nature communications

Article | Article in Press

https://doi.org/10.1038/s41467-026-79408-6

Structurally guided wavefront neural encoder for broad-range generalizable 3D holography

Received: 8 January 2026

Accepted: 20 October 2026

Published online: 16 November 2026

Check for updates

Yanan Zhang[1,5], Xingpeng Yan[1,2,5] ✉, Jiaqi Li[1], Zijian Shi[3], Yang Yu[1], Yang Zhang[1,4] & Tao Jing[1]

Deep learning has advanced computer-generated holography, but learned models remain coupled to fixed wavelengths and distances, limiting generalization. Here we present the structurally guided wavefront neural encoder (SGWN-Encoder), which integrates diffraction physics with neural networks and employs a structural propagation strategy that analytically strips the parameter-sensitive global phase. This reveals a scale-normalized description of diffraction, letting the network learn transferable light-field evolution rules and generalize across propagation distances and visible wavelengths by direct inference. A dataset-free paradigm built on statistically diverse random fields teaches the model the structural mapping from complex amplitude to phase-only holograms instead of external image distributions. Without fine-tuning, a single SGWN-Encoder maintains a distance-averaged RGB PSNR of 31.9 dB and SSIM of 0.91 over the primary 2–200 mm range, retaining reconstruction up to 500 mm. Numerical and optical experiments on full-color 3D scenes establish a physically grounded route toward adaptive holographic displays.uction reaction, highlighting the potential of DBCata as a powerful tool for catalyst design and optimization.

Computer-generated holography (CGH)[1–4] is capable of reconstructing wavefront information of target light fields. By providing depth cues, including binocular disparity and motion parallax, it can alleviate the vergence-accommodation conflict inherent in traditional 3D displays. Consequently, CGH is widely studied as a route toward next-generation true 3D displays[5–7]. It has application potential in fields such as virtual reality (VR)[8,9], augmented reality (AR)[10–14], automotive head-up displays[15,16], holographic printing[17], real-time holographic acquisition[18], and high-precision optical manipulation[19,20].

However, efficient generation of high-quality holograms remains a significant challenge. Traditional iterative optimization algorithms, such as the Gerchberg-Saxton algorithm[21] and Wirtinger holography[22], ensure reconstruction quality but are constrained by the high temporal costs of repetitive iterations. Conversely, non-iterative algorithms, including point-source[23], polygon-based[24], and layer-based methods[25], avoid cyclic optimization but still encounter substantial computational burdens when processing complex 3D scenes. Furthermore, they must address the encoding challenge of mapping complex amplitude light fields to phase-only holograms[26], which frequently comes at the cost of sacrificing the space-bandwidth product or introducing reconstruction artifacts. In general, existing traditional methods struggle to overcome the inherent computational bottleneck[27], making it difficult to balance real-time performance and reconstruction fidelity. Recently, the introduction of deep learning (DL)[28–30] has transformed the field of holography, advancing both digital holographic reconstruction[31] and CGH[32]. For CGH,

[1]Department of Information Communication, Army Arms University of PLA, Beijing, China. [2]National Key Laboratory of Intelligence Parallel Technology, Army Arms University of PLA, Beijing, China. [3]State Key Laboratory of Ultrafast Optical Science and Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. [4]School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. [5]These authors contributed equally: Yanan Zhang, Xingpeng Yan. ✉e-mail: yanxingpeng@tsinghua.org.cn

representative learning-based generation algorithms, including end-to-end networks and neural rendering-inspired methods[33–38], have made significant progress. Among them, end-to-end models, exemplified by TensorHolo[39,40], facilitate real-time generation by directly learning the scene-to-hologram mapping. In contrast, physics-driven models like HoloNet[41] incorporate diffraction algorithms, such as the angular spectrum method (ASM)[42,43], into the network, leveraging explicit physical constraints to improve reconstruction quality. Despite these advancements, existing deep learning-based methods remain constrained by a generalization bottleneck.

Specifically, the training of existing network models relies heavily on predefined physical parameters, such as propagation distances and wavelengths. This dependency constrains the network to learn static mapping relationships under specific physical configurations, rather than transferable light-field propagation rules. Consequently, once the optical setup in practical applications changes, the pre-trained weights may no longer match the new optical configuration, leading to degradation in reconstruction quality. Although simultaneous generalization across multiple physical parameters remains challenging, recent works have made progress in improving adaptability to physical configurations. Asano et al. proposed conditional neural holography (CNH)[44], which extends the HoloNet[41] architecture by using a zone-plate encoder to transform distance parameters into 2D physical feature maps for guided hologram generation. Although CNH achieves depth adaptation over a range of 200–500 mm, its reported degradation at near-field distances (< 200 mm) suggests a residual reliance on interpolation within specific physical priors. Similarly, Liu et al. proposed the SFO-solver[45], combining physical diffraction mechanisms with Fourier neural operators. By introducing propagation distance as a conditional input, they achieved continuous depth adaptation within a small range (e.g., 30 mm); however, this remains a local fitting within a limited parameter space. Furthermore, IncepHoloRGB[46] proposed by Yu et al. achieved single-model full-color generation by embedding a three-wavelength angular spectrum propagation model, yet its diffraction distance remains fixed. A unified treatment of both distance and wavelength adaptation remains insufficiently explored. This bottleneck is related to the mismatch between the dynamic evolution of optical diffraction and the static inductive bias of standard convolutional architectures and parameter-conditioned models. As light waves propagate in free space, their phase evolution rate and diffraction kernel scale undergo continuous scaling transformations with distance z and wavelength λ. In contrast, fixed local filters and limited parameter conditioning make it difficult to capture such physical feature drifts that span multiple orders of magnitude. Therefore, simple parameter conditioning often yields local interpolation rather than physical scale invariance.

To address these challenges, we propose the structurally guided wavefront neural encoder (SGWN-Encoder), a computational framework that combines structural propagation with an encoding network to achieve holographic generation across varying optical configurations with a single model. To reduce the mismatch between the dynamic evolution of diffraction and the static bias of neural networks, we propose a structural propagation strategy. By separating the distance-dependent phase-reference term while preserving wavefront structural information, this strategy normalizes the input feature space and provides stable complex-field inputs to the encoder. The key insight of this design lies in exploiting the scale structure of the diffraction process, allowing the network to learn a scale-normalized encoding rule associated with the dimensionless propagation factor z/λ under the reference configuration. This supports direct inference within the evaluated scalar-diffraction regime without retraining. Moreover, we establish a dataset-free learning paradigm that utilizes statistically diverse random fields to drive physical closed-loop training. This approach encourages the network to focus on learning the structural mapping from complex amplitude wavefronts to phase-only holograms, rather than fitting the specific distributions of external image datasets. Leveraging these mechanisms, the SGWN-Encoder operates across the evaluated visible wavelengths and propagation distances from 2 to 500 mm without fine-tuning. Over the primary 2–200 mm working range, our method maintains a distance-averaged RGB peak signal-to-noise ratio (PSNR) of 31.9 dB and a structural similarity (SSIM) of 0.91. Extended evaluations up to 500 mm show a gradual roll-off consistent with the expected finite-aperture diffraction limit. These results indicate that this single-model generalization operates within the valid bounds of the scalar-diffraction model and hardware constraints, rather than implying applicability beyond those conditions. Combined with the layer-by-layer replacement method (L2RM)[47], we realize full-color reconstruction of complex 3D scenes in both numerical simulations and optical experiments, indicating the potential of this approach for adaptive holographic computation.

Results

Structural propagation

To construct a holographic encoder adaptable to changing physical configurations, a key requirement is to address the discrepancy between optical diffraction models and neural network optimization mechanisms. This mismatch primarily arises from the rapid phase wrapping inherent in free-space propagation and the neural network's requirement for stable feature distributions. The ASM is a standard diffraction model that is accurate within the scalar Fourier-optics regime, but its full transfer function contains a global piston phase that changes rapidly with propagation distance and wavelength. As shown in Fig. 1a, applying the full transfer function H causes the propagated phase to contain both a structural phase component and a global phase offset. This offset is independent of the spatial frequencies (f_x, f_y), but it depends on z/λ and is therefore wavelength-dependent. Small distance or wavelength changes can introduce large phase wrapping that is not associated with object structure. For data-driven learning, this produces parameter-sensitive feature variations that increase the burden on the network.

