Neural Network-Empowered Hologram Compression for Computational Near-Eye Displays

Hyunmin Ban, Wenbin Zhou, Xiangyu Meng, Feifan Qu, Yifan Peng
The University of Hong Kong, Hong Kong SAR, China

Abstract

Holography enhances VR and AR displays by providing realistic 3D imagery, but current computer-generated hologram (CGH) algorithms face high computational demands. This work presents an efficient hologram generation and compression method using a pre-trained wave propagation model and a filter-free design. The approach reduces data redundancy, simplifies hardware, and achieves near real-time decoding, enabling practical use in compact AR/VR systems.

Author Keywords

Computer-generated holography; Compression; Deep Learning; VR/AR; Near-eye displays.

1. Introduction

Computer-generated holography (CGH) has emerged as a technology in virtual and augmented reality (VR/AR), offering true 3D vision that enhances immersive experiences. Its ability to accurately render depth and perspective makes it a promising candidate for next generation display systems [1, 2]. Recent advancements in deep learning have been instrumental in improving holographic display quality and accelerating hologram generation, addressing some of the computational challenges associated with CGH [3, 4].

However, practical deployment of CGH faces several critical challenges. High-resolution holograms require substantial data volumes, which impose significant computational demands. Efficient compression methods are necessary to reduce data transmission and storage requirements without compromising the quality of reconstructed holographic images [5]. Moreover, the mismatch between simulated holographic images and their optical display counterparts often leads to sub-optimal visual quality, limiting the effectiveness of CGH in real-world applications [3]. Efforts have been made to address this issue alongside advancements in compact form factor displays that eliminate the need for additional filtering systems [6].

Traditional approaches, including video codecs like HEVC and JPEG, have been adapted for hologram compression [7]. These methods, illustrated in Figure 1(b), encode holograms using conventional video codecs. However, as these codecs are primarily designed for natural images, they often result in quality degradation when applied to holographic data. Neural network-based techniques, shown in Figure 1(c), have recently emerged as a promising alternative, demonstrating improved compression performance [8, 9]. This includes our proposed method, which optimizes phase-only holograms for low bitrate and high quality while maintaining a lightweight decoding process, making it suitable for edge devices. In contrast, Figure 1(a) is an alternative approach where the target image is compressed using standard codecs, and the hologram is inferred at the edge device. This method places a heavy computational burden on the edge side. While neural network-based techniques address many of these issues, they still face challenges in adapting to real optical display conditions and leave room for further improvements in

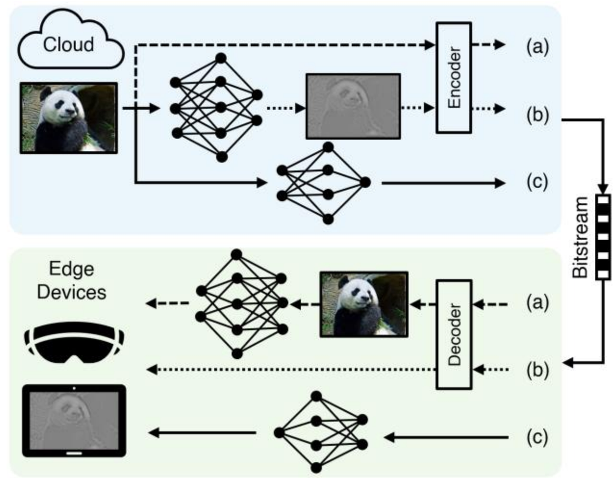

Figure 1 Comparison of hologram generation and transmission schemes: (a) Target image compression using image codecs leads to high computational load during hologram inference. (b) Holograms encoded with video codecs suffer quality degradation. (c) The proposed method optimizes phase-only holograms for low bitrate and high display quality, enabling a lightweight edge solution.

compression efficiency.

This work addresses these gaps by presenting an end-to-end framework that integrates hologram generation and compression using a pre-trained, camera-calibrated wave propagation model. This model ensures high display fidelity while optimizing for RGB input, enabling practical and efficient solutions for next-generation holographic displays.

2. Joint Hologram Generation and Compression

We propose a neural network framework as shown in Figure 2 for training phase-only hologram (POH) generation and compression for holographic near-eye displays. The proposed framework integrates hologram generation and compression to enable efficient phase-only hologram generation and transmission.

