Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Light-Field-to-Bits Computing via an Online-Trainable Cascaded Optoelectronic Neural Network

Hongqiang Zhou

hhzhou@bjut.edu.cn

Beijing University of Technology

Likuo Jia

Beijing University of Technology

Tianlong Man

Beijing University of Technology https://orcid.org/0000-0002-2463-7458

Xunbo Yu

Beijing University of Posts and Telecommunications

Sheng Shen

China Mobile Communication Co.

Liangcai Cao

Tsinghua University https://orcid.org/0000-0002-8099-2948

Yuhong Wan

Beijing University of Technology

Article

Keywords:

Posted Date: October 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11002782/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is no conflict of interest

Light-Field-to-Bits Computing via an Online-Trainable Cascaded Optoelectronic Neural Network

Likuo Jia[1#], Hongqiang Zhou[1#], Tianlong Man[1], Xunbo Yu[2†], Sheng Shen[3]，Liangcai Cao[4], Yuhong Wan[1‡]*

1. School of Physics and Optoelectronic Engineering, Beijing University of Technology, Beijing, 100124, China
2. State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876, China
3. China Mobile Group Design Institute Co., Ltd., Beijing, 100080, China
4. State Key Laboratory of Precision Measurement Technology and instruments, Department of Precision instrument, Tsinghua University, Beijing, 100084, China

[#]Authors contributed equally to this work.

Abstract

Optical computing encodes and processes information through the continuous evolution of optical fields, whereas digital electronic systems operate on discrete logical states. Conventional optical-to-electronic interfaces convert measured optical responses into electronically accessible data, with task-specific interpretation subsequently performed in the electronic domain. Here, we introduce and experimentally validate a proof-of-concept light-field-to-bits (LFB) computing architecture that maps information-bearing light fields onto predefined 4-bit codewords. The proposed architecture comprises two cascaded diffractive layers that route optical energy at the output plane into eight predefined regions of interest (ROIs) arranged as four differential pairs, with each pair encoding one bit through relative-intensity comparison.

[*] hhzhou@bjut.edu.cn

[†] yuxunbo@126.com

[‡] yhongw@bjut.edu.cn

We physically implement the first diffractive layer and computationally emulate the second, using a CMOS sensor to capture the intermediate intensity distribution. We further develop a differentiable physics-assisted surrogate model within a camera-in-the-loop framework. The experimental prototype achieves 4-bit codeword accuracies of 95.31% for geometrically perturbed hexadecimal inputs and 93.36% for traffic-sign inputs subjected to combined viewpoint distortion and noise. Meanwhile, the optical front-end exploits the intrinsic parallelism of diffractive propagation, performing high-dimensional optical transformations equivalent to ~51.2 Giga-Operations (GOPs) of multiply accumulate (MAC) operations within a ~1.16 ns optical time-of-flight, while the electronic stage only performs lightweight spatial aggregation at the scale of tens of Mega-FLOPs. This strategy establishes the feasibility of LFB computing as an interface between continuous optical information processing and discrete digital decision-making for latency-sensitive machine vision and edge sensing.

Introduction

The growing computational demands of autonomous systems, robotics, and wireless communications are placing increasing pressure on conventional electronic computing architectures [1-4]. In particular, data movement between physically separated memory and processing units imposes substantial energy and latency costs [4,5]. Optical computing offers a complementary approach by exploiting the broad bandwidth, intrinsic parallelism, and low propagation latency of light [6-9]. These physical attributes make optical computing well suited to large-scale parallel information processing.

Among the various implementations of optical computing, optical neural networks have attracted growing interest because they enable trainable transformations to be implemented directly in optical hardware [10,11]. Diffractive optical neural networks, in particular, perform trainable linear transformations by tailoring the spatial modulation of one or more diffractive layers [12,13]. They have been demonstrated in image classification, phase retrieval, and optical mode processing [12,14,15]. However, passive diffractive propagation provides limited intrinsic nonlinearity, while static diffractive elements offer restricted reconfigurability [10,16,17]. These limitations have motivated hybrid optoelectronic architectures, in which optical propagation performs feature transformation and electronic computation provides programmability, nonlinear operations, and task-specific processing [18,19]. Such hybrid systems have supported increasingly complex applications in machine vision, large-scale learning, and optical generative modelling [20-23].

Beyond computational capability, the representation of optical outputs is also important for connecting optical processors with downstream digital systems. Existing approaches generally follow two representative routes. In class-oriented readout, task results are represented by class-specific optical responses [12,24,25]. In feature-oriented readout, optical features are transferred to electronic processors for further inference [19,20]. These approaches have enabled effective optical inference, but the optical outputs are typically expressed as class responses or intermediate features. Additional interpretation or processing is therefore required before the task result is expressed as a compact digital representation. A compact and efficient route for linking

task-level optical computation with downstream digital representations remains comparatively underexplored.

A further practical challenge arises when numerically trained diffractive systems are transferred to physical hardware. Diffractive optical networks are commonly optimized using numerical propagation models before physical deployment. However, the response of a real optical system can deviate from its ideal model because of device nonidealities, nonuniform illumination, optical misalignment, wavefront aberrations, and environmental fluctuations [26-28]. These discrepancies create a simulation-to-reality gap between the modeled and measured optical responses, which can degrade the performance of parameters optimized entirely in simulation. Hardware-aware and *in situ* training strategies have been developed to incorporate physical responses into the optimization process [28-31]. For experimentally implemented diffractive systems, accounting for the actual hardware response is therefore an important consideration during optimization.