To confront this challenge, we propose structural propagation. Based on the mathematical decomposition in Fig. 1b and Supplementary Material Note S1, we factor the transfer function as $H = H_g H_s$ and use the structural term H_s to construct the network input. The removed factor H_g is a unit-modulus phase reference, so removing it does not change the intensity distribution because $|H_g| = 1$. For direct optical intensity measurements and standard evaluation metrics (e.g., PSNR and SSIM), the global phase factor can be regarded as an unobservable phase gauge for intensity reconstruction. Therefore, structural propagation normalizes the diffraction calculation by fixing the phase reference frame, rather than modifying the physical laws themselves.

This operation first constructs a more stable learning environment. As shown in Fig. 1c, structural propagation forms a distance-normalized input field before encoding. To visualize this effect, Fig. 1d presents a one-dimensional propagation test using an amplitude line profile extracted from the target image. The same input field was propagated over small distance offsets ($\Delta z \in [0, 10]$ μm) using either the full ASM transfer function or the structural transfer function. The panels plot the center-field real response, selected phase profiles, and sampled complex-field values in the (Re, Im) plane. The comparison demonstrates that, relative to standard propagation, removing the piston phase aligns the phase profiles across distance offsets and yields a more compact complex-field distribution. This reduction in distance-induced feature variance provides the encoder with a consistent representation, allowing the network to process the structural component of the wavefront independent of global phase offsets.

It also uses the scale structure of diffraction. By introducing the normalized spatial frequency $\tilde{f} = \lambda f$, we rewrite H_s as a dimensionless

https://doi.org/10.1038/s41467-026-79408-6

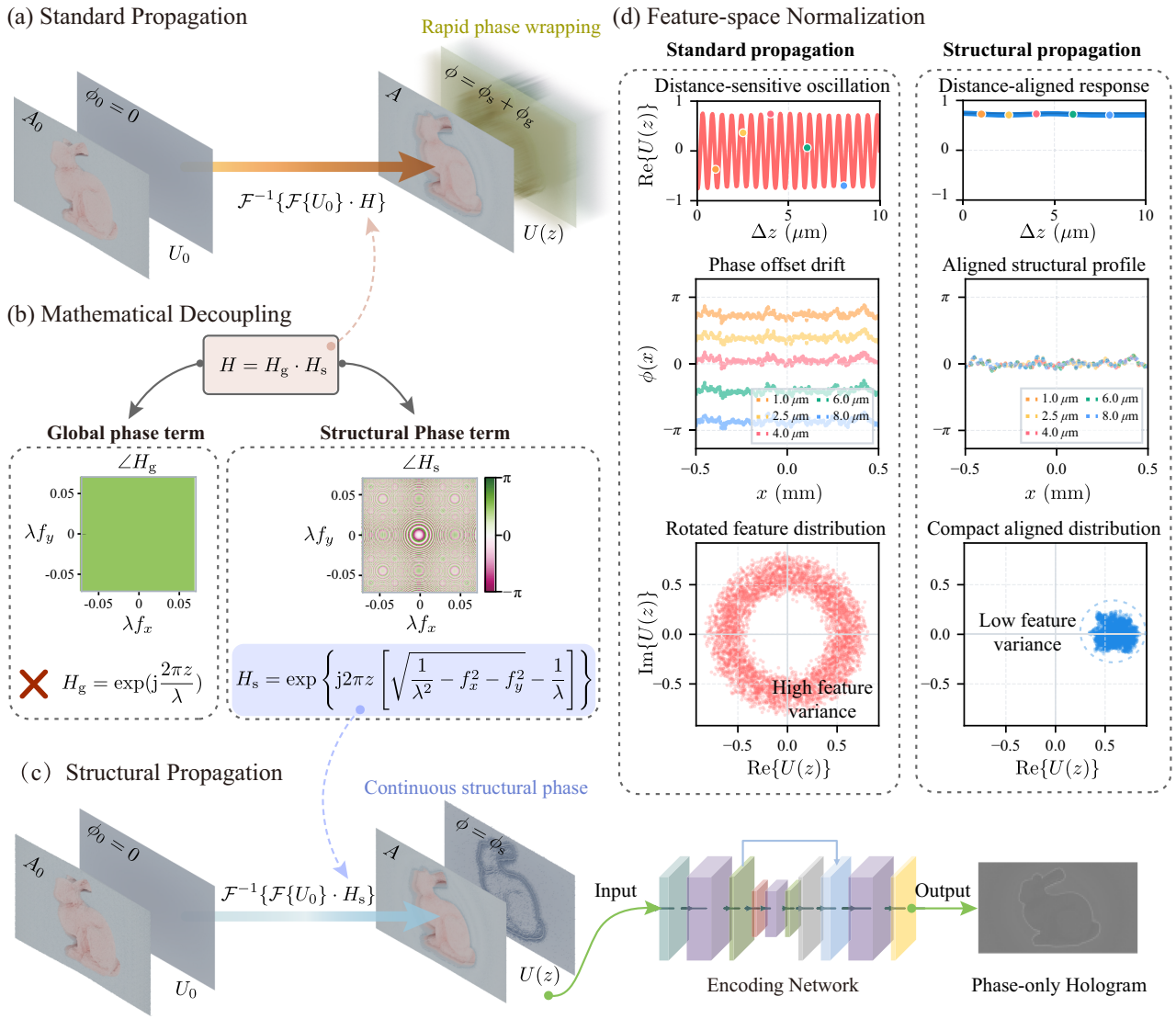

Fig. 1 | Structural propagation and feature-space normalization for parameter-adaptive hologram encoding. a In the standard angular-spectrum formulation, the full transfer function H introduces both a spatial-frequency-dependent structural phase and a distance- and wavelength-dependent global phase reference. The latter changes rapidly with z/λ, producing parameter-sensitive phase wrapping in the network input. **b** The transfer function is factorized into a unit-modulus global phase term H_g and a structural term H_s. Removing H_g fixes the phase reference frame while preserving the intensity distribution and structural diffraction features. **c** H_s is used to generate a normalized complex-field input for the encoding network. **d** Feature-space comparison showing that removing the global phase reference yields more compact and aligned complex-field distributions before encoding. The rabbit example is the Stanford Bunny from the Stanford 3D Scanning Repository (Stanford University Computer Graphics Laboratory, graphics.stanford.edu/data/3Dscanrep).

diffraction kernel controlled by the dimensionless factor $\mathcal{P} = z/\lambda$. After removing H_g, diffraction patterns under different physical configurations can be expressed in a normalized form governed by a scaling law. This implies that different combinations of (z, λ) can be described by a dimensionless evolution rule, specifically a geometric scaling process determined by the ratio z/λ. Under this framework, physical parameter changes during inference are represented primarily as smooth geometric expansions or contractions of wavefront textures rather than as large phase-reference shifts in the network input. This structural decoupling allows the network to reduce the memorization of absolute physical values and instead learn scale-normalized wavefront evolution rules.