Phase Hologram Generation: Our proposed network predicts the phase-only hologram ϕ, enabling the desired target amplitude A_t at the target plane. To reduce inter-color-channel redundancy, we use the RGB channels as the input of our network. The initial U-Net [10] predicts the phase distribution θ_t from A_t, forming a complex field, which is backward-propagated to the spatial light modulator (SLM) plane using angular spectrum method (ASM) propagation, P_{-d}. The three-color channels, processed separately due to their wavelength dependency, are jointly handled by the network. This approach generates six components (real and imaginary parts for each channel), which are then encoded into a latent representation y. The encoder effectively captures critical

ISSN 0097-996X/25/5602-1139-$1.00 © 2025

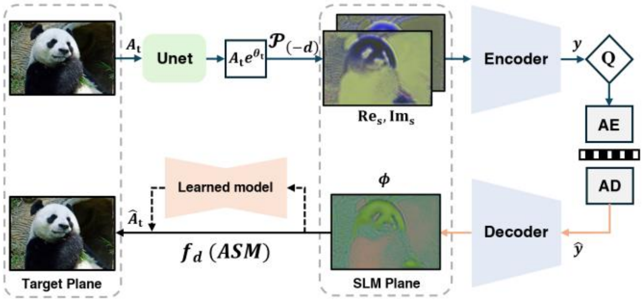

Figure 2 Overview of the proposed network, which inputs the target amplitude and outputs a phase-only hologram. The navy-blue arrow (top) represents the encoding process on the cloud/server side, and the orange arrow (bottom) indicates the decoding process on the edge device. The hyperprior encoder and decoder [11], used in the actual implementation for compression, have been omitted from the figure for simplicity, focusing on the primary flow of the encoder, decoder, and arithmetic encoder/decoder.

information from the complex field, providing a compact structure for efficient compression. To balance performance and complexity, we propose two network scales, the proposed model and a version with reduced feature channels of the network denoted as proposed-small. The latent variable y is decoded into a complex wavefield and constrained to a phase-only representation for the SLM. The POH ϕ is reconstructed using forward ASM wave propagation, optimized by minimizing the distortion metric between the simulated amplitude $\hat{A}_t$ and the target amplitude A_t.

Phase Hologram Compression: The compression pipeline compresses holographic data by quantizing the latent feature y using a quantization operator, enabling efficient data compression similar to learned image compression frameworks [11, 12]. To enhance compression, entropy coding is applied to the quantized feature $\hat{y}$, requiring a probability model for effective encoding. A hyperprior model, introduced by Balle et al. [11], captures spatial dependencies in y through side information z. This improves the entropy model, allowing nearly lossless compression using arithmetic coding. The system is optimized with a rate-distortion loss, expressed as:

$$L = E_{A_t \sim p(A_t)}\left[-\log_2 p_{\hat{y}}(\hat{y}) - \log_2 p_{\hat{z}}(\hat{z}) + \lambda L_D\left(A_t, \hat{A}_t\right)\right], \quad (1)$$

where $p_{\hat{y}}(\hat{y})$ and $p_{\hat{z}}(\hat{z})$ represent the probabilities of the quantized latent $\hat{y}$ and side information $\hat{z}$, balancing compression rate and reconstruction quality through the distortion term L_D. Our model design leverages the hyperprior encoder and decoder architectures from ELIC [12], excluding the space-channel context model to achieve a lightweight and efficient structure.

The propagation, f_d, as shown in Figure 2, can be implemented with the ASM propagation, but we can also employ a modified camera-calibrated, learned wave propagation model to address mismatches between simulated images and physical displays. This camera-calibrated model [4] integrates an ASM propagator with two U-Net architectures. The model accurately predicts the reconstructed POH for new input scenes, applying learned corrections. Its parameters are frozen and integrated with the proposed model, allowing POH generation and compression without additional feedback from the physical display.

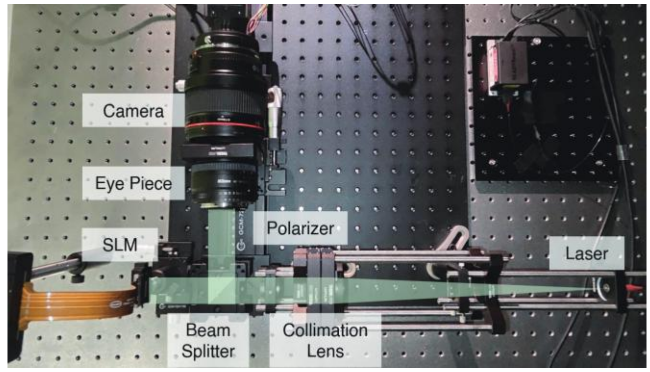

Figure 3 Holographic display prototype used in our experiments. Notably, the setup excludes optical filtering components, which are typically positioned between the SLM and the eyepiece, to explore unfiltered holography scenarios.