Here, we introduce a LFB computing architecture that maps information-bearing optical fields onto predefined multi-bit codewords through cascaded diffractive transformations. The two diffractive layers route task-relevant optical energy into four pairs of predefined regions at the output plane. Each pair represents one bit through relative-intensity comparison, enabling a 4-bit codeword to be decoded without reconstructing the input image or generating conventional class scores. To experimentally validate this concept under realistic hardware conditions, we implement the first diffractive layer physically and emulate the second computationally, with a

CMOS camera recording the intermediate optical intensity. We further employ a camera-in-the-loop optimization framework with a physics-assisted surrogate model. The measured optical response is retained in the forward process, while the surrogate combines analytical diffraction propagation with learned hardware corrections to provide gradients for updating the physical phase parameters. Using hexadecimal and traffic-sign inputs, the system demonstrates robust 16-class-to-4-bit encoding and achieves complete-codeword accuracies of 95.31% and 93.36%, respectively, under perturbation-expanded test conditions. The localized output distribution further provides a pathway toward sparse photodetector readout in future implementations.

Results

Operational principle of the LFB architecture

The proposed architecture provides an interface between continuous light field and discrete digital signals. As conceptually illustrated in **Fig. 1**, the input digital information (hexadecimal digits 0–F) is encoded into the spatial amplitude of the light field, structured as 4-quadrant tiled patterns to correspond spatially with the 4-bit output channels. This information-carrying wavefront propagates through a cascaded diffractive network comprising two modulation layers. The network extracts the high-dimensional optical features and routes them into target ROIs on the detection plane.

To encode these focused optical features into deterministic 4-bit binary codewords (0000–1111), a dual-focus contrast readout mechanism is implemented. Instead of relying on absolute intensity thresholds, which are highly susceptible to systemic power fluctuations and environmental noise, this mechanism determines bit values through the

relative intensity contrast. In the ultimate hardware vision, this readout is executed by a sparse array of eight photodetectors (PDs) spatially positioned at the predefined ROIs and organized into four differential pairs. For each bit channel (b_i, where $i \in \{0,1,2,3\}$), the corresponding pair consists of an upper-left detector (PD_{UL}) and a lower-right detector (PD_{LR}). The optical energy (I_{UL}) captured by PD_{UL} represents the logic state '1', while the energy (I_{LR}) captured by PD_{LR} represents the logic state '0'. Consequently, the binary value of each bit is strictly defined by the comparative contrast: $b_i = 1$, if $I_{UL} > I_{LR}$, and $b_i = 0$, otherwise.

While the ultimate target architecture envisions sparse photodetector arrays to reduce the dependence on full-frame quantization and minimize analog to digital conversion (ADC) channels, the experimental validation in this work is conducted using an optoelectronic proof-of-concept prototype.

***In situ* optimization via a differentiable physics-assisted proxy**

While numerical simulations provide a theoretical foundation, physical deployments inevitably suffer from a "reality gap" induced by optical misalignments, phase aberrations, and non-uniform illumination. To overcome these systemic discrepancies, a camera-in-the-loop optimization strategy is employed [27]. Unlike static offline training, this approach treats the physical optical path as the forward propagation, ensuring the optimization process is strictly grounded in the physical conditions of the experimental setup.

As illustrated in **Fig. 2**, the system operates through a collaborative optoelectronic loop. The forward pass (top row) consists of two stages. First, the input amplitude field

(E_{in}) is modulated by the physical phase layer (Φ_{phys}). The light physically propagates through free space, denoted by the transfer function P_{phys}, and the resulting intermediate intensity distribution (I_{mid}) is recorded by the CMOS sensor. Subsequently, in the digital domain, the captured intensity map is normalized and converted into a non-negative amplitude feature. It is then combined with a trainable electronic phase mask (Φ_{elec}) to form a computational complex field for subsequent angular spectrum method (ASM) propagation [32], generating the final output distribution (I_{out}).

A key component of the optimization framework is the differentiable physics-assisted surrogate model (bottom row). The network parameters are optimized via the Adam algorithm to minimize the task loss (L_{task}), defined as the mean squared error (MSE) between the prediction (I_{out}) and the ground-truth target (I_{gt}). Gradients for the digital layer (Φ_{elec}) can be obtained directly through standard automatic differentiation (∇L_{task}). However, gradients cannot be directly obtained through the physical optical path using standard automatic differentiation. To resolve this, we construct a physics-assisted surrogate model to serve as a differentiable digital proxy. Rather than relying solely on a numerical ASM, the proposed surrogate model augments the analytical ASM framework (U_{ASM}) with learnable correction terms, including amplitude attenuation $\alpha(x, y)$ and phase aberration $\phi_{ab}(x, y)$. Unlike data-driven black-box mappings, this formulation preserves the underlying physical propagation model while introducing minimal learnable deviations to account for hardware-induced imperfections. The detailed implementation of the surrogate model and its layer-wise

architecture and physical interpretation are summarized in *Supplementary note* 1, which explicitly maps each computational component to its physical counterpart. By dynamically aligning with the measured optical responses, this surrogate approximates the actual physical propagation process, establishing a reliable differentiable pathway. Consequently, the error gradients bypass the non-differentiable physical barrier, backpropagating from the digital domain to accurately update the physical phase parameters (Φ_{phys}). The contribution of the surrogate-assisted optimization is further quantified through a controlled ablation study in *Supplementary Note* 3 and Fig. S2a–d, where numerical and physical implementations trained with and without the surrogate model are compared under both normal and perturbed input conditions.

Experimental validation of LFB encoding

To validate the efficacy of the proposed "LFB" online-trainable optoelectronic system, we utilized the experimental setup illustrated in **Fig.3** to encode spatial patterns of hexadecimal digits (0-F) into 4-bit binary codewords (0000-1111). The system consists of three core components: an amplitude-type SLM that dynamically encodes the digital input into the spatial wavefront, a subsequent phase-type SLM that executes the trainable diffractive modulation, and a CMOS sensor. Following the sensor capture, a personal computer (PC) is employed to execute the secondary electronic diffractive layer. Synchronized via a custom control program, this hybrid optoelectronic pipeline establishes the physical foundation for *in situ* error backpropagation. This hardware-in-the-loop control allows the network to automatically adapt to the imperfect physical

hardware conditions. Ultimately, it shifts the burden of system accuracy from extreme mechanical precision to adaptive algorithmic optimization.