Structurally guided dataset-free learning paradigm

Structural propagation defines how the network input is constructed; a separate question is what training signals should drive the learning. The training of traditional holographic neural networks typically relies on real or synthetic scene data as a source of supervision signals. Specifically, such methods often use images or RGB-D scenes[39,48] as proxies for data generation, producing corresponding complex amplitude wavefronts via equivalent wavefront modeling and forward propagation to construct training samples. However, since the neural network directly processes complex amplitude wavefronts rather than the scene itself during both training and inference, this indirect generation strategy, mediated by specific scenes, can entangle the learning of diffraction mappings with specific object structural priors, such as the low-frequency smoothness inherent in natural images. This implicit reliance on specific data distributions, recognized as a bottleneck in data-driven computational imaging[49], can restrict the network's ability to model transferable diffraction structures, leading to performance degradation when inferring out-of-distribution data, particularly binary targets containing abundant high-frequency step signals (as verified in Supplementary Material Note S6.1). In light of this, we construct a structurally guided dataset-free learning

https://doi.org/10.1038/s41467-026-79408-6

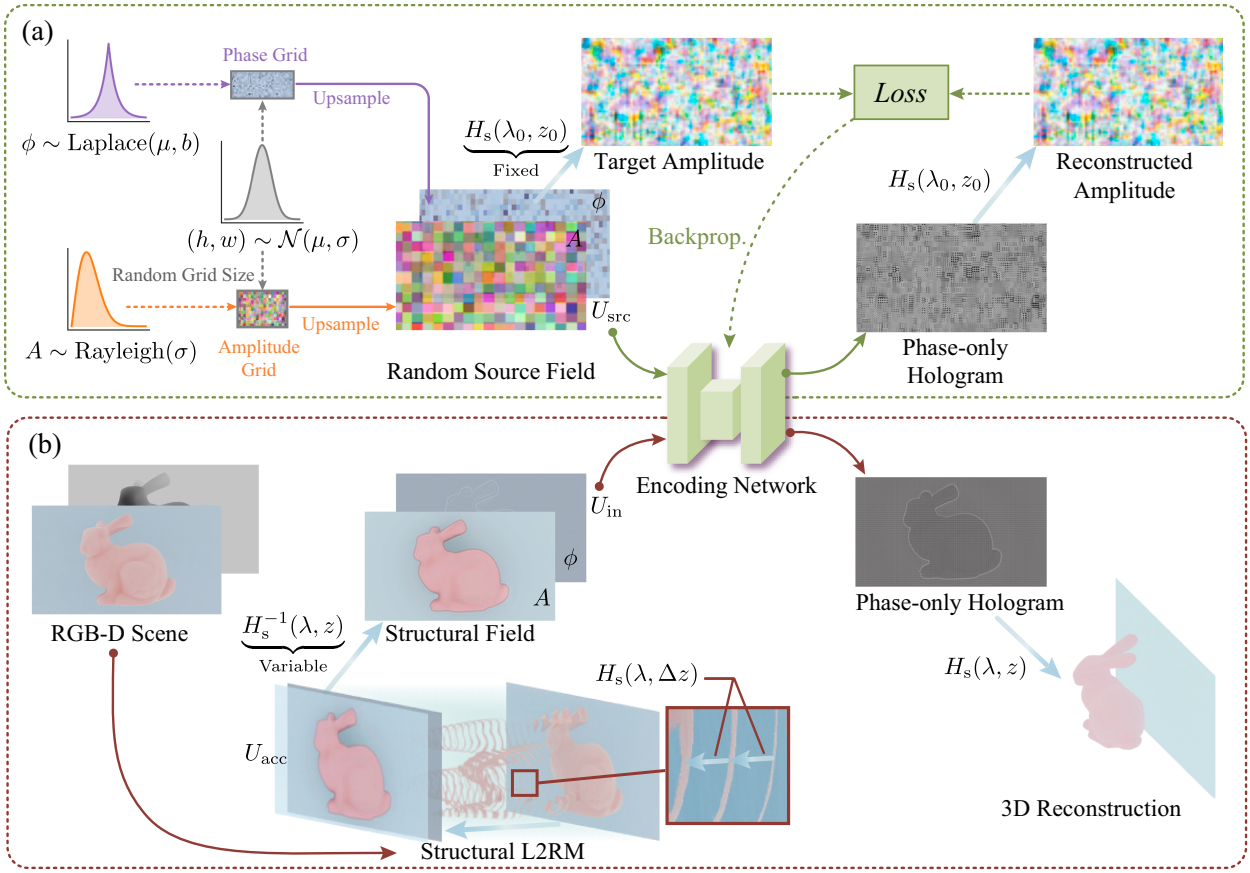

Fig. 2 | Structurally guided dataset-free learning and inference. a Dataset-free training constructs random source fields and optimizes the encoder through a structural-propagation closed loop under a single reference configuration (λ_0, z_0). The target and learning paths share the same synthetic source field. **b** During real-scene inference, the structural layer-by-layer replacement method (L2RM) accumulates RGB-D layers and back-traces the field to the hologram plane, enabling single-model hologram generation across the evaluated wavelengths and distances without retraining. The rabbit example is the Stanford Bunny from the Stanford 3D Scanning Repository (Stanford University Computer Graphics Laboratory, graphics.stanford.edu/data/3Dscanrep).

paradigm that reduces dependence on external image data (as shown in Fig. 2a).

Under this paradigm, the network is trained under only a single physical configuration (λ_0, z_0). The core insight is that diffraction propagation can be reformulated via the structural propagation operator as a geometric evolution process governed by scale-normalized structural parameters, so that diffraction processes under different physical configurations share a common scale-normalized form under compatible scalar-diffraction conditions. The goal of training is therefore not to fit specific instances of wavelengths or propagation distances, nor to traverse them during training, but to learn a scale-normalized structural mapping from complex amplitude to phase-only holograms that is decoupled from specific physical scales, and to apply this mapping to other compatible configurations through the scale-normalized representation. This setting reduces optimization variation caused by multi-configuration data and encourages the network to focus on learning the structural mapping of diffraction propagation rather than fitting numerical features under specific parameters.

In the training closed loop, a random source field U_{src} (generated with the multi-scale modulation strategy and statistical priors detailed in Methods) is simultaneously fed into dual pathways. In the target path, U_{src} propagates forward via the structural propagation operator to generate the corresponding diffraction field as the supervision signal. In the learning path, the same source field is directly input into the encoding network to predict a phase-only hologram, which is then numerically reconstructed via the structural propagation operator. Network parameters are optimized by minimizing the discrepancy between the reconstruction results of the two paths. Since no object semantics are introduced during training and physical parameters remain fixed, the network is biased toward learning stable mapping relationships under the dimensionless formulation defined by structural propagation. For the reported model, training was performed under a single reference configuration ($\lambda_0 = 520$ nm and $z_0 = 5$ mm; see Methods for the full training and optimization settings).

In the inference stage (as shown in Fig. 2b), the structural propagation operator is embedded into the L2RM framework to process real 3D scene light fields with varying wavelengths and propagation distances (detailed algorithmic flow is provided in Supplementary Material Note S3). Through the layer-wise accumulation and inverse back-tracing of structural propagation, light fields under different physical configurations are mapped into the normalized encoder input field U_{in} used during training. Since the statistical properties of this input field remain compatible with the dimensionless light-field structure learned from single-configuration training, the encoding network can apply the learned structural encoding mapping, enabling direct hologram inference for both 2D planes and 3D volumes across the evaluated wavelengths and diffraction distances without retraining.

Broad-range 2D generalization

All experiments below employed a single pre-trained SGWN-Encoder with a fixed set of weights, which directly generates holograms across wavelengths (R, G, B) and propagation distances without fine-tuning

or parameter-specific retraining. The core validation of our proposed SGWN-Encoder lies in reducing the dependence on fixed physical parameters. To this end, we compared SGWN-Encoder with representative neural holography baselines, including SFO-solver [45], HoloNet [41], and CNH [44], under the evaluation datasets and baseline protocols described in Supplementary Material Note S5. HoloNet represents a physics-driven neural holography model trained under fixed diffraction parameters, SFO-solver conditions hologram generation on the propagation distance z and is evaluated around its reported training window (approximately 85–115 mm), and CNH introduces a conditioned zone-plate parameterization with reported operation mainly over 200–500 mm and degradation at near-field distances below 200 mm. These baselines therefore provide complementary references for fixed-parameter, locally conditioned, and distance-adaptive neural holography.

