SUPPLEMENTARY INFORMATION: Towards Edge Holography via Implicit Neural Representation and Compression

Hyunmin Ban* Wenbin Zhou† Yifan Peng‡

The University of Hong Kong

S1 Mathematics Elaboration

S1.1 Near-field Wave Propagation in Free Space

To model the propagation between two parallel planes, we utilize the well-established angular spectrum method (ASM) [7]. More generally, let $u(x,y;z=0)$ denote the complex wavefront at the source plane $z=0$ (assuming the uniform amplitude for brevity) and $u(x,y;z=z_0)$ denote the wavefront at the target plane. We can numerically simulate the propagation by a convolution, which can be further accelerated by the Fourier transform as:

$$u(x,y;z=z_0) = \mathscr{F}^{-1}\{\mathscr{F}\{u(x,y;z=0)\}H(f_x,f_y)\}, \tag{S1}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ are the Fourier transform and inverse Fourier transform, respectively. $H(f_x,f_y)$ is the transfer function of propagation. In this work, H is defined as:

$$H(f_x,f_y) = \mathrm{e}^{\mathrm{j}2\pi\frac{z}{\lambda}\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}}, \tag{S2}$$

where λ indicates the wavelength employed.

S2 Algorithm Details

S2.1 Camera-calibrated Wave Propagation Model Architecture Details

We adopt the learned wave propagation model akin to the one studied by Choi et al. [3]. As shown in Fig. S1, the input phase is preceded by the first UNet, which addresses the SLM crosstalk/phase non-uniformity and predicts the complex-valued field on the SLM plane. Then, the ASM is adopted to propagate this field to 5 target planes. Next, the resulting 5 target fields are passed through the second UNet and a phase low-pass filter to account for the optical aberrations and limited bandwidth of the system. Eventually, the output fields are supervised by the captured dataset introduced in Sec. S4.2.

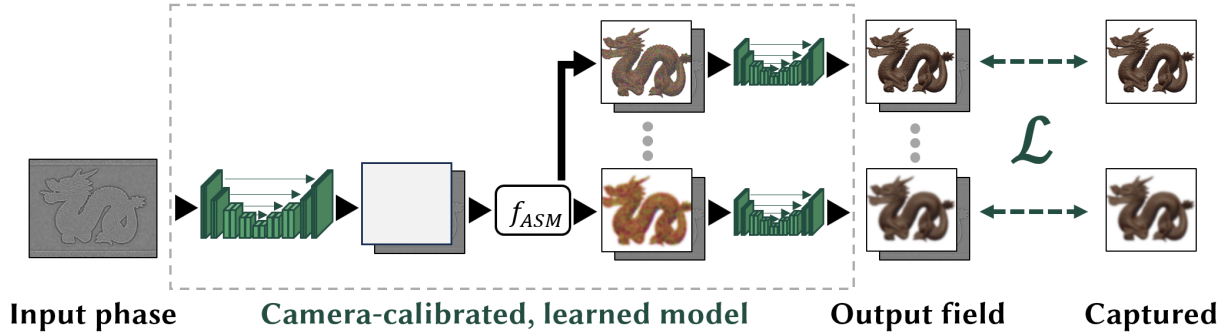

Figure S1: Illustration of the camera-calibrated, learned wave propagation model training for 3D (multi-plane) CGH.

Both UNets have 2 input and 2 output channels that accommodate the real and imaginary components of the fields. The first UNet comprises 8 sequential downsampling layers using strided convolutions, followed by 8 corresponding upsampling layers implemented through transposed convolutions. The initial input layer is followed by 32 feature channels, which double with each subsequent downsampling operation, reaching a maximum of 512 channels. The second UNet consists of 5 downsampling and 5 upsampling layers, beginning with 8 feature channels, which double progressively with each downsampling stage, culminating in a maximum of 128 channels. Both sub-networks apply instance normalization, Leaky ReLU activations for downsampling blocks, and ReLU non-linearities for upsampling blocks.