Building upon the experimental platform, we first validated the proposed LFB encoding strategy using hexadecimal optical patterns. Sixteen hexadecimal classes (0–F) were assigned to their corresponding 4-bit binary codewords (0000 – 1111). Representative results are shown in **Fig. 4a**, where each input pattern is directly converted into a spatial intensity distribution. The surrogate simulation and the hybrid optoelectronic execution both produce the expected optical routing patterns, enabling direct binary decoding without reconstructing the input image or introducing a conventional electronic classifier.

To further examine whether this LFB mapping can be extended beyond regular hexadecimal symbols, we introduced a traffic-sign-based visual encoding task. In this task, 16 traffic-sign classes were assigned to predefined 4-bit codewords, forming a nontrivial mapping from semantically distinct visual patterns to compact binary outputs. Representative traffic-sign encoding results are shown in **Fig. 4c**. Compared with hexadecimal patterns, the traffic-sign inputs contain richer spatial structures and semantic differences, yet the hybrid system still produces clear differential ROI intensity contrasts for binary decoding. The full qualitative conversion results for all 16 classes in both the hexadecimal and traffic-sign tasks, together with their corresponding learned phase distributions, are provided in *Supplementary Note* 2 and *Supplementary Fig. S*1.

Comprehensive robustness against input perturbations and hardware degradation.

Beyond representative encoding results, we further quantified the robustness of the proposed LFB encoder using perturbation-expanded test sets. For the hexadecimal encoding task, each of the 16 classes was evaluated using 16 perturbed test instances generated through scaling, in-plane rotation, and translational shift, resulting in a total of 256 test instances. As summarized by the confusion matrix in **Fig. 4b**, the hybrid optoelectronic system correctly decoded 244 out of 256 instances, corresponding to a complete 4-bit codeword accuracy of 95.31%. The strongly diagonal distribution indicates that the learned differential ROI routing remains stable under moderate geometric deformation of the incident amplitude patterns.

We then evaluated the traffic-sign-based encoding task under viewpoint perturbations. In this test, the hybrid system was evaluated under 16 simultaneous pitch-yaw viewpoint combinations, with both pitch and yaw sampled from −25° to +25°. A hybrid input noise was additionally applied to emulate practical optical degradation. As shown in **Fig. 4d**, the system correctly decoded 239 out of 256 traffic-sign instances, corresponding to a complete 4-bit codeword accuracy of 93.36%. Although several sparse off-diagonal elements appear in the confusion matrix, most errors correspond to limited codeword-level confusions, indicating that the differential ROI readout remains robust under combined viewpoint distortion and input noise.

In addition to these task-level perturbation tests, we systematically characterized the physical tolerance limits of the system under individual perturbation dimensions,

including rotational aberration, translational misalignment, defocusing-induced scaling, and edge-deployment hardware degradation. As shown in **Fig. 5a–d**, the classification accuracy gradually degrades when the perturbation magnitude exceeds the learned operating regime. These results together demonstrate that the proposed LFB encoder maintains reliable binary codeword conversion under both input-level geometric deformation and hardware-level signal degradation.

Beyond lateral misalignments, the architecture exhibits robustness to longitudinal defocusing. To visualize this extended axial tolerance, we reconstructed the 3D propagation of the diffracted light field along the Z-axis. As visualized in **Fig. 5e**, the 3D stacked slices depict the continuous spatial distribution of the optical energy from Z = 15 cm to Z = 25 cm. Although the designated focal plane for sensor capture is positioned at Z = 20 cm, the optical field retains structured and spatially localized intermediate features over this 10-cm axial range. Rather than rapidly dispersing into an unstructured distribution, the field retains distinct intermediate spatial clusters. Notably, these localized clusters do not yet represent the final 4-bit readout ROIs. Rather, they act as pre-conditioned physical features that are subsequently focused into the final differential pairs by the secondary electronic layer. Ultimately, these comprehensive robustness evaluations demonstrate that the operational burden is transferred from mechanical alignment to a robust optical routing mechanism.

Conclusion and discussion

In this work, we present an online-trainable LFB computing architecture that maps information-bearing optical fields onto predefined digital codewords through

differential readout. Employing a camera-in-the-loop optimization strategy with a physics-assisted surrogate model based on the angular spectrum method, we enable joint optimization of the physical and digital parameters in the hybrid prototype. The effectiveness of the surrogate-assisted optimization in mitigating hardware model mismatch and improving robustness is further validated through extensive ablation studies in *Supplementary Note* 3, which compare surrogate-assisted and non-surrogate training under both clean and perturbed conditions. This framework shifts the operational burden from demanding mechanical precision to adaptive algorithmic optimization.

The efficacy of this architecture is experimentally validated through a dual-layer implementation comprising a physical optical diffractive layer and a secondary electronic diffractive layer. The physical layer extracts high-dimensional features, while the electronic layer aggregates them to execute a conversion of hexadecimal optical patterns (0–F) into 4-bit binary codewords. This deterministic spatial routing endows the system with physical robustness. As demonstrated, the network maintains zero-error encoding even under aggressive hardware degradation—specifically, operating with up to a 40% CMOS dead-pixel ratio combined with a 2-bit ADC quantization width. Furthermore, the optical field maintains structured and spatially localized features over this 10-cm axial range.