In contrast, our method does not memorize image features at specific distances but learns a scale-normalized encoding rule from complex amplitude fields to phase-only holograms. Quantitatively, within the primary 2–200 mm portion of the range shown in Fig. 3a, the reported single model achieves a distance-averaged RGB PSNR of 31.9 dB and an SSIM of 0.91 on the 16-image 4K evaluation set. At the extended 500 mm boundary, the model retains recognizable reconstruction capability with a PSNR of 22.9 dB and an SSIM of 0.71. The corresponding lateral PSF analysis in Fig. 3b provides an object-independent physical-resolution evaluation and shows FWHM broadening with propagation distance. This gradual metric roll-off is consistent with finite-aperture diffraction and sampling-bandwidth limits, rather than indicating an abrupt model-specific breakdown. Accordingly, the 2–500 mm interval should be regarded as the main validated operating range, while shorter and longer distances are treated as boundary cases in Supplementary Material Note S8 rather than as open-ended extrapolation. A more complete lateral and axial PSF analysis, including wavelength-dependent ASM references, SBR statistics, and multi-depth axial responses, is provided in Supplementary Material Note S7. Furthermore, the normalized capability overview in Fig. 3c summarizes the system-level attributes of SGWN-Encoder: a single lightweight model supports native 4K resolution, shared RGB inference, and 3D RGB-D compatibility through structural L2RM, requiring substantially fewer trainable parameters than the compared methods. In this comparison, parameter compactness is normalized by inverse log parameter count. In terms of generation efficiency, the complete 2D single-channel pipeline runs at 43.45 FPS at native 4K resolution on a single NVIDIA A100 GPU (measurement settings are provided in Methods; detailed RGB and 3D runtime breakdowns are given in Supplementary Material Note S11).

The visual results of complex 2D natural scene reconstruction (Fig. 3d) and high-frequency binary targets (Fig. 3e) further translate numerical trends into consistent perceptual quality. Using the same SGWN-Encoder model, consistent texture details, intensity distributions, and background suppression are maintained across the near-field (5 mm), mid-field (100 mm), and far-field (200 mm). All baseline results are aligned with the ground truth according to their original reconstruction pipelines and valid output regions before RGB synthesis and metric computation. All RGB comparisons used wavelength-specific settings for each method; SGWN-Encoder shares weights across color channels, whereas the baselines follow their native RGB or channel-specific model configurations, as detailed in Supplementary Material Note S5.2. At its representative working distance of 100 mm, the SFO-solver effectively recovers natural-scene textures, but exhibits strong chromatic deviation on binary targets containing high-frequency step structures. HoloNet, evaluated at its fixed 200 mm configuration, shows high structural similarity and good texture recovery on natural images; however, for high-frequency targets, its reconstructions are affected by visible color aberrations, background speckle, and local interference artifacts. In addition, CNH shows limited fidelity on both natural scenes and binary targets at its operating distance, accompanied by detail blurring and color distortion. Overall, these visual results indicate that structural propagation preprocessing helps mitigate parameter-dependent phase-reference variations, enabling SGWN-Encoder to achieve more stable visual reconstruction quality across both natural textures and high-frequency structures. This is further corroborated by ablation studies in Supplementary Material Note S6, demonstrating that both the dataset-free learning strategy and structural propagation contribute to the observed advantages over baseline methods.

To ensure the statistical reliability of these findings, we extended our evaluation beyond the primary 16-image test set. As detailed in Supplementary Material Notes S9--S10, training stability across independent random seeds and large-scale statistical evaluations on the DIV2K-valid (N = 100), HFP-20 (High-Frequency Patterns 20; N = 20), and the SFO-benchmark (N = 10) support that the reported broad-range generalization is robust to initialization randomness and remains consistent across diverse image distributions.

3D scene reconstruction

In the previous sections, we evaluated the performance of the proposed SGWN-Encoder in distance and wavelength adaptability. This section evaluates its capability for reconstructing complex 3D scenes. We employed the encoder within the L2RM framework to perform holographic calculations for a color 3D kitchen scene comprising 8 depth layers, covering a range from 10 to 24 mm with a 2 mm interval, as shown in Fig. 4a (additional verification under diverse configurations is provided in Supplementary Material Note S6.3).

To investigate the mechanism enabling 3D generalization with a single model, we constructed a control experiment to visualize the input light fields received by the network under different propagation modes. As shown in Fig. 4c, when standard propagation was used during inference preprocessing, the layer-wise calculation of L2RM exposed a phase-reference mismatch: the distance-dependent piston phase introduced by the standard propagation operator amplified phase discontinuities at inter-layer occlusions and triggered strong amplitude ringing. Following the cumulative iteration over eight layers, the input field back-propagated to the hologram plane contained dense wrapped fringes in phase and pronounced ringing in amplitude. This perturbed input caused the network, which was trained under the single-plane dataset-free setting, to process phase-reference artifacts, leading to visible color distortion and spectral ripples in the standard propagation reconstruction results shown at the bottom of Fig. 4c. Notably, chromatic ringing artifacts were particularly pronounced in flat regions such as walls, yielding a PSNR of only 24.7 dB.

In contrast, Fig. 4d shows that structural propagation provides a more stable input representation for 3D encoding. By using the structural operator during inter-layer propagation and back-propagation, the field fed to the network retains the structural phase and amplitude information while avoiding the distance-dependent phase-reference variation introduced by the standard formulation. Specifically, the phase component changes from densely wrapped fringes to a more continuous structural distribution, while the amplitude component exhibits reduced ringing and better preserves scene textures. Consequently, the network can apply the learned complex-amplitude-to-phase mapping more consistently to the 3D input. The encoder outputs a phase-only hologram via a single forward inference, as shown at the bottom of Fig. 4d, with the reconstruction PSNR increasing to 36.1 dB.

Furthermore, the digital refocusing experiment in Fig. 4d indicates that the generated hologram contains valid 3D depth cues. By simulating the focusing process of the human eye at different depths, we can observe depth-of-field effects: when focusing on the bear sticker in the near-field at 10 mm, the towel in the mid-ground and the frying pan in the background exhibit natural defocus blur; as the focal