*e-mail: hmban@connect.hku.hk
†e-mail: zhouwb@connect.hku.hk
‡e-mail: evanpeng@hku.hk

S3 BENCH-TOP UNFILTERED HOLOGRAPHIC DISPLAY PROTOTYPE

We present additional details of our holographic display setup, as shown in main manuscript Fig. 4. The detailed specifications of the main components are listed in Tab. S1. Specifically, we use a fiber-coupled laser and a phase-only LCoS-SLM. The collimated laser light illuminates the SLM through a beam-splitter. After SLM modulation, light is magnified by an eyepiece and imaged by a camera. A linear polarizer optimizes the beam's polarization for the SLM, and a neutral density filter controls the intensity at 70% power, at which level laser output is stable. An Arduino controls the camera lens focus for multi-plane image acquisition.

Table S1: Specifications of our holographic display prototype built using off-the-shelf parts.

Experimental devices	Parameters
Fiber-coupled laser	Wavelengths: 639.0, 524.9, 457.0 nm
Collimation lens	Focal length: 300 mm
Phase-only SLM	Holoeye Pluto 2.1
	Resolution: 1,920 $\times$ 1,080
	Pixel size: 8 μm
Eyepiece	Focal length: 50 mm
Camera lens	Focal length: 35 mm
	Focusing range: 0.45 m$\sim$infinity
Camera sensor	FLIR GS3-U3-51S5M-C
	Resolution: 2,448 $\times$ 2,048
	Pixel size: 3.45 μm

S3.1 Supplementary Video of Conceptual Edge Holography and Experimental Results

We provide a supplementary video, named *edgeholo.mp4*, over-viewing the motivation of our edge INR-CGH paradigm and its experimental results compared with several representative baselines. Code and training data will also be released.

S3.2 Clarification on Abrupt Laser Imperfections in Red and/or Blue Outputs

Careful readers may notice that several of our experimental reconstructions, for instance, those in Fig. 8 to Fig. 10 of main text, exhibit reasonably fine details and "3D (RGBD)" behavior, yet still suffer from certain amount of color artifacts. We acknowledge that this is unfortunately due to a last-minute abrupt performance drop in the laser's red and/or channels. Nevertheless, we have demonstrated through green-channel reconstruction results (e.g., Fig. 6) that our INR-CGH can realize a good compromise between image fidelity and compression rate, evidenced by both visual results and qualitative assessments (refer to Sec. 5 of main text).

As such, we have chosen to not hide it and still show these color results for visualization purposes. We believe this subtle engineering issue and reproducibility concern can be easily resolved within the revision period. We would like to kindly note that experimentally showcasing perfect color results is orthogonal to the scope of this work. Our focus lies in exploring INR-represented hologram generation and compression for the future deployment of edge holographic displays within the cloud-edge computing mechanism.

S4 ADDITIONAL TRAINING DETAILS

S4.1 Distance Mapping and Discretization

In our experiments, all 3D holographic results are positioned and captured at 5 target planes, equally spaced in the dioptric space throughout 0.84 $m \sim$ *infinity* from the camera. The inter-plane distance is 0.31 diopter, which corresponds to the depth of field of the human eye [2, 6] and can therefore be considered approximately continuous in depth. We measured these distances, corresponding to 10.00, 10.07, 10.14, 10.21, and 10.28 cm away from the SLM physically and infinity, 3.52, 1.74, 1.14, and 0.84 m away from the camera, mapped by the eyepiece.

S4.2 Training Details of the Camera-calibrated, Learned Wave Propagation Model

To train the camera-calibrated, learned wave propagation model, we sequentially display many phase-only holograms on the SLM and capture the resulting images on 5 target planes using the focus-tuneable camera in our holographic display prototype. To train a model better aligns with the phase patterns produced by the proposed algorithm, we generate 10,500 phase patterns and captures the resulting 52,500 images at 5 target planes. The phase patterns are derived from 2D images in the DIV2K [1] dataset at various depths, utilizing the proposed INR-based method optimized for 2,000 iterations with naive ASM propagator. After the first round of capturing and training, the trained wave propagation model is adopted to replace the ASM and generate a second round of phase patterns, facilitating further refinement of the learned neural network-based propagator.