Although the optical propagation itself provides massive parallelism within a nanosecond-scale time-of-flight, the current system-level throughput remains limited by the refresh rate of the SLM and the readout delay of the CMOS camera. Therefore,

the estimated optical operations should be interpreted as an equivalent computational capacity of the diffractive propagation process rather than the measured end-to-end throughput of the present prototype. By offloading high-dimensional feature extraction to the physical optical domain, our architecture reduces the amount of digital computation required for the optical transformation. For $160{,}000$ spatial modes, a single optical forward pass passively executes ~ 51.2 Giga-Operations (GOPs) in equivalent MACs within a ~ 1.16 ns time-of-flight. In contrast, the secondary electronic layer performs lightweight spatial aggregation, reducing digital complexity to $O(N \log N)$ (tens of Mega-FLOPs). Although the current system-level throughput is bottlenecked to ~ 3.07 TOPS by the 60 Hz SLM-CMOS refresh rate, future integration with on-chip metasurfaces and GHz-level sparse photodetectors could better exploit the intrinsic parallelism of optical propagation for ultra-high-throughput computing. By replacing the macroscopic SLM and the electronic diffractive layer with on-chip metasurfaces and the CMOS with eight PD arrays at the target ROIs, these hardware constraints can be substantially reduced. Such nanophotonic integration would reduce the system footprint and could reduce the sensing and data-conversion overhead associated with full-frame acquisition. While the current tabletop SLM–CMOS platform provides a proof-of-concept validation of the physics-assisted optoelectronic encoding mechanism, the same design principle may ultimately be extended toward compact and high-speed chip-scale implementations. Such a transition would primarily involve replacing the SLM with a planar nanophotonic element and substituting the full-frame CMOS sensor with a sparse array of discrete photodetectors.

First, the optical computing layer could be miniaturized using dielectric metasurfaces. Compared with the micrometer-scale pixel pitch and centimeter-scale propagation distances of the present SLM-based platform, metasurfaces offer substantially finer spatial sampling and can implement complex wavefront transformations within a much smaller footprint [33,34]. In this scenario, the phase profiles obtained through surrogate-assisted optimization would need to be translated into fabricable meta-atom geometries while accounting for discrete phase levels, inter-element coupling, material dispersion, and fabrication constraints. The resulting device could serve as a compact flat-optical computational layer, although its achievable efficiency and bandwidth would depend on the selected material platform and structural design.

Second, the differential readout provides a natural route toward sparse optical detection. Because the trained optical front-end directs energy into eight predefined regions corresponding to four differential bit channels, a future implementation could replace the full-frame CMOS sensor with eight discrete high-speed photodiodes. Each detector pair could be connected to transimpedance amplifiers and differential comparators to generate the corresponding binary electrical output [35]. In principle, such a sparse readout architecture would reduce the number of sensing and digitization channels required for inference and could alleviate the data-conversion overhead associated with full-frame image acquisition [35,36]. Its actual operating speed and energy consumption would be determined by the photodetector bandwidth, amplifier design, comparator circuitry, and optical power budget; therefore, GHz-level operation

and low energy-per-bit performance should be regarded as prospective system-level targets rather than demonstrated results of the present prototype.

Overall, the SLM–CMOS system should be viewed as an experimentally accessible validation platform for a broader sparse-readout optoelectronic architecture. Metasurfaces integration and discrete photodetector readout represent promising future directions for reducing system footprint and readout complexity, but their performance advantages will require dedicated device-level design, fabrication, packaging, and experimental verification. Ultimately, this architecture paves the way for latency-critical applications in high-throughput machine vision, all-optical network routing, and edge-aware sensing.

Materials and Methods

Experimental details of the optoelectronic hybrid pipeline

The experimental validation was conducted using a custom-built, free-space optoelectronic computing platform. A coherent laser ($\lambda = 532\ nm$) was utilized as the illumination source. The beam was spatially filtered and expanded to provide uniform plane-wave illumination. The digital input data (400×400 pixels) were loaded onto the first amplitude-type SLM, which served as the data encoder. The incident beam was pre-polarized via a linear polarizer before illuminating this first SLM. The modulated wavefront was then relayed through a standard 4f optical system, where a spatial filter was placed at the Fourier plane to eliminate higher-order diffraction artifacts. Upon exiting the 4f system, the beam passed through a vertical linear polarizer. This step was strictly implemented because the accurate amplitude encoding generated by the

amplitude-type SLM is manifested within the vertically polarized component. Subsequently, to guarantee a pure phase-modulation regime without unwanted amplitude coupling at the next stage, a half-wave plate was employed to adapt this vertical polarization, strictly aligning it with the active liquid crystal axis of the second, phase-type SLM. Following this rigorous polarization control, the beam freely propagated over a physical distance of 5 cm to illuminate the phase-only SLM (pixel pitch: $8\,\mu m$). This phase SLM functioned as the primary physical diffractive layer, modulating the wavefront with the trained 400×400 spatial phase parameters. The modulated light field freely propagated over a macroscopic distance of $Z = 20\,cm$. The resulting intermediate intensity distribution at the focal plane was captured by a CMOS image sensor. The captured spatial intensity map was then transferred to a PC, where the secondary electronic diffractive layer computationally focused the extracted physical features onto the target dual-focus readout zones.

Before CITL optimization, the camera region was calibrated to establish a reproducible spatial correspondence between the measured optical field and the digital array. To achieve pixel-to-pixel registration, a computer-generated hologram encoding a predefined calibration pattern was loaded onto the phase SLM. This pattern comprised a precisely defined array of sparse circular reference dots spanning a target region of 400 × 400 pixels, with an exact center-to-center spacing of 80 pixels between adjacent dots. Following phase modulation, the calibration field underwent free-space diffractive propagation over a distance of 20 cm to be reconstructed at the camera plane. By capturing these distinct fiducial markers with the CMOS sensor, we calculated the exact

coordinate mapping matrix. This initial calibration ensures the precise spatial registration of the captured optical computing region, which is a fundamental prerequisite for accurately applying the subsequent electronic phase modulation in the digital domain.

Physics-assisted surrogate model and forward propagation

To dynamically track and compensate for the discrepancy between the ideal numerical propagation model and the real optical hardware, we constructed a differentiable physics-assisted surrogate model. The ideal free-space propagation was described by the ASM [32]. Given an input light field ($U_1(x, y)$), the propagated field after a distance (z) can be expressed as:

$$U_2(x, y) = \mathcal{F}^{-1}\{\mathcal{F}[U_1(x, y)] \cdot H(f_x, f_y)\} \tag{1}$$

where ($\mathcal{F}$) and ($\mathcal{F}^{-1}$) denote the Fourier transform and inverse Fourier transform, respectively. The free-space transfer function is given by

$$H(f_x, f_y) = \exp\left[j\frac{2\pi z}{\lambda}\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right] \tag{2}$$

where (λ) is the wavelength, and (f_x) and (f_y) are the spatial frequencies along the *x*- and *y*-directions.