https://doi.org/10.1038/s41467-026-79408-6

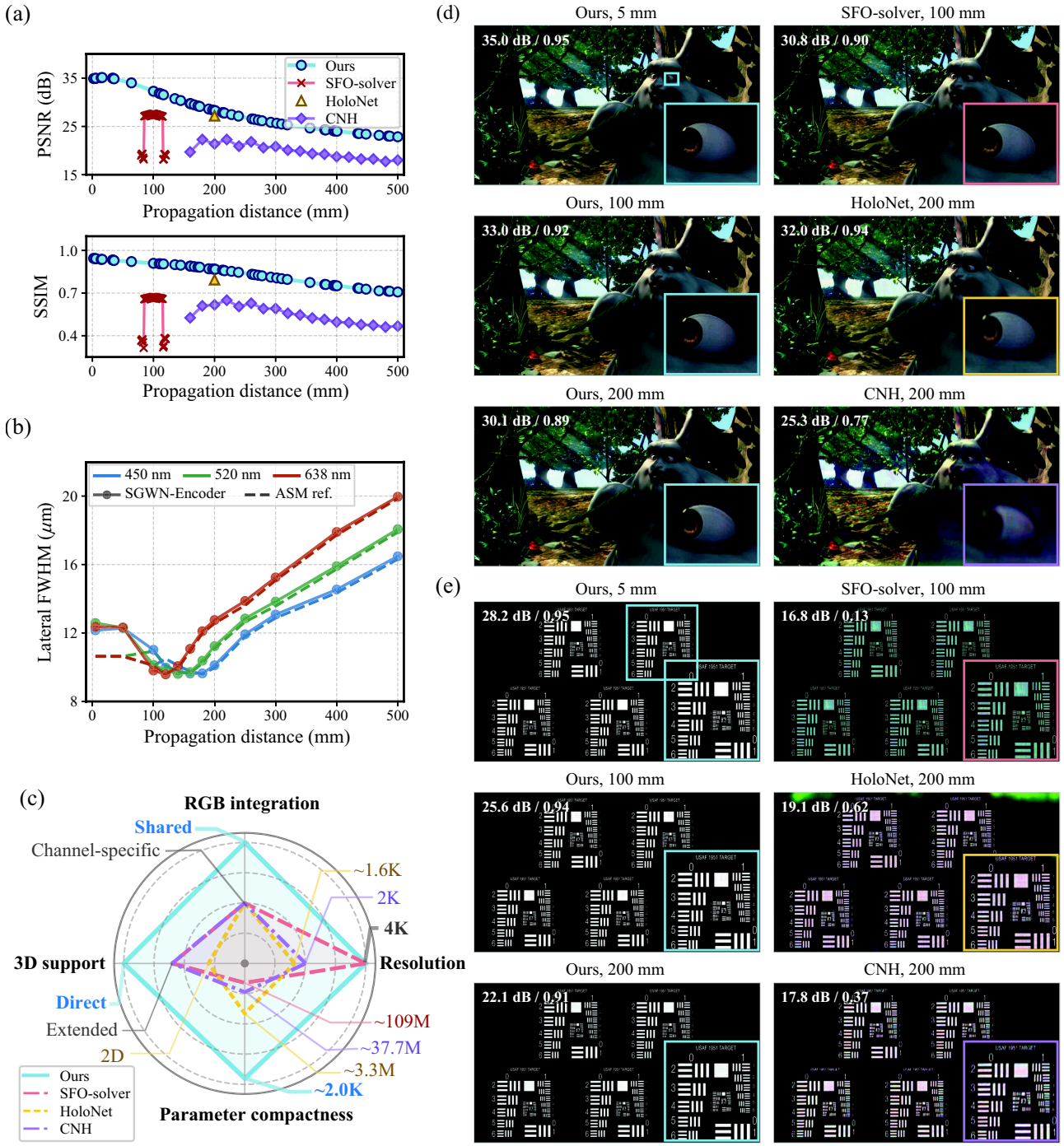

Fig. 3 | Broad-range 2D hologram generation and resolution characterization. **a** Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) over propagation distances from 2 to 500 mm compared with representative baselines. **b** Lateral point-spread function (PSF) full width at half maximum (FWHM) as an image-content-independent resolution metric, with angular spectrum method (ASM) references shown as dashed curves. **c** Normalized comparison of resolution, RGB integration, 3D support, and parameter compactness, where parameter compactness is normalized by inverse log parameter count. **d**,**e** Representative reconstructions of a natural scene and a high-frequency binary target; values denote RGB PSNR and SSIM. The natural-scene target in (**d**) is reproduced from www.bigbuckbunny.org(© 2008, Blender Foundation) under a Creative Commons licence (https://creativecommons.org/licenses/by/3.0/).

plane shifts to the mid-field at 16 mm and the far-field at 22 mm, the in-focus regions shift accordingly. This depth-of-field behavior, together with the axial PSF analysis in Supplementary Material Note S7, supports that the structural propagation strategy within the L2RM framework can process 3D depth information and encode it into a phase-only hologram using a single model. (For the comparative analysis of structural propagation in 2D scenarios, please refer to Supplementary Material Note S6.2.)

Optical reconstruction and experimental verification

For optical verification, we built a full-color holographic display setup (Fig. 5(a)); the optical configuration, registration strategy, and acquisition procedure are detailed in Methods and Supplementary Material Note S4. Based on this experimental platform, we loaded the generated holograms onto the spatial light modulator (SLM)[50] for optical reconstruction, with results shown in Fig. 5b–g. First, in the distance generalization test (Fig. 5b, c), the model showed depth adaptability

https://doi.org/10.1038/s41467-026-79408-6

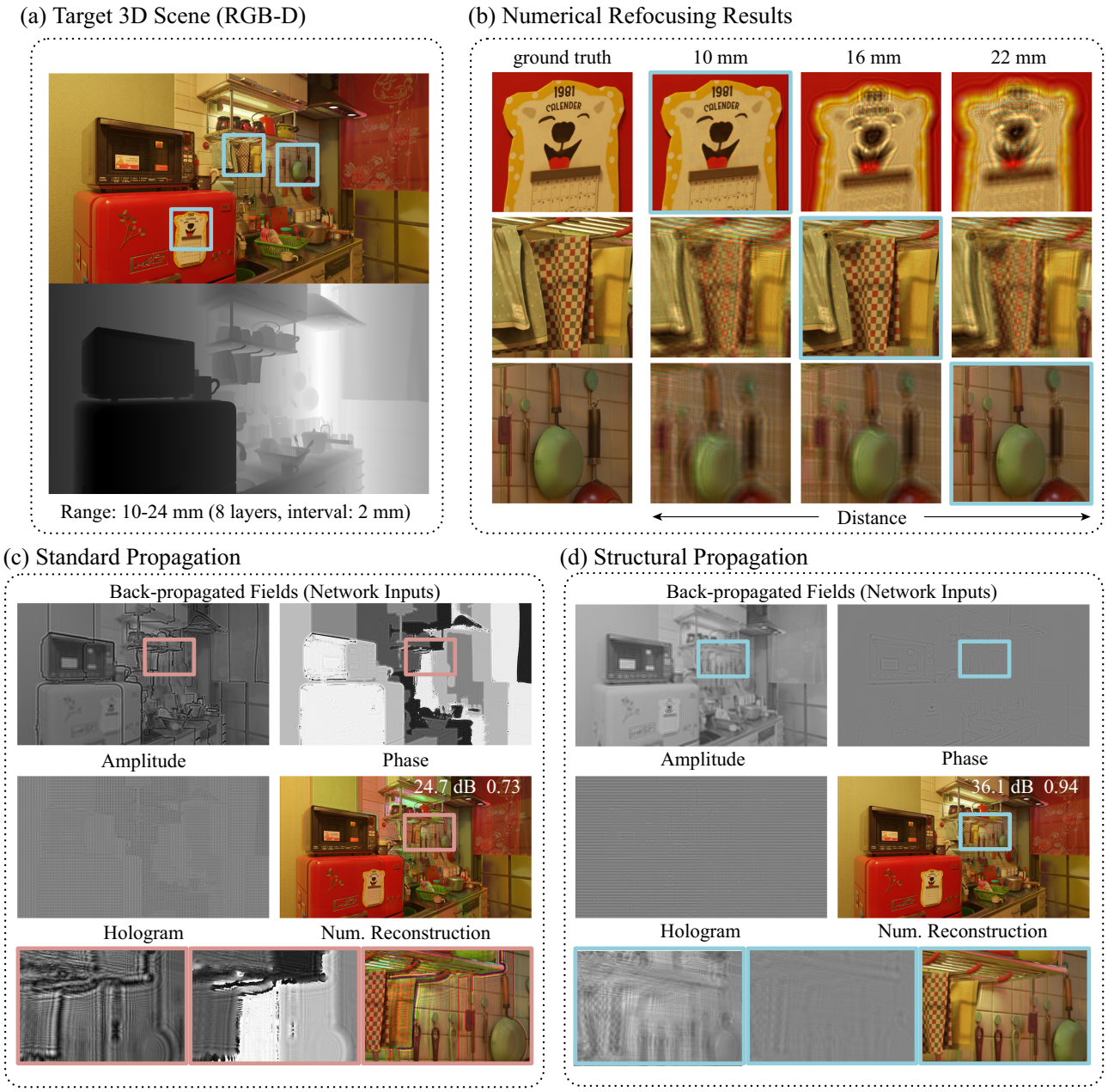

Fig. 4 | 3D holographic reconstruction and refocusing. a Target color scene and depth map, discretized into eight layers from 10 to 24 mm. **b** Digital refocusing at 10, 16, and 22 mm shows depth-selective focus and defocus from a single hologram. **c**,**d** Network input fields and reconstructions using standard propagation and structural propagation; structural propagation reduces phase-reference artifacts, with PSNR increasing from 24.7 to 36.1 dB.

over the evaluated range. Although a slight degradation in sharpness and minor interference fringes at the edges were observed as the propagation distance increased to 200 mm, physically attributed to the reduced effective numerical aperture and the enlargement of coherent speckle grain size, the overall structural features remained clearly legible, and the contrast was maintained. The three color channels also remained registered in physical space at long distances, reproducing vivid and natural colors without noticeable dispersion artifacts. These results support the experimental consistency of the encoder under multi-wavelength operation and the tested long-distance settings.

