy consumption. This approach offers three compelling advantages: (1) Unprecedented energy efficiency: The high parallelism and low power consumption of optical computing make it exceptionally suitable for core deep learning operations. This optical model generates images of comparable quality through a single exposure at an energy cost of merely 30–60 mJ. This is significantly lower than the thousands of iterations required by digital diffusion models, with an estimated energy cost of between 13.25 and 145.8 J per image. (2) Ultra-Fast inference: By physically instantiating network weights as permanent micro-/nano-structures on a diffractive optical element, optical models eliminate the need for repeated memory access, thereby circumventing the latency inherent to sequential digital architectures. Furthermore, owing to the speed of light propagation, image generation is completed within nanoseconds. Despite the computational power of modern GPUs like the NVIDIA RTX 5090, the computation of a single diffusion step in digital models still requires hundreds of milliseconds. (3) Scalability and flexibility: The same optical system can be reconfigured for different tasks by loading pre-trained phase patterns and decoder states, eliminating the need for physical modifications.

This research demonstrates for the first time that optical systems are capable of not only performing discriminative tasks such as classification and regression, but also accomplishing generative tasks, i.e., creating entirely new data. This kind of "optical creativity" opens a new technical pathway for AI-generated content, with particular promise for applications demanding real-time response and high energy efficiency. Within data centers, optical AI accelerators are expected to initially serve as co-processors, responsible for handling computationally intensive and energy-sensitive workloads. In the field of AR/VR, combining optical generative networks with integrated photonics allows for the creation of low-power, real-time scenes and images on edge devices. Looking ahead, as technology matures, it will be possible to instantly generate high-fidelity images, videos, and 3D assets from text or simple sketches. This progression will redefine design processes across film, gaming, and digital art, offering creators powerful real-time assistance and inspiring entirely new forms of artistic creation.

Moreover, this research presents a compelling paradigm for "Physical AI," illustrating that specific computations can be executed directly within the physical domain, bypassing digital conversion altogether. This resurgence of analog computing, augmented by modern deep learning techniques, has the potential to redefine the fundamental boundaries of computation.

CHALLENGES TO FULLY INTEGRATED OPTICAL AI

Despite these promising advances, the long-term impact of optical generative models hinges on achieving end-to-end optical computing—where the entire pipeline, from data preprocessing and training to inference, is executed optically without relying on energy-intensive electronic components. Deep physical neural networks (PNNs) offer a viable pathway by leveraging physical transformations and blurring the conventional separation between software and hardware. For instance, Pai et al. designed a photonic neural network chip that demonstrated an optical implementation of the backpropagation algorithm, the most widely used training method for conventional neural networks.[3] Inspired by the forward-forward algorithm, Momeni et al. demonstrated a backpropagation-free training method for wave-based PNNs,[4] presenting a promising alternative for training optical large language models (LLMs). However, the realization of large-scale optical AI systems still faces a series of formidable practical integration barriers. Due to the significant differences in material and mode properties among optical components (e.g., lasers, waveguides, and fiber arrays), efficient coupling requires submicron to nanometer-level alignment accuracy, which poses a significant challenge to the manufacturing process. Although active alignment can achieve the required precision, its complexity and high cost in high-density array applications make it unsuitable for mass production. In terms of scalability, the primary bottlenecks include shortages of high-performance tunable/modulation devices and broadband light sources, a lack of advanced waveguide materials suitable for high-density integration, and the absence of high-yield heterogeneous integration processes along with unified manufacturing standards.

In the foreseeable future, the hybrid architecture will remain the most viable option.[5] Within this framework, the optics component handles large-scale linear computations and forward propagation, leveraging its speed and parallelism. Meanwhile, electronics focuses on non-linear activations, weight updates and complex control logic. This division of labor exploits the distinct strengths of each domain. It maintains high energy efficiency while circumventing the current challenges associated with implementing all-optical nonlinear computation. Consequently, future development efforts should prioritize the creation of scalable photonic integrated circuits and actively explore hybrid optoelectronic architectures, with the ultimate goal of realizing fully integrated, low-power, end-to-end optical AI systems.

In summary, the research by Chen et al. illuminates a new computational future: one where light supplants electricity, and physical processes replace digital computations.[2] While the realization of end-to-end optical AI remains a challenging journey, each step forward brings us closer to a more efficient, sustainable, and intelligent computational era.

REFERENCES

1. Li H., Yu Z., Zhao Q., et al. (2022). Accelerating deep learning with high energy efficiency: From microchip to physical systems. *The Innovation* 3:100252. DOI:10.1016/j.xinn.2022.100252
2. Chen S., Li Y., Wang Y., et al. (2025). Optical generative models. *Nature* 644:903–911. DOI:10.1038/s41586-025-09446-5
3. Pai S., Sun Z., Hughes T.W., et al. (2023). Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* 380:398–404. DOI:10.1126/science.ade8450
4. Momeni A., Rahmani B., Malléjac M., et al. (2023). Backpropagation-free training of deep physical neural networks. *Science* 382:1297–1303. DOI:10.1126/science.adi8474
5. Bente I., Taheriniya S., Lenzini F., et al. (2025). The potential of multidimensional photonic computing. *Nat. Rev. Phys.* 7:439–450. DOI:10.1038/s42254-025-00843-3

FUNDING AND ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (81930048, 82330061), Guangdong Science and Technology Commission (2019BT02X105), Hong Kong Research Grant Council (15125724, 15217721, C7074-21GF), Shenzhen Science and Technology Innovation Commission (JCYJ20220818100202005), and Hong Kong Polytechnic University (P0039517, P0043485, P0045762, and P0048314). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

H.H.: Conceptualization, Investigation, Writing - original draft. C.W.: Visualization, Writing - original draft. Z.W.: Resources, Visualization. T.Z.: Project administration, Writing - review & editing. P.L.: Supervision, Funding acquisition, Writing - review & editing. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.