In the experimental system, however, the actual optical response deviates from the ideal ASM model because of nonuniform illumination, imperfect SLM phase modulation, optical misalignment and wavefront aberrations. To bridge this reality gap, a physics-assisted surrogate model was incorporated into the forward propagation process. The input field is first propagated to the SLM plane, where the loaded phase is corrected by a learnable SLM response module. Then, a residual convolutional network

predicts the amplitude correction and phase aberration associated with the experimental system. These learned corrections are imposed on the complex field before it is further propagated to the output plane. The layer-wise architecture and physical roles of each module are summarized in *Supplementary Table* S1, which explicitly maps each computational block to its physical counterpart.

During training, the parameters of the SLM correction module and the residual CNN are optimized by minimizing the discrepancy between the surrogate-predicted intensity and the experimentally captured intensity. In this way, the surrogate model preserves the differentiability and physical interpretability of ASM-based propagation while learning the amplitude attenuation, phase aberration, and modulation errors induced by the real optical hardware. The optoelectronic hardware-degradation model and the theoretical analyses of computational complexity and throughput are provided in *Supplementary Notes* 4 and 5, respectively.

Funding

National Natural Science Foundation of China (62305009, 62675006). China Mobile Communications Corporation (CMCC) (C2025001 - A2025268).

Disclosures

All the authors declare no conflict of interest.

Code Availability

The code used to generate the results of this study is available from the corresponding author upon reasonable request.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

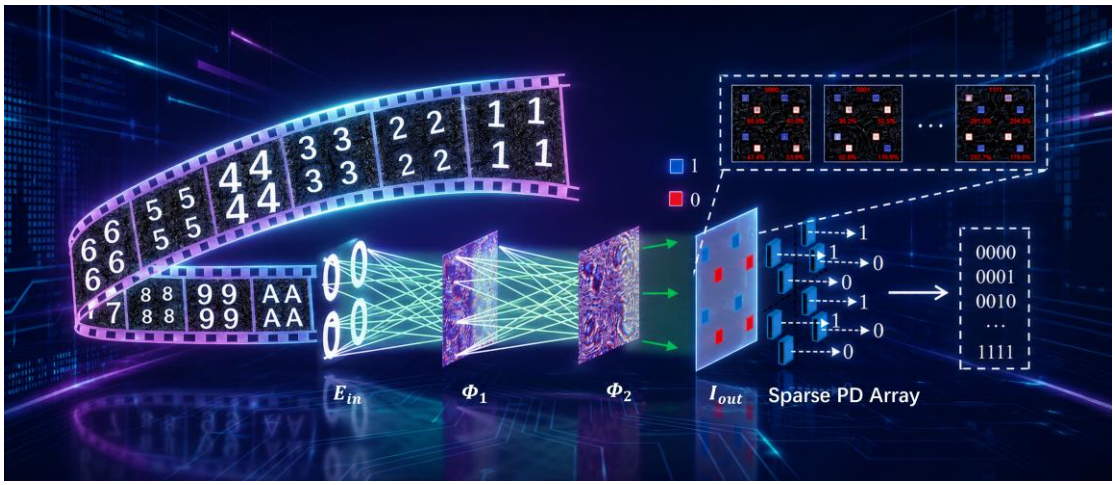


Fig. 1 Conceptual architecture of the "Light-Field-to-Bits" encoder. The system provides a direct "LFB" mapping from hexadecimal input patterns (0-F) to 4-bit binary codewords (0000-1111). The input digital information is initially encoded into the spatial amplitude of the light field(E_{in}). The wavefront then undergoes a cascaded diffractive modulation (Φ_1 and Φ_2). The network deterministically routes the optical energy into specific regions of interest (ROIs) on the output detection plane (I_{out}). A sparse photodetector (PD) array is conceptualized to capture the focused optical distribution. The eight PDs are grouped into four differential pairs. Within each pair, the detector receiving the higher optical intensity directly dictates the output logic state of the corresponding bit.