Furthermore, regarding the depth reconstruction of the 3D kitchen scene (Fig. 5d, e), by controlling the rail to position the acquisition module at 10, 16, and 22 mm, respectively, we successfully reproduced optical depth-of-field effects consistent with numerical simulations: objects at the focal plane appear sharp, while background areas exhibit natural physical defocusing. In contrast, Fig. 5e shows that artifacts associated with standard propagation became more apparent in real optical reconstruction. We observed that its results exhibited stronger color distortion and ringing artifacts than the numerical simulations. In the optical setup, this was accompanied by more visible channel mismatch, resulting in severe color deviation and speckle noise. The reproduction of 3D focusing relationships in real space supports that the single model can encode 3D structural depth into phase-only holograms and translate the numerical reconstruction behavior to physical display.

To further evaluate the model under complex configurations, we conducted extended tests. As shown in Fig. 5f, in the large depth-of-field harbor scene (comprising 64 dense depth layers spanning a physical range of 50–113 mm), the generated holograms preserved

https://doi.org/10.1038/s41467-026-79408-6

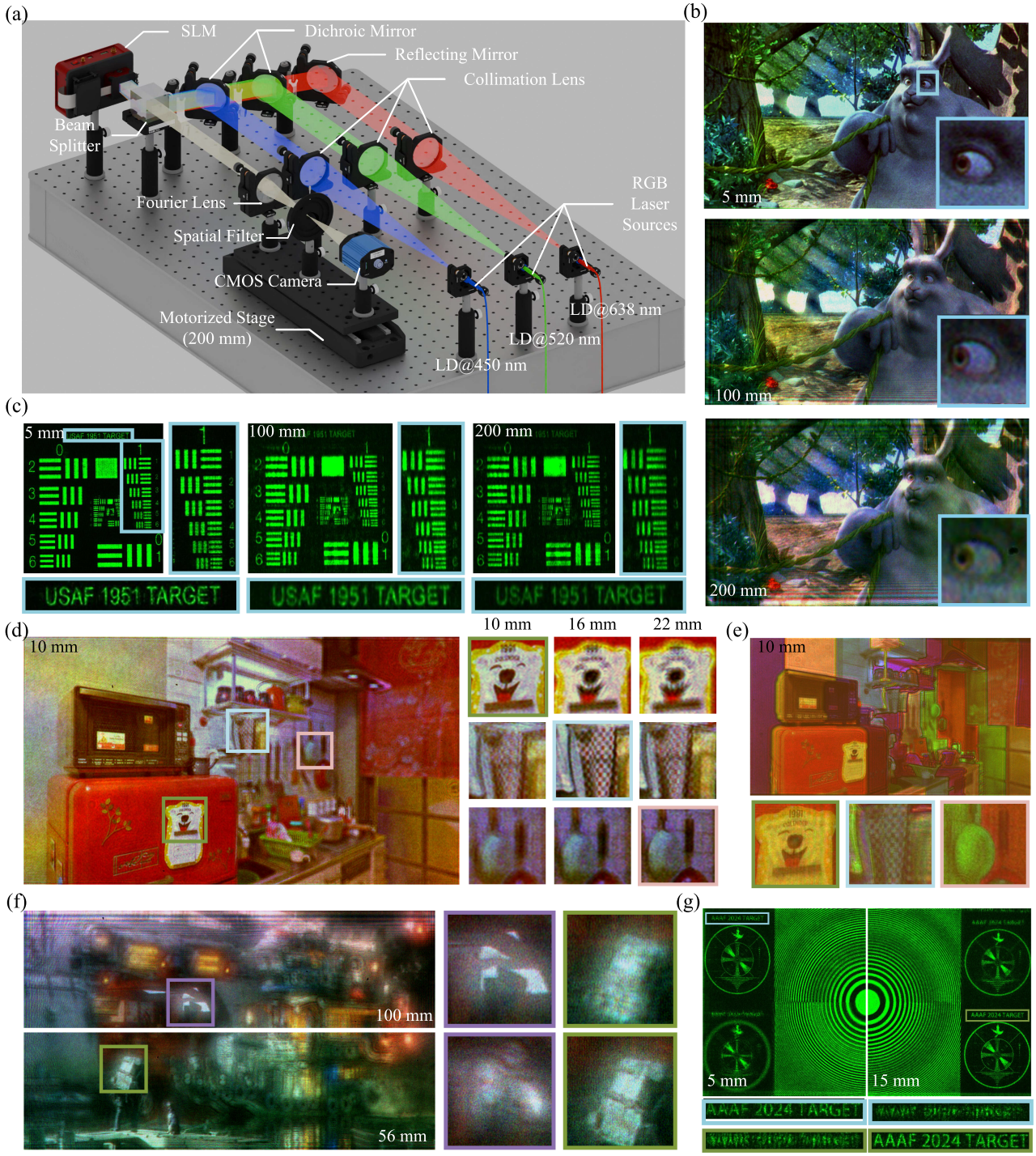

Fig. 5 | Full-color optical reconstruction. a Experimental holographic display setup with RGB illumination and spatial filtering. **b**,**c** 2D optical reconstructions of natural scenes and binary targets at representative near-, mid-, and far-field distances. **d**,**e** Kitchen-scene 3D reconstruction with structural and standard propagation. **f** Large-depth harbor-scene reconstruction at 100 and 56 mm. **g** 3D binary-target reconstruction showing focus isolation between depth layers. a CAD render based on Thorlabs, Inc., models EXULUS-HD3HP, DDS220, B6090B, MB1030/M, and other standard Thorlabs optomechanical components available via manufacturer website (https://www.thorlabs.com/). The natural-scene targets are frames from the open movie Big Buck Bunny, reproduced from www.bigbuckbunny.org(© 2008, Blender Foundation) under a Creative Commons Attribution 3.0 licence (https://creativecommons.org/licenses/by/3.0/). The harbor scene in (f) is the Blender 2.83 LTS splash screen "PartyTug 6:00AM" by Ian Hubert, reproduced from https://download.blender.org/archive/gallery/blender-splash-screens/blender-2-83-lts/(© Ian Hubert) under a Creative Commons Attribution-ShareAlike 4.0 licence (https://creativecommons.org/licenses/by-sa/4.0/). SLM, spatial light modulator.

extended depth cues. The distant harbor structures at 100 mm and the near-field dock-side figure at 56 mm could be independently focused, with the right insets showing correct focus and defocus effects over long distances. Meanwhile, the 3D binary target reconstruction results in Fig. 5g show that even for binary structures with high-frequency step signals, the method can maintain sharp edges and achieve correct focus isolation between different depth layers. For detailed depth configurations of these original scenes and

corresponding numerical simulation comparisons, please refer to Supplementary Material Note S6.3.

Discussion

This paper proposes a physics-adaptive holographic computing framework, SGWN-Encoder, addressing the parameter generalization challenge in deep learning holography. By factorizing the propagation operator and analytically removing the spatial-frequency-independent global phase from the network input, we construct a more stable physical learning space and reduce the dependence of network weights on fixed optical parameters. Building on this, the dataset-free learning paradigm reduces the inductive bias of specific datasets by using statistically diverse random fields to drive training, encouraging the model to learn physical wavefront-encoding rules rather than specific object priors. These two components work together to support generalization across physical configurations and object structures.

From a physical perspective, a key feature of our method is its exploitation of the intrinsic scale invariance of diffraction rather than direct fitting to each physical setting. As demonstrated in our comparative experiments, methods, such as SFO-solver[45], HoloNet[41], and CNH[44], while incorporating physical models or parameter-conditioning mechanisms, remain sensitive to the operating ranges defined by their training or calibration configurations. Consequently, their reported performance is tied to specific distance windows, fixed focal planes, or calibrated depth ranges. In contrast, structural propagation and dataset-free learning allow a single model to maintain competitive performance across diffraction distances from 2 to 500 mm. This shift from data fitting to physical encoding reduces the need for parameter-specific retraining and supports the integrated application of holographic technology in next-generation display terminals. Particularly for near-eye display systems based on holographic waveguides and automotive head-up displays, this method can, in principle, adapt to changing depth planes and field-of-view requirements without parameter-specific retraining, which may be useful for depth fusion between virtual information and real-world scenes in driving and near-eye display scenarios. Runtime analysis further shows that complete native-4K 2D single-channel hologram synthesis takes 23.02 ms (43.45 FPS) on an NVIDIA A100 GPU, supporting near-real-time single-channel 2D generation, although multi-layer wave propagation in structural L2RM currently remains the primary latency bottleneck (see Supplementary Material Note S11 for detailed runtime breakdown).