S4.3 Evaluation Sample Images

To assess the compression performance for *three* compression pipelines: INR + CNNpropCNN (ours), SGD + CNNpropCNN + JPEG, and SGD + CNNpropCNN + JPEG2000, we conduct experiments over a subset of ten natural images selected from the investigated dataset. These sample images, corresponding to Fig. 5 in the main manuscript, are shown in Fig. S2.

S5 TOWARDS COMPACT FORM-FACTOR WEARABLE HOLOGRAPHIC DISPLAYS

Recent advances in compact wearable holographic displays emphasize lightweight, high-resolution 3D visual experiences by merging holography with emerging optics/photonics [8]. Early foundational work highlighted the potential of holographic techniques to replace conventional bulk optics through wavefront reconstruction [5], enabling optical see-through capabilities and glasses-like form factors, though limited by speckle and color fidelity issues.

Figure S2: Testing images employed in compression performance evaluation for three compression pipelines: INR + CNNpropCNN (ours), SGD + CNNpropCNN + JPEG, and SGD + CNNpropCNN + JPEG2000.

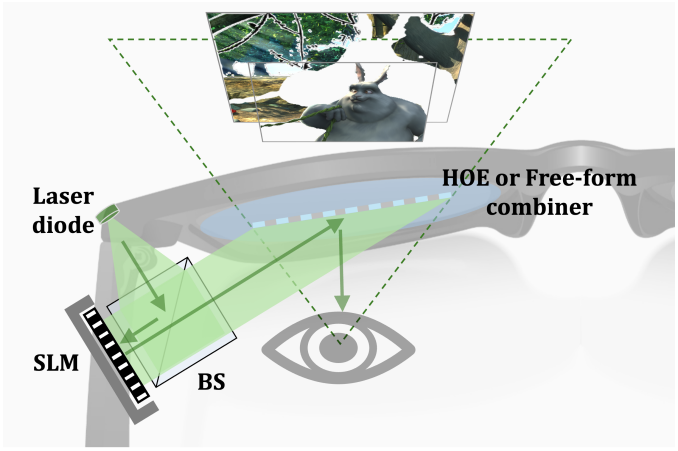

Figure S3: Conceptual illustration of compact form-factor wearable holographic displays, where the laser diode, SLM, and BS could be integrated into the wearable frame, and the extra physical filtering part is not necessary, as demonstrated in our benchtop prototype.

A significant leap was achieved with waveguide holography [4], which integrates exit-pupil-expanding waveguides with SLMs to deliver 3D holographic images while mitigating VAC. By modeling coherent light propagation in waveguides and leveraging pupil replication for étendue expansion, this approach could enable software-steerable eyeboxes without mechanical adjustments, though noticeable challenges in calibration and computational complexity remain. The most recent breakthrough by Gopakumar et al. gopakumar2024full synergized inverse-designed metasurface gratings with neural holography algorithms. Their co-designed system has replaced bulky collimation optics with dispersion-compensating metasurface waveguides, enabling full-color 3D content with depth-accurate defocus blur through AI-calibrated camera-in-the-loop optimization [3].

We note that previous work focusing on compact form factors has neglected the fact of demanding a potentially large computational processor. The algorithms employed to synthesize phase-only holograms typically struggle to run in real time (i.e., $\geq$30 Hz in RGB colors), even on state-of-the-art consumer-level GPUs (e.g., RTX 3090). Notably, the objective of our work is to explore a solution for data-efficient hologram data communication between edge and cloud devices, while enabling real-time decoding on the edge end. We envision that this approach has the potential to facilitate the practical deployment of near-eye holographic displays with efficient computational processors in the near future. Thus, we kindly note that experimentally demonstrating an actual compact form-factor holographic display is orthogonal to the scope of this article, and we will leave it for future endeavors (refer to Fig. S3 as an example).