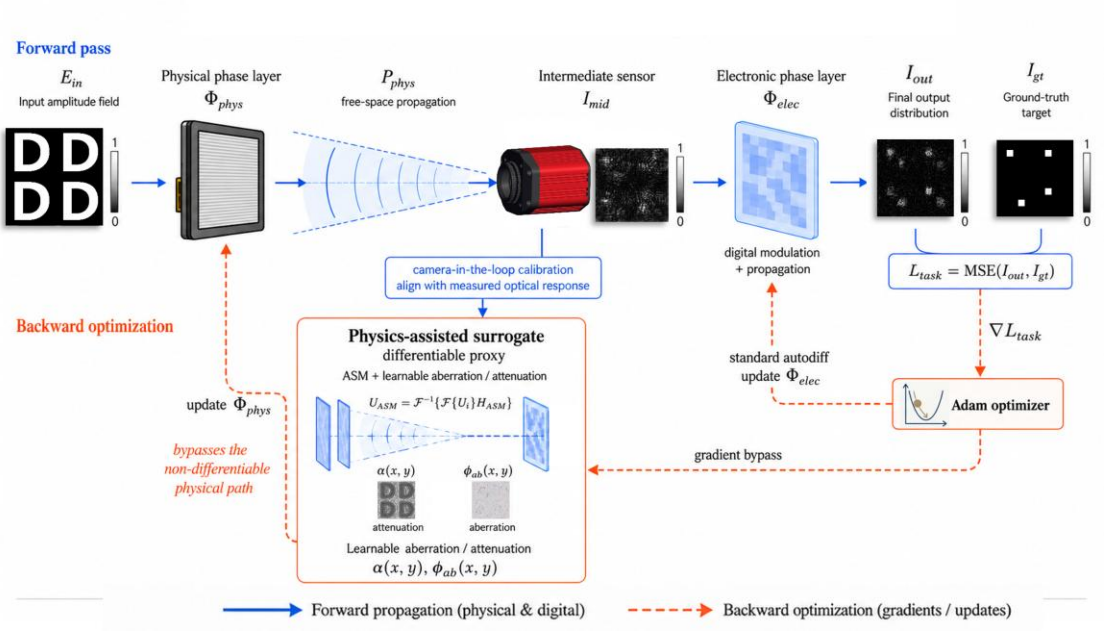


Fig. 2 *In situ* camera-in-the-loop optimization framework driven by a physics-assisted surrogate model. The system architecture operates through a collaborative optoelectronic loop comprising a forward propagation pass (solid blue arrows) and a backward optimization pathway (dashed orange arrows). In the forward pass, the input amplitude field (E_{in}) is modulated by the physical phase layer (Φ_{phys}) and undergoes free-space diffractive propagation (P_{phys}). The intermediate intensity (I_{mid}) is captured by a CMOS sensor and subsequently processed by the electronic phase layer (Φ_{elec}) to yield the final output distribution (I_{out}). During backward optimization, the task loss (L_{task}) is computed by evaluating the deviation between the final output distribution (I_{out}) and the ground-truth target (I_{gt}). The error gradients bypass the actual optical path and backpropagate through a differentiable physics-assisted surrogate.

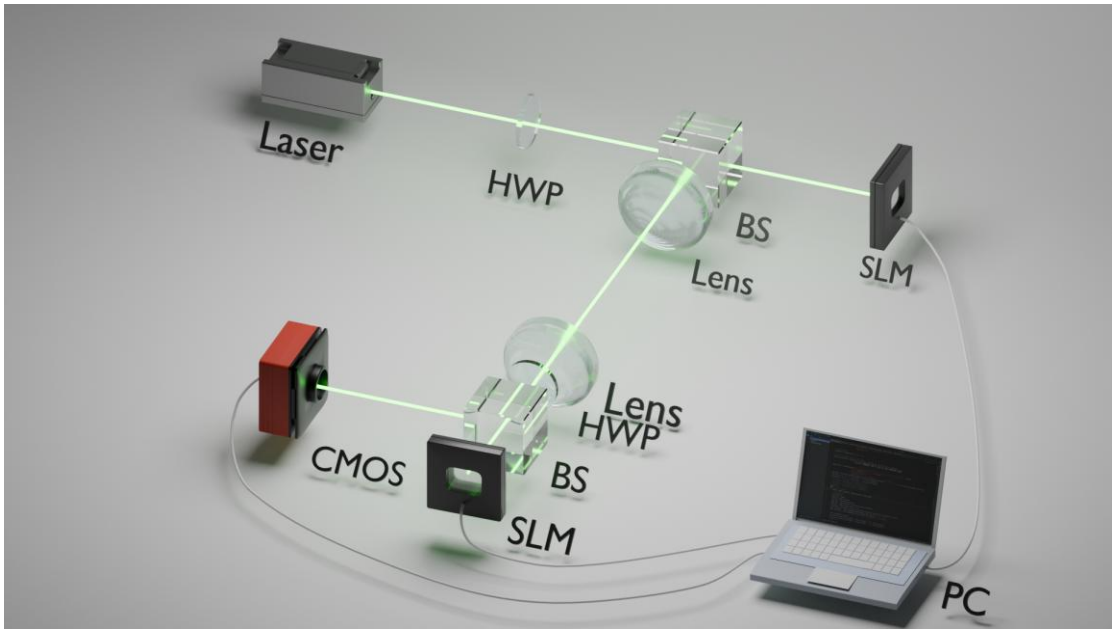


Fig. 3 Schematic of the experimental optical setup. The system employs a cascaded reflective architecture. A 532 *nm* wavelength laser serves as the coherent light source. The beam is expanded and polarization-tuned by a half-wave plate (HWP) before illuminating the amplitude-type SLM, which encodes the input information. An optical relay system comprising beam splitters (BS) and lenses directs the signal to the phase-type SLM, which executes the trainable phase modulation. The intermediate diffracted intensity is captured by a CMOS sensor and transferred to a PC for implementation of the secondary electronic diffractive layer. PC: Personal computer.