Nevertheless, there remains room for optimization. First, the propagation model used in this work should be understood as a scalar Fourier-optics model based on the ASM, rather than a full-wave electromagnetic solver[51]. It neglects evanescent waves, polarization-dependent vectorial effects, multiple scattering, material dispersion within the object, and sub-wavelength or very-high-curvature field variations. Therefore, the generated hologram should not be interpreted as a faithful full-wave image of a general object light field; its validity is limited to the sampled spatial-frequency bandwidth, finite aperture, and scalar-diffraction conditions considered in our simulations and optical experiments. Accordingly, the single-model generalization reported in this work should be understood as physically bounded. Structural propagation removes the spatial-frequency-independent piston phase, but H_s still contains propagation phase variations determined by z/λ and spatial frequency; therefore, at very high spatial frequencies or extreme distances, numerical errors and finite-aperture degradation will increase. The applicable range remains limited by finite aperture, sampled bandwidth, and phase-only encoding capacity. Second, regarding inference efficiency, although the L2RM framework accurately reconstructs 3D depth information, its computational cost increases linearly with the number of depth layers. Real-time generation remains a challenge when dealing with large depth-of-field scenes containing dense slices. Future work will explore more efficient parallel wavefront accumulation strategies to further enhance the generation rate of 3D holograms. Third, constrained by the pixel pitch of current SLMs and the inherent physical limits of diffraction principles, the field of view of the reconstructed images remains relatively small. Although this bottleneck stems primarily from hardware specifications, extending the effective visual range through algorithm-level field stitching or eye-tracking-based foveated viewing techniques remains a key step toward immersive applications. Furthermore, while structural propagation reduces ringing artifacts introduced by numerical computation, optical speckle noise inherent to laser holography and residual errors between the simulation model and the real optical path still exist. To further bridge this gap, we plan to introduce a camera-in-the-loop training strategy in future research. By feeding captured real-world optical reconstructions back into the network for fine-tuning, the model could learn and compensate for system-specific aberrations, which, combined with improved multi-wavelength calibration, is expected to improve the final visual experience.

In summary, the SGWN-Encoder shows the feasibility of a single-model hologram-generation framework within the evaluated scalar-diffraction and hardware regimes. By combining structural propagation with dataset-free training, the framework supports hologram generation across the tested wavelengths, propagation distances, and 3D scene configurations without parameter-specific retraining. Future work will focus on improving 3D generation speed, enlarging the effective field of view, and incorporating camera-in-the-loop calibration to further reduce the gap between numerical modeling and optical display. By embedding physical diffraction structure into the learning process, this framework provides a foundation for adaptive holographic computation, with potential applications spanning near-eye displays, automotive HUDs, and spatial computing interfaces.

Methods

Training configuration and dataset-free optimization

The encoder architecture, an improved lightweight U-Net with trigonometric enhancement blocks, is described in Supplementary Material Note S2. The SGWN-Encoder was trained at a resolution of 1080×1920. Random source fields were generated online to provide broad coverage of structural features under the single reference configuration: a multi-scale dynamic grid modulation strategy was used to characterize the continuous variation of light fields from smooth envelopes to complex textures, and statistical physical priors were introduced by setting the amplitude component to follow a Rayleigh distribution (simulating fully developed speckle-like amplitude fluctuations) and the phase component to follow a Laplace distribution (characterizing sparse phase changes caused by structural edges and singularities). Each epoch used 400 online-generated random complex fields, and the network was trained for 50 epochs, corresponding to 20,000 independent training fields. Optimization used Adam with a batch size of 1 and an initial learning rate of 1×10^{-3} with step decay. The training convergence curves, checkpoint selection rule, loss definition, and full hyperparameter table are provided in Supplementary Material Note S2.

Evaluation datasets, metrics, and runtime

Reconstruction quality was quantified using PSNR and SSIM computed on RGB-synthesized reconstructions. The primary evaluation used a 16-image 4K test set; extended statistical evaluations covered the DIV2K-valid (N = 100), HFP-20 (High-Frequency Patterns 20; N = 20), and SFO-benchmark (N = 10) sets. Evaluation datasets, baseline protocols, and per-method RGB configurations are described in Supplementary Material Note S5, and the extended statistical analyses are reported in Supplementary Material Notes S9--S10.

Hologram synthesis time was measured at native 4K resolution (3840 × 2160) on a single NVIDIA A100 GPU. The complete 2D single-channel pipeline takes 23.02 ms per hologram (43.45 FPS), including structural propagation and network inference. The RGB 2D pipeline takes 68.97 ms (14.50 FPS). Detailed RGB and 3D runtime breakdowns are provided in Supplementary Material Note S11.

Optical experimental setup

For full-color optical reconstruction, the RGB channels were captured separately and synthesized offline. The three-color beams were registered at the -1 diffraction order through optical alignment, and spatial filtering was applied to suppress zero-order direct light. To capture reconstructed images at varying depths in the real optical path, the integrated image acquisition module was mounted on a precision motorized rail with a 200 mm travel range and translated along the optical axis to record the 3D light field at different depths. Detailed optical configuration and multi-wavelength alignment strategies are provided in Supplementary Material Note S4.

Data availability

The datasets generated and analyzed in this study, including the evaluation images, scene inputs, phase-only holograms, numerical and optical reconstructions, and the source metrics supporting all figures, have been deposited in figshare and are publicly available at https://doi.org/10.6084/m9.figshare.32881460[52]. The pre-trained model weights and random field generation parameters are available in the code repository listed below.

Code availability

The source code for the SGWN-Encoder, including the implementation of the structural propagation operator, the dataset-free training pipeline, and the pre-trained model weights, is publicly available at https://github.com/YananZhang1030/SGWN-Encoderand archived on Zenodo[53].

References

1. Gabor, D. A new microscopic principle. *Nature* **161**, 777 (1948).
2. Goodman, J. W. Digital Image Formation From Electronically Detected Holograms. In *Computerized Imaging Techniques*, vol. 0010, 176–181 (SPIE, 1967).
3. Park, J.-H. Recent progress in computer-generated holography for three-dimensional scenes. *J. Inf. Disp.* **18**, 1–12 (2017).
4. Sui, X., He, Z., Chu, D. & Cao, L. Non-convex optimization for inverse problem solving in computer-generated holography. *Light. Sci. Appl.* **13**, 158 (2024).
5. Blanche, P.-A. Holography, and the future of 3D display. *Light.: Adv. Manuf.* **2**, 446–459 (2021).
6. Chang, C. et al. Split Lohmann computer holography: Fast generation of 3D hologram in single-step diffraction calculation. *Adv. Photonics Nexus***3**, 036001 (2024).
7. Wang, D. et al. Decimeter-depth and polarization addressable color 3D meta-holography. *Nat. Commun.* **15**, 8242 (2024).
8. Xiong, J., Hsiang, E.-L., He, Z., Zhan, T. & Wu, S.-T. Augmented reality and virtual reality displays: Emerging technologies and future perspectives. *Light. Sci. Appl.* **10**, 216 (2021).
9. Park, J.-H. & Lee, B. Holographic techniques for augmented reality and virtual reality near-eye displays. *Light.: Adv. Manuf.* **3**, 137–150 (2022).
10. Chang, C., Bang, K., Wetzstein, G., Lee, B. & Gao, L. Toward the next-generation VR/AR optics: A review of holographic near-eye displays from a human-centric perspective. *Optica* **7**, 1563–1578 (2020).
11. Martinez, C., Krotov, V., Meynard, B. & Fowler, D. See-through holographic retinal projection display concept. *Optica* **5**, 1200–1209 (2018).
12. Sando, Y., Satoh, K., Barada, D. & Yatagai, T. Holographic augmented reality display with conical holographic optical element for wide viewing zone. *Light. Adv. Manuf.* **3**, 26–34 (2022).
13. Gopakumar, M. et al. Full-colour 3D holographic augmented-reality displays with metasurface waveguides. *Nature* **629**, 791–797 (2024).
14. Jang, C., Bang, K., Chae, M., Lee, B. & Lanman, D. Waveguide holography for 3D augmented reality glasses. *Nat. Commun.* **15**, 66 (2024).
15. Skirnewskaja, J. & Wilkinson, T. D. Automotive holographic head-up displays. *Adv. Mater.* **34**, 2110463 (2022).
16. Moon, W., Jeon, H., Na, S., Kim, H. & Hahn, J. Accommodation-capable holographic waveguide head-up display with extended field of view. *Displays* **81**, 102617 (2024).
17. Yan, X. et al. Hexagonal-hogel holographic stereograms based on neural graphics primitives. *Appl. Soft Comput.* **182**, 113500 (2025).
18. Li, Z.-S. et al. Real-time holographic camera for obtaining real 3D scene hologram. *Light. Sci. Appl.* **14**, 74 (2025).
19. Leite, I. T. et al. Three-dimensional holographic optical manipulation through a high-numerical-aperture soft-glass multimode fibre. *Nat. Photonics* **12**, 33–39 (2018).
20. Shen, Y. et al. Optical vortices 30 years on: OAM manipulation from topological charge to multiple singularities. *Light. Sci. Appl.* **8**, 90 (2019).
21. Gerchberg, R. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **37**, 237–246 (1972).
22. Chakravarthula, P., Peng, Y., Kollin, J., Fuchs, H. & Heide, F. Wirtinger holography for near-eye displays. *ACM Trans. Graph.* **38**, 213:1–213:13 (2019).
23. Lucente, M. E. Interactive computation of holograms using a look-up table. *J. Electron. Imaging* **2**, 28–34 (1993).
24. Matsushima, K., Nakamura, M. & Nakahara, S. Silhouette method for hidden surface removal in computer holography and its acceleration using the switch-back technique. *Opt. Express* **22**, 24450–24465 (2014).
25. Chen, J.-S. & Chu, D. P. Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications. *Opt. Express* **23**, 18143–18155 (2015).
26. Arrizón, V. & Sánchez-de-la-Llave, D. Double-phase holograms implemented with phase-only spatial light modulators: performance evaluation and improvement. *Appl. Opt.* **41**, 3436–3447 (2002).
27. Pi, D., Liu, J. & Wang, Y. Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display. *Light. Sci. Appl.* **11**, 231 (2022).
28. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
29. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778 (CVPR, 2016).
30. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
31. Bai, Z. et al. Diff-Holo: a residual diffusion model with complex transformer for rapid single-frame hologram reconstruction. *IEEE Trans. Comput. Imaging* **11**, 689–703 (2025).
32. Yu, X. et al. On the use of deep learning for computer-generated holography. *iScience* **28**, 112507 (2025).
33. Wu, J., Liu, K., Sui, X. & Cao, L. High-speed computer-generated holography using an autoencoder-based deep neural network. *Opt. Lett.* **46**, 2908–2911 (2021).

34. Choi, S., Chao, B., Yang, J., Gopakumar, M. & Wetzstein, G. Gaussian wave splatting for computer-generated holography. *ACM Trans. Graph.* **44**, 57:1–57:13 (2025).
35. Dong, Z., Ling, Y., Li, Y. & Su, Y. Motion Hologram: Jointly optimized hologram generation and motion planning for photorealistic 3D displays via reinforcement learning. *Sci. Adv.* **11**, eads9876 (2025).
36. Liu, K., Wu, J., He, Z. & Cao, L. 4K-DMDNet: diffraction model-driven network for 4K computer-generated holography. *Opto-Electron. Adv.* **6**, 220135–220135 (2023).
37. Yan, X. et al. Generation of multiple-depth 3D computer-generated holograms from 2D-image-datasets trained CNN. *Adv. Sci.* **12**, 2408610 (2025).
38. Chang, C. et al. Conversion of 2D picture to color 3D holography using end-to-end convolutional neural network. *PhotoniX* **6**, 30 (2025).
39. Shi, L., Li, B., Kim, C., Kellnhofer, P. & Matusik, W. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).
40. Shi, L., Li, B. & Matusik, W. End-to-end learning of 3D phase-only holograms for holographic display. *Light. Sci. Appl.* **11**, 247 (2022).
41. Peng, Y., Choi, S., Padmanaban, N. & Wetzstein, G. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* **39**, 185:1–185:14 (2020).
42. Matsushima, K., Schimmel, H. & Wyrowski, F. Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves. *JOSA A* **20**, 1755–1762 (2003).
43. Yu, X., Xiahui, T., Yingxiong, Q., Hao, P. & Wei, W. Band-limited angular spectrum numerical propagation method with selective scaling of observation window size and sample number. *JOSA A* **29**, 2415–2420 (2012).
44. Asano, Y., Yamamoto, K., Fushimi, T. & Ochiai, Y. Conditional neural holography: a distance-adaptive CGH generator. *Opt. Express* **33**, 16671–16683 (2025).
45. Liu, N., Liu, K., Yang, Y., Peng, Y. & Cao, L. Propagation-adaptive 4K computer-generated holography using physics-constrained spatial and Fourier neural operator. *Nat. Commun.* **16**, 7761 (2025).
46. Yu, X. et al. IncepHoloRGB: multi-wavelength network model for full-color 3D computer-generated holography. *Opto-Electron. Adv.* **8**, 250130–11 (2025).
47. Yan, X. et al. Generating multi-depth 3D holograms using a fully convolutional neural network. *Adv. Sci.* **11**, 2308886 (2024).
48. Agustsson, E. & Timofte, R. NTIRE 2017 challenge on single image super-resolution: dataset and study. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1122–1131 (IEEE, 2017).
49. Wang, F., Czarske, J. W. & Situ, G. Deep learning for computational imaging: from data-driven to physics-enhanced approaches. *Adv. Photonics* **7**, 054002 (2025).
50. Yang, Y., Forbes, A. & Cao, L. A review of liquid crystal spatial light modulators: devices and applications. *Opto-Electron. Sci.* **2**, 230026–29 (2023).
51. Goodman, J. W. Introduction to Fourier Optics (Roberts and Company, 2005).
52. Zhang, Y. et al. SGWN-Encoder: data for the manuscript Structurally Guided Wavefront Neural Encoder for Broad-Range Generalizable 3D Holography. figshare. https://doi.org/10.6084/m9.figshare.32881460 (2026).
53. Zhang, Y. et al. SGWN-Encoder: code for the manuscript Structurally Guided Wavefront Neural Encoder for Broad-Range Generalizable 3D Holography. Zenodo. https://doi.org/10.5281/zenodo.23039120 (2026).

Acknowledgements

The authors thank all members of the research group for their support.

Author contributions

X.Y. and Y.N.Z. conceived the concept and designed the SGWN-Encoder architecture. Y.N.Z. derived the theoretical model of structural propagation and performed the numerical simulations. Y.N.Z., J.L., and Y.Z. constructed the optical setup and conducted the optical experiments. Z.S. and Y.Y. assisted with the data analysis and baseline comparisons. T.J. contributed to the code optimization and figure preparation. X.Y. supervised the project. X.Y. and Y.N.Z. wrote the manuscript with inputs from all authors.

Funding

X.Y. discloses support for the research of this work from the National Natural Science Foundation of China (No. 62375286) and the Outstanding Young Scholar Support Project (No. 2025CJ01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-79408-6.

Correspondence and requests for materials should be addressed to Xingpeng Yan.

Peer review information *Nature Communications* thanks Aleksi Tamminen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.