3. Implementation

Software: The model was trained using the DIV2K dataset [13], consisting of 800 Full HD images with data augmentation through flipping. Images were preprocessed to 1,600×880 pixels, then zero-padded to Full HD resolution. We trained our network in two stages, pre-training for POH generation over 40 epochs (learning rate 1×10^{-4}) without latent quantization and full network training for compression over 60 epochs, optimizing a joint rate-distortion loss with quantized latent representation. The camera-calibrated wave propagation model was trained for 50 epochs (learning rate 5×10^{-4}). Compression was trained with various λ values, balancing bitrate and quality (λ ranges: {0.0005, 0.002, 0.007, 0.02, 0.08} for proposed-small). The model was compared against H.266/VVC [14] in intra-mode with RGB444 input and QPs {25, 30, 35, 40}, using 100 test images from the DIV2K validation dataset. Experiments used an SLM pixel pitch of 8 μm, wavelengths of 639 nm, 524.9 nm, and 445.8 nm, and a 10 cm propagation distance. All tests ran on an NVIDIA RTX 4090 GPU, with accurate inference times measured using GPU synchronization.

Hardware: Our benchtop holographic display prototype was constructed to validate the proposed method as shown in Figure 2. Generated POHs were displayed on a phase-only SLM (HOLOEYE Pluto 2) with a resolution of 1,080×1,920 pixels and an 8 μm pixel pitch. The propagation distance was set to 10 cm, based on prior studies. Illumination was provided by an RGB laser (Fisba ReadyBeam) directed through optical elements, including a collimating lens, ND filter, polarizer, and beam splitter, before incidence on the SLM. Modulated wavefronts were captured by a FLIR Grasshopper 3 sensor (1,200×1,920, RAW16). The prototype excluded conventional optical filters to evaluate challenging scenarios, utilizing a Nikon 50 mm eyepiece and a Canon 35 mm camera lens. Calibration procedures were performed for the SLM and camera, including voltage response tuning and image alignment via affine transformations. For this proof-of-concept, a green laser (524.9 nm) was used, with full-color displays.

4. Results

Figure 4 highlights the visual quality comparison, where the proposed model significantly outperforms DPRC [8] at similar bpp (bits per pixel), preserving more high-frequency details. It achieves substantial bitrate savings over NHVC [9] while maintaining comparable image quality, especially in sharpness and detail retention. Compared to the HoloNet [3] + VVC scheme, the model

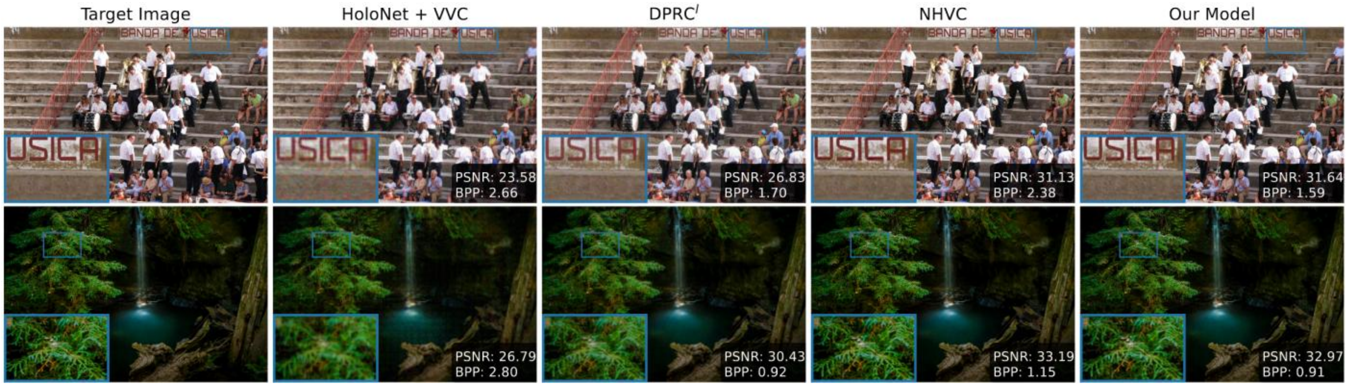

Figure 4 Visual and quantitative comparison of POH compression methods across two scenes. Each row presents different methods with their corresponding PSNR and bpp values. Red boxes highlight zoomed-in regions to emphasize visual details.