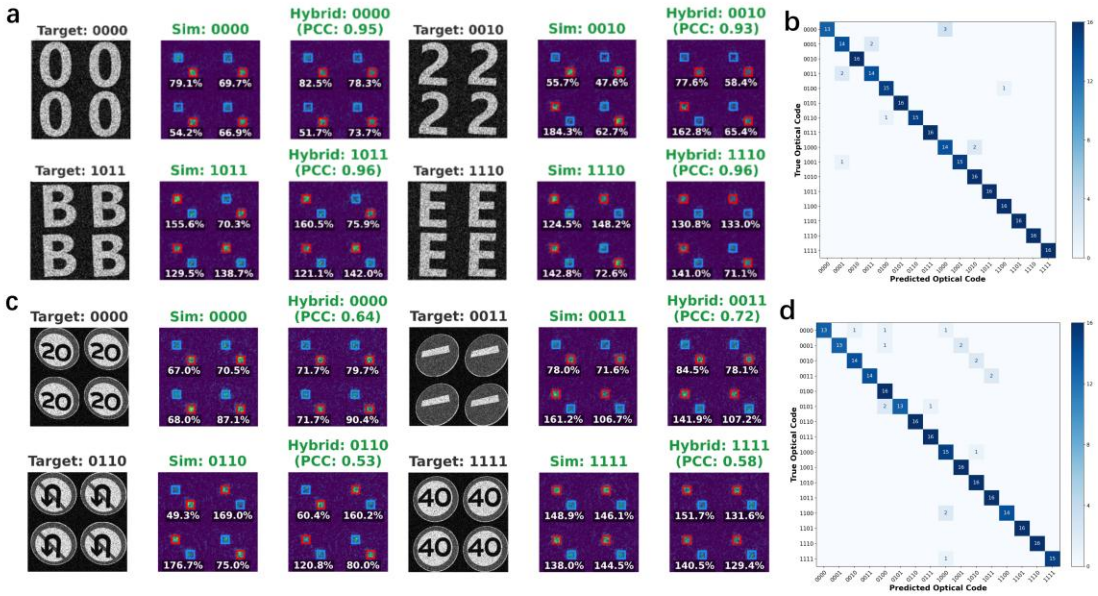


Fig. 4 | LFB encoding of hexadecimal and traffic-sign visual inputs.
(a) Representative encoding results for hexadecimal optical patterns. Each input class is converted into a predefined 4-bit binary codeword through differential ROI intensity comparison. The target input, surrogate-simulated output (**Sim.**), and hybrid optoelectronic output (**Hybrid**) are shown. (b) Confusion matrix for the hexadecimal encoding task using 16 perturbed test instances per class, generated through scaling, in-plane rotation, and translational shift. (c) Representative encoding results for traffic-sign-based visual inputs. Each traffic-sign class is assigned to a predefined 4-bit binary codeword, demonstrating the extension from regular symbolic inputs to semantically distinct visual patterns. (d) Confusion matrix for the traffic-sign encoding task under 16 simultaneous pitch-yaw viewpoint combinations from −25° to +25° with additional hybrid input noise. The traffic-sign task achieves 93.36% complete 4-bit codeword accuracy over 256 test instances.

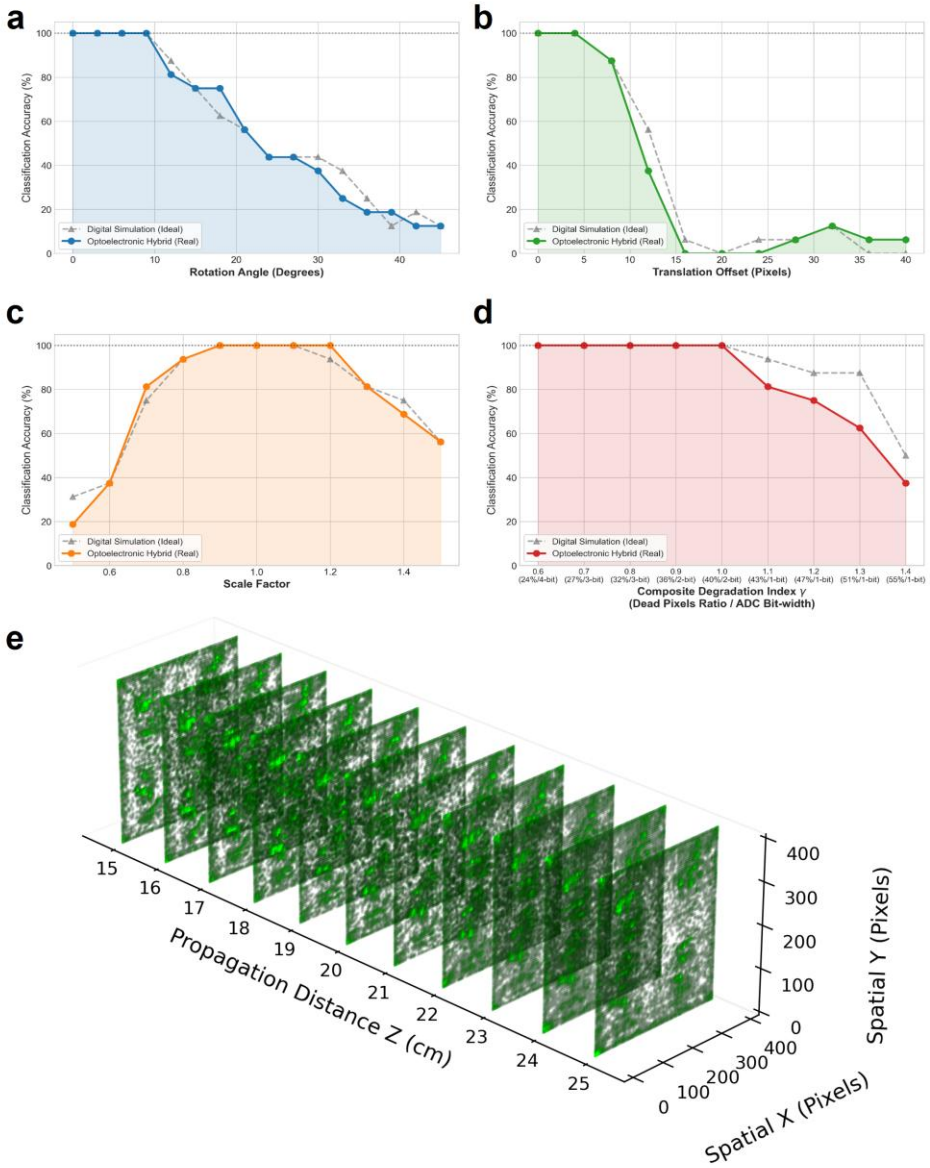


Fig. 5 Comprehensive robustness and physical tolerance evaluation of the "LFB" optoelectronic encoder. (a–c) Complete 4-bit codeword accuracy of the system under various spatial input perturbations, comparing the ideal digital simulation (dashed grey lines) with the experimental optoelectronic hybrid implementation (solid colored lines). (a) rotation, (b) translation offset, and (c) scale factors. (d) System robustness against composite optoelectronic hardware noise. The x-axis represents a composite degradation index γ, which pairs the CMOS dead-pixel ratio with the ADC quantization bit-width. (e) Volumetric evolution of the diffractive light field along the propagation axis.