S6 Exploration on Color-channel Joint Compression

To investigate the feasibility of jointly modeling RGB phase channels in a single implicit neural representation (INR), we extend our INR-CGH framework to regress three output channels—corresponding to red, green, and blue phases—using a shared network. This design aims to exploit inter-channel redundancy for more efficient representation and compression.

In this preliminary experiment, we use three CNNpropCNN optimized for individual channels to the output, applied independently to each color channel. Due to GPU memory constraints, we reduce the INR's hidden dimension to 256 channels. For comparison, the single-channel baseline uses an INR with 180 hidden channels per color, trained separately.

As shown in Fig. S4, the jointly trained INR model may result in imperfect color balance compared to the separately optimized channels. This is likely due to the simplistic parameter sharing, which does not fully capture the distinct propagation behavior of each color channel. Despite the drop in fidelity, this approach achieves a high compression gain, highlighting a potential trade-off between quality and compactness. While the current joint modeling setup is limited, it serves as a baseline for exploring more sophisticated multi-channel strategies in future work.

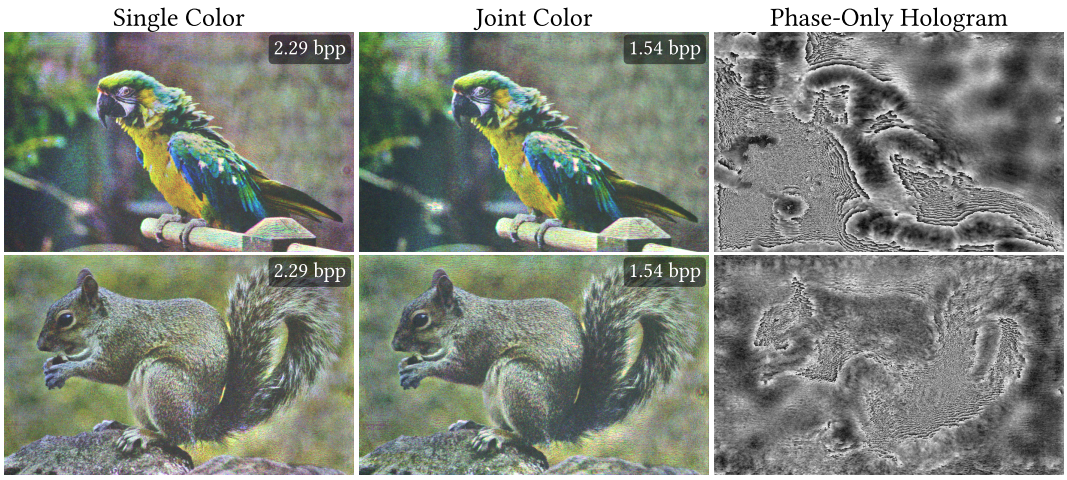

Figure S4: Preliminary visualization of RGB phase-only holography generated using a shared INR model for cross-channel processing. Left: images reconstructed from independently optimized single-channel INR. Middle: images reconstructed from a shared INR jointly optimizing all color channels. Compression ratios in bpp are presented. Right: Corresponding green-channel phase-only holograms from the joint INR model.

S7 Phase-Frequency Characteristics Across CGH Pipelines

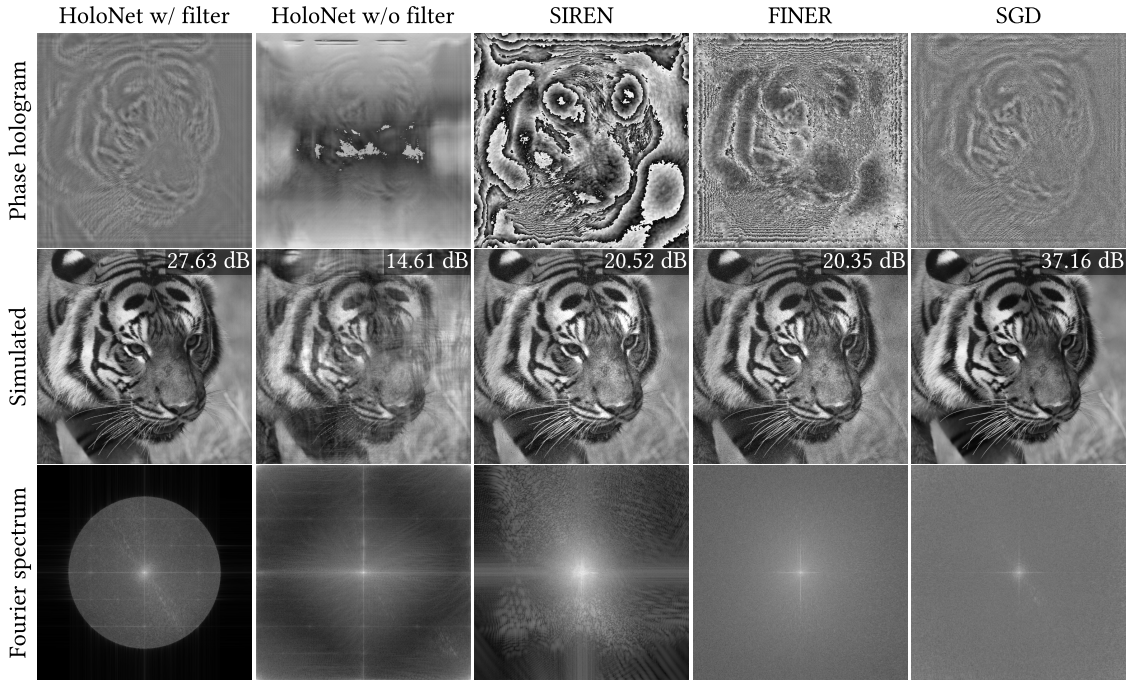

Figure S5: Comparison of simulation results for CNN-based (HoloNet), INR-based (SIREN and FINER), and iterative (SGD) CGH pipelines. From top to bottom: phase holograms, reconstructed amplitudes, and the logarithmic magnitudes of the Fourier spectrum of the reconstructed amplitudes. For HoloNet, results are shown under two conditions: with and without a Fourier aperture filter applied during hologram generation. Overall, both FINER and SGD exhibit broader spectral coverage with more pronounced high-frequency components.

Figure S5 presents a preliminary qualitative analysis of frequency characteristics across representative CGH pipelines. We include HoloNet results with and without Fourier aperture filtering to illustrate the typical gap between conventional filtered simulation configurations and the intended unfiltered scenario. Consistent with the observations presented in the main text, the proposed FINER-based INR yields reconstructions with broader spectral support compared to the CNN-based baseline, and qualitatively aligns closely to the iterative SGD solution in terms of high-frequency content.

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 2

[2] F. W. Campbell. The depth of field of the human eye. *Optica Acta: International Journal of Optics*, 4(4):157–164, 1957. 2

[3] S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein. Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. *ACM Transactions on Graphics (TOG)*, 40(6):1–12, 2021. 1, 3

[4] C. Jang, K. Bang, M. Chae, B. Lee, and D. Lanman. Waveguide holography for 3d augmented reality glasses. *Nature Communications*, 15(1):66, 2024. 3

[5] A. Maimone, A. Georgiou, and J. S. Kollin. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (Tog)*, 36(4):1–16, 2017. 2

[6] S. Marcos, S. A. Burns, E. Moreno-Barriusop, and R. Navarro. A new approach to the study of ocular chromatic aberrations. *Vision Research*, 39(26):4309–4323, 1999. 2

[7] K. Matsushima and 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, Oct 2009. doi: 10.1364/OE.17.019662 1

[8] X. Xia, D. Ma, X. Meng, F. Qu, H. Zheng, Y. Yu, and Y. Peng. Off-axis holographic augmented reality displays with hoe-empowered and camera-calibrated propagation. *Photonics Research*, 13(3):687–697, 2025. 2