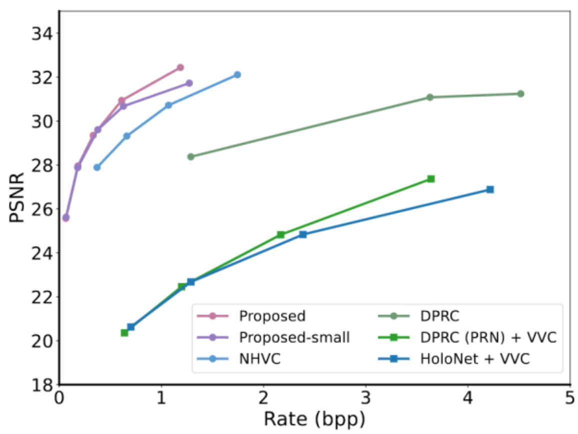

Figure 5 Rate-distortion (RD) curve comparison of POH compression methods. The proposed model is evaluated against various phase hologram compression neural networks and hologram generation combined with state-of-the-art video codecs (VVC), using PSNR for reconstruction quality and bpp for compression efficiency.

provides superior visual quality at lower bitrates with fewer compression artifacts. Zoomed-in areas show that the proposed method effectively minimizes distortion, delivering clearer and more detailed reconstructions.

Rate-distortion (RD) curves illustrate the trade-off between image quality and compression efficiency, with higher PSNR and SSIM values at the same bpp reflecting better performance. In Figure 5, the RD-curve comparison demonstrates that the proposed models consistently outperform existing methods, achieving bitrates from 0.065 bpp (compression ratio ×369) to 1.186 bpp (compression ratio ×20) with PSNR values ranging from 25.6 dB to 32.4 dB. At lower bitrates, both networks, proposed and proposed-small, exhibit similar performance, while at higher bitrates, our model achieves superior quality due to its larger feature capacity.

In Table 1, BD-rate calculations confirm significant bitrate savings, with proposed models achieving a -72.56% and a -74.32% reduction compared to baseline methods like DPRC and NHVC. The models also demonstrate substantial improvements in decoding time, with proposed-small achieving 45 ms per channel (134 ms for three channels), significantly outperforming DPRC and NHVC. This balance of low bitrate, high quality, and fast decoding makes the proposed methods particularly effective for hologram generation and compression.

Table 1. Rate distortion and model decoding time comparison between neural network-based compression models. We compare decoding time for a single channel, with three-channel joint decoding time for our proposed model shown in (·).

Method	BD-rate (%)	Decoding Time (ms)
DPRC	0.00	215
NHVC	-49.38	179
Proposed-small	-72.56	45 (134)
Proposed	-74.32	83 (250)

Figure 6 shows experimental results of holograms at a different compression rate, with ratios ranging from ×67 to ×237 for the first-row images and ×31 to ×118 for the second-row images. Despite high compression, the optical captures maintain high fidelity, demonstrating effective compression and quality preservation. In this experiment, we used single-color channel optimization with half the feature channels of proposed-small, with potential for higher efficiency through joint RGB optimization.

5. Conclusion

In this work, we have proposed a neural network-based framework for efficiently generating and compressing POHs with RGB input. By leveraging a pre-trained, camera-calibrated wave propagation model, our approach addresses the practical challenges of real-world holographic displays, compensating for hardware imperfections and ensuring high display fidelity. Importantly, the proposed method achieves these results without relying on optical filters, enabling a compact and filter-free design that is particularly advantageous for VR/AR applications, where form factors play a critical role. With a decoding time of approximately 40ms per color channel, the proposed framework is well-suited for real-time holography on edge devices, providing a balance of high performance and efficiency, making it a promising solution for next-generation holographic systems that require practical deployment in compact VR/AR environments.

Looking ahead, this work establishes a foundation for future advancements in holography. Improving the learned camera-calibrated model to better address mismatches in unfiltered

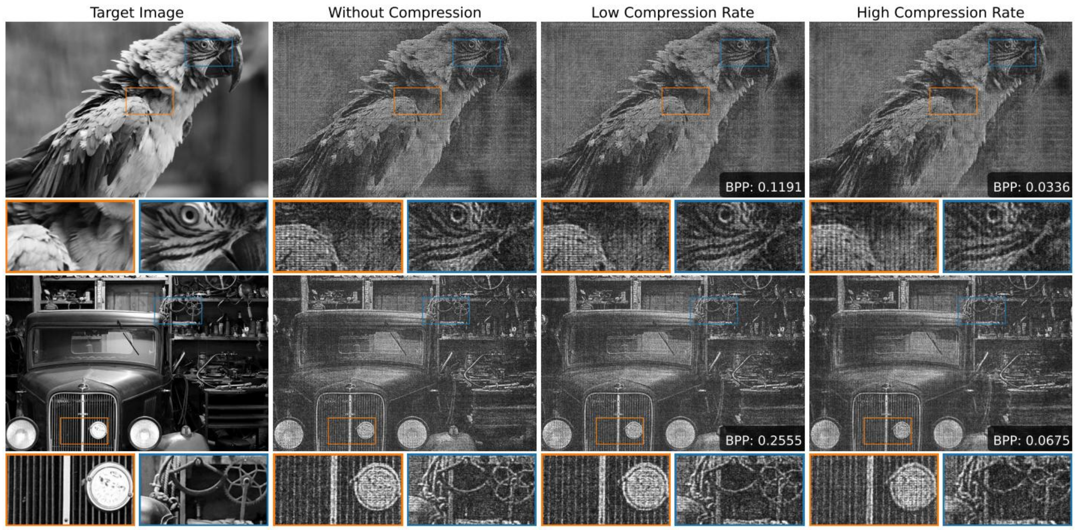

Figure 6 Experimental results of the proposed model tested on a benchtop holographic display with only the green laser source active. The compression rate corresponds to a λ value of 0.003 and 0.0005, with the bit-per-pixel (bpp) value indicated. Green laser images have been converted to grayscale for visualization purposes.

environments could further enhance display fidelity. Integrating our approach with hologram video compression techniques [9], offers the potential for efficient streaming and rendering of holographic videos. Additionally, expanding the model to support 3D holography by training with diverse 3D datasets could significantly broaden its applicability, providing enhanced depth perception and immersive experiences in advanced VR/AR systems. By addressing computational efficiency, display fidelity, and design constraints, this work contributes to bridging the gap between holographic technology and its practical deployment, laying the groundwork for future innovations in holographic displays.

6. Acknowledgements

This work was partially supported by the National Science Foundation of China (62322217) and the Research Grants Council of Hong Kong (ECS 27212822, GRF 17208023).

7. References

1. 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. 2020 Nov 20;7(11):1563-78.

2. Maimone A, Georgiou A, Kollin JS. Holographic near-eye displays for virtual and augmented reality. ACM Transactions on Graphics (Tog), 36(4):1-6, 2017.

3. Peng Y, Choi S, Padmanaban N, Wetzstein G. Neural holography with camera-in-the-loop training. ACM Transactions on Graphics (TOG). 2020 Nov 26;39(6):1-4.

4. Choi S, Gopakumar M, Peng Y, Kim J, Wetzstein G. Neural 3D holography: learning accurate wave propagation models for 3D holographic virtual and augmented reality displays. ACM Transactions on Graphics (TOG), 40(6):1-2, 2021

5. Blinder D, Ahar A, Bettens S, Birnbaum T, Symeonidou A, Ottevaere H, Schretter C, Schelkens P. Signal processing challenges for digital holographic video display systems. Signal Processing: Image Communication. 2019 Feb 1;70:114-30.

6. Gopakumar M, Kim J, Choi S, Peng Y, Wetzstein G. Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays. Optics letters. 2021 Dec 1;46(23):5822-5.

7. Oh KJ, Ban H, Choi S, Ko H, Kim HY. HEVC extension for phase hologram compression. Optics Express. 2023 Mar 13;31(6):9146-64.

8. Wang Y, Chakravarthula P, Sun Q, Chen B. Joint neural phase retrieval and compression for energy-and computation-efficient holography on the edge. ACM Transactions on Graphics. 2022 Jul;41(4).

9. Ban H, Choi S, Cha JY, Kim Y, Kim HY. NHVC: Neural Holographic Video Compression with Scalable Architecture. In 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) 2024 Mar 16 (pp. 969-978). IEEE.

10. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InMedical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 2015 (pp. 234-241). Springer International Publishing.

11. Ballé J, Minnen D, Singh S, Hwang SJ, Johnston N. Variational image compression with a scale hyperprior. arXiv preprint arXiv:1802.01436. 2018 Feb 1.

12. He D, Yang Z, Peng W, Ma R, Qin H, Wang Y. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 (pp. 5718-5727).

13. Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study. InProceedings of the IEEE conference on computer vision and pattern recognition workshops 2017 (pp. 126-135).

14. Bross B, Wang YK, Ye Y, Liu S, Chen J, Sullivan GJ, Ohm JR. Overview of the versatile video coding (VVC) standard and its applications. IEEE Transactions on Circuits and Systems for Video Technology. 2021 Aug 2;31(10):3736-64.