References

1. C. Mao, Z. Mu, Q. Liang, *et al.*, “Deep learning in physical layer communications: evolution and prospects in 5g and 6g networks” IET Commun. **17**(16), 1863-1876 (2023).
2. W. Shi, J. Cao, Q. Zhang, *et al.*, “Edge computing: vision and challenges” IEEE Internet Things J. **3**(5), 637-646 (2016).
3. S. Levine, C. Finn, T. Darrell, *et al.*, “End-to-end training of deep visuomotor policies” J. Mach. Learn. Res. **17**(39), 1-40 (2016).
4. J. L. Hennessy, D. A. Patterson, “A new golden age for computer architecture” Commun. ACM **62**(2), 48-60 (2019).
5. D. Ielmini, H. S. P. Wong, “In-memory computing with resistive switching devices” Nat. Electron. **1**(6), 333-343 (2018).
6. H. J. Caulfield, S. Dolev, “Why future supercomputing requires optics” Nat. Photonics **4**(5), 261-263 (2010).
7. D. A. Miller, “Attojoule optoelectronics for low-energy information processing and communications” J. Lightwave Technol. **35**(3), 346-396 (2017).
8. G. Wetzstein, A. Ozcan, S. Gigan, *et al.*, “Inference in artificial intelligence with deep optics and photonics” Nature **588**(7836), 39-47 (2020).
9. B. J. Shastri, A. N. Tait, T. Ferreira De Lima, *et al.*, “Photonics for artificial intelligence and neuromorphic computing” Nat. Photonics **15**(2), 102-114 (2021).
10. T. Fu, J. Zhang, R. Sun, *et al.*, “Optical neural networks: progress and challenges” Light: Science & Applications **13**(1), 263 (2024).
11. Y. Shen, N. C. Harris, S. Skirlo, *et al.*, “Deep learning with coherent nanophotonic circuits” Nat. Photonics **11**(7), 441-446 (2017).
12. X. Lin, Y. Rivenson, N. T. Yardimci, *et al.*, “All-optical machine learning using diffractive deep neural networks” Science **361**(6406), 1004-1008 (2018).
13. O. Kulce, D. Mengu, Y. Rivenson, *et al.*, “All-optical synthesis of an arbitrary linear transformation using diffractive surfaces” Light: Science & Applications **10**(1), 196 (2021).
14. Z. Lu, S. Duan, H. Zhang, *et al.*, “Diffractive optical neural network for dual-wavelength vectorial vortex beams de-/multiplexing” Opt. Lett. **50**(12), 4082-4085 (2025).
15. Y. Li, Y. Luo, D. Mengu, *et al.*, “Quantitative phase imaging (QPI) through random diffusers using a diffractive optical network. Light: Advanced Manufacturing 4, Article 17 (2023).
16. Y. Zuo, B. Li, Y. Zhao, *et al.*, “All-optical neural network with nonlinear activation functions” Optica **6**(9), 1132-1137 (2019).
17. C. Liu, Q. Ma, Z. J. Luo, *et al.*, “A programmable diffractive deep neural network based on a digital-coding metasurface array” Nat. Electron. **5**(2), 113-122 (2022).
18. T. Zhou, X. Lin, J. Wu, *et al.*, “Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit” Nat. Photonics **15**(5), 367-373 (2021).
19. D. Mengu, Y. Luo, Y. Rivenson, *et al.*, “Analysis of diffractive optical neural networks and their integration with electronic neural networks” IEEE J. Sel. Top. Quantum Electron. **26**(1), 1-14 (2020).
20. Y. Chen, M. Nazhamaiti, H. Xu, *et al.*, “All-analog photoelectronic chip for high-speed vision tasks” Nature **623**(7985), 48-57 (2023).

21. X. Yuan, Y. Wang, Z. Xu, *et al.*, “Training large-scale optoelectronic neural networks with dual-neuron optical-artificial learning” Nat. Commun. **14**(1), 7110 (2023).
22. S. Chen, Y. Li, Y. Wang, *et al.*, “Optical generative models” Nature **644**(8078), 903-911 (2025).
23. J. Qiu, G. Lu, T. Liu, *et al.*, “Optoelectronic generative adversarial networks” Commun. Phys. **8**(1), 162 (2025).
24. M. S. S. Rahman, J. Li, D. Mengu, *et al.*, “Ensemble learning of diffractive optical networks” Light: Science & Applications **10**(1), 14 (2021).
25. Y. Li, J. Li, A. Ozcan, “Nonlinear encoding in diffractive information processing using linear optical materials” Light: Science & Applications **13**(1), 173 (2024).
26. D. Mengu, Y. Zhao, N. T. Yardimci, *et al.*, “Misalignment resilient diffractive optical networks” Nanophotonics **9**(13), 4207-4219 (2020).
27. Y. Peng, S. Choi, N. Padmanaban, *et al.*, “Neural holography with camera-in-the-loop training” ACM Trans. Graph. **39**(6), 185 (2020).
28. L. G. Wright, T. Onodera, M. M. Stein, *et al.*, “Deep physical neural networks trained with backpropagation” Nature **601**(7894), 549-555 (2022).
29. Z. Xue, T. Zhou, Z. Xu, *et al.*, “Fully forward mode training for optical neural networks” Nature **632**(8024), 280-286 (2024).
30. S. Bandyopadhyay, A. Sludds, S. Krastanov, *et al.*, “Single-chip photonic deep neural network with forward-only training” Nat. Photonics **18**(12), 1335-1343 (2024).
31. J. C. A. Rocha, U. G. Būtaitė, J. Carpenter, *et al.*, “Self-configuring high-speed multi-plane light conversion” Nat. Commun. **17**(1), 73 (2026).
32. K. Matsushima, T. Shimobaba, “Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields” Opt. Express **17**(22), 19662-19673 (2009).
33. E. Goi, X. Chen, Q. Zhang, *et al.*, “Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip” Light: Science & Applications **10**(1), 40 (2021).
34. H. H. Zhu, J. Zou, H. Zhang, *et al.*, “Space-efficient optical computing with an integrated chip diffractive neural network” Nat. Commun. **13**(1), 1044 (2022).
35. A. Song, S. N. M. Kottapalli, R. Goyal, *et al.*, “Low-power scalable multilayer optoelectronic neural networks enabled with incoherent light” Nat. Commun. **15**(1), 10692 (2024).
36. L. Mennel, D. K. Polyushkin, D. Kwak, *et al.*, “Sparse pixel image sensor” Sci. Rep. **12**(1), 5650 (2022).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIOptoelectronicNeuralNetwork20260909.docx