Computational Visual Media
https://doi.org/10.26599/CVM.2025.9450497

Research Article

Multi-color compressive hologram synthesis with learned wave propagation

Hyunmin Ban[1], Wenbin Zhou[1], Xiangyu Meng[1], and Yifan (Evan) Peng[1,2] (✉)

Abstract Holographic displays are a promising technology for delivering immersive, true 3D visualization in virtual and augmented reality applications. However, generating high-fidelity phase-only holograms remains challenging, especially with the demand for efficient compression to handle the substantial data inherent in high-resolution holographic streaming. Existing techniques often struggle to balance the trade-off between optical display quality and compression efficiency, and jointly optimizing these aspects is still in its infancy. This work presents a learning-empowered multi-color hologram compression scheme that utilizes a pre-trained, camera-calibrated wave propagation model, especially for unfiltered holographic display configurations with compact form factors. In particular, the inter-color processing leverages the inherent redundancy across color channels, allowing for efficient compression. By incorporating the learned camera-calibrated wave propagation model into our training process, we can achieve superior optical display quality and compression rates. Experiments demonstrate that our method realizes a reduction in bits per pixel (bpp) of 44% to 74% over representative baselines at the same quality level. We envision the proposed compressive hologram synthesis scheme establishing a new benchmark for high-fidelity holographic reconstruction at lower bitrates, marking a significant advance towards the deployment of holography-empowered visual media systems.

1 Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China. E-mail: H. Ban, hmban@connect.hku.hk; W. Zhou, zhouwb@connect.hku.hk; X. Meng, mengxy22@connect.hku.hk; Y. Peng, evanpeng@hku.hk (✉).

2 Department of Computer Science, The University of Hong Kong, Hong Kong SAR, China.

Manuscript received: 2024-10-19; accepted: 2025-06-20

Keywords computer-generated hologram (CGH); image compression; learned propagation; virtual reality (VR); augmented reality (AR)

1 Introduction

Computer-generated holography (CGH) has attracted significant attention in recent years due to its potential applications in virtual reality (VR) [1, 2] and augmented reality (AR) [3, 4]. Holography can provide true 3D vision, positioning it as a promising technology for enhancing immersive experiences. Recently, deep learning has been widely adopted to improve its display quality and reduce hologram generation time [5, 6]. However, the practical deployment of holography faces several critical challenges. One is the substantial volume of data required, particularly for high-resolution digital holograms [7]. This leads to a considerable computational load, underscoring the importance of efficient compression to alleviate data transmission and storage requirements. Notably, compression must be performed without compromising the quality of the reconstructed holographic images, a task that becomes more daunting when transitioning from simulated environments to hardware displays. The mismatch between simulated holographic images and their optical display counterparts often results in sub-optimal visual experience [5], limiting the effectiveness of holographic applications in real-world scenarios. Efforts have been made to address this mismatch between simulations and optical displays [6, 8]. Additionally, we note that in the CGH community, efforts have been directed towards exploring compact form factor holographic displays that do not require incorporating an extra filtering

system. These approach is known as unfiltered holography [9].

Conventional video codecs such as HEVC [10] and JPEG [11] have been repurposed for hologram compression in recent studies. For instance, Oh et al. [12] applied an HEVC extension for phase holograms, taking into account the periodic nature of phase values. Blinder et al. [13] explored holographic video compression using HEVC with global motion compensation to reduce temporal redundancy. Zhou et al. [14] proposed a JPEG-aware end-to-end optimized hologram compression method, while Jiao et al. [15] proposed a hologram compression framework that combines JPEG encoding and neural network-based post-processing to enhance reconstruction quality. Shi et al. [16] integrated HEVC with a generative adversarial network (GAN) for complex hologram compression. While these approaches lay a foundation, they do not fully achieve the desired levels of compression efficiency and visual quality. In contrast, neural network-based techniques have recently emerged as a promising avenue for hologram generation and compression [17, 18], offering efficient GPU inference and the potential for superior reconstruction quality. Wang et al. [17] introduced a neural framework for phase-only hologram compression, which was later extended to video compression by Ban et al. [18], while Jia et al. [19] proposed a variant considering foveation. These methods demonstrated the inefficiency of schemes (a) and (b) in Fig. 1 while proposing scheme (c) as a more effective solution. However, many of the reported models struggle with optimization in less-than-ideal display conditions and lack support for multi-color holograms with three-channel RGB input. The necessity for both effective compression and display-aware optimization remains a crucial hurdle in advancing the practicality of CGH. Beyond compression, the accurate generation of holograms is of paramount importance. To this end, compression techniques must be aware of the distinct needs of hologram generation, ensuring that the resulting holograms provide high display quality. Addressing both hologram generation and compression concurrently is essential for unlocking the complete capabilities of holography. In this work, we utilize a pre-trained, camera-calibrated wave propagation model to enhance display fidelity and

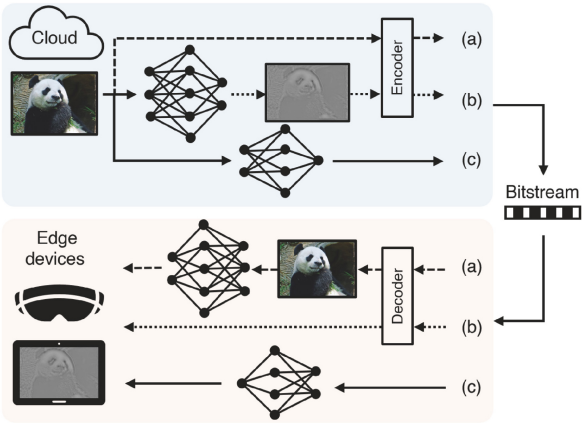


Fig. 1 Comparison of hologram generation and transmission schemes in visual computing applications. (a) The target image is compressed using a video or image codec (e.g., JPEG) and sent to the edge device, where a neural network infers the hologram, causing high computational load and latency. (b) The hologram is generated and encoded using conventional video codecs, but quality is likely to be degraded as these codecs are optimized for natural images. (c) Our proposed method generates and optimizes the phase-only hologram for both low bitrate and high display quality, offering a lightweight solution for edge devices.

streamline the compression of computer-generated holograms, with a particular focus on optimizing for RGB input. This work makes the following main contributions:

- A neural network-empowered compressive hologram synthesis scheme, HoLIC (*holo-learned image compression*), which efficiently generates compressed holograms with three-channel RGB input, leveraging inter-color redundancy to improve compression performance and reduce decoding complexity.
- A camera-calibrated wave propagation model is incorporated in our learning scheme to strike a balance between compression efficiency and display fidelity, suitable for practical applications. This model adeptly represents and compensates for imperfections in displays while operating at a low bit rate.
- A demonstration of the effectiveness of our network model in synthesizing phase-only holograms for three channels in simulation, as well as unfiltered holographic display configurations, ensuring improved fidelity and a compact display form factor.

2 Related work

2.1 Computer-generated holography

Recent advances in holographic near-eye displays

have focused on enhancing support for focus cues [20, 21]. Our work primarily focuses on advancing wave propagation models to unlock the full potential of practical holographic displays. Many existing CGH methods focus on generating phase-only holograms (POHs) by encoding complex amplitude into phase through algorithmic approaches, such as double phase-amplitude coding (DPAC) and its variants [22, 23]. These have been widely utilized due to their efficiency. In addition, the application of artificial intelligence, particularly deep neural networks, in holography has seen considerable development to infer POHs in real time [24, 25]. On the other hand, iterative optimization methods, such as the Gerchberg-Saxton (GS) algorithm [26, 27] and stochastic gradient descent (SGD)-based approaches [5, 28], involve iteratively updating POHs to make the simulated results more closely resemble the original targets. Although these methods typically require more time than direct encoding techniques, they can produce more accurate reconstructions. Our work builds on these methods by jointly optimizing the generation and compression of POHs.

2.2 Camera-calibrated holographic displays

Recent advances in holographic displays have highlighted the limitations of directly applying ideal wave propagation models such as the angular spectrum method (ASM) [29] to physical systems. These models often do not account for the non-idealities of real-world systems, including phase distortion of the SLM, optical aberration, and limited diffraction efficiency [3, 4]. This mismatch between simulated and physical image formation results in degraded image quality. To address these challenges, online calibration involves fine-tuning the POH based on live feedback from a camera, which typically achieves the best quality but requires re-calibration for every image [5, 8]. In contrast, offline calibration uses a pre-captured dataset to fit a learned optical system model, which generalizes to a variety of target images [6, 21]. Our also work incorporates an offline-calibrated and learned model into hologram compression and generation, aiming to deliver compelling visual display quality.

2.3 Learned image compression

Learned lossy image compression models have gained significant attention in recent years due to their ability to outperform traditional video codecs in terms of compression efficiency. Foundational work by Balle et al. [30] introduced an auto-encoder architecture where the encoder maps an input image to a compact latent representation, followed by quantization and entropy coding [31]. To enhance compression efficiency, a hyper-prior model is employed to capture spatial dependencies, allowing for better rate estimation. Minnen et al. [32] further improved this by introducing an auto-regressive context model, enabling more accurate probability predictions for entropy coding. Although effective, this model relies on sequential decoding, making it less practical for high-resolution images due to its slow decoding speed. Recent advances in entropy modeling, such as channel-wise entropy models [33, 34], offer parallel decoding techniques with improved decoding speed compared to the sequential method. Extending this state-of-the-art, our compressive hologram synthesis scheme focuses on a more general model, with a fast and lightweight architecture, offering broader applicability without relying on complex context models.

3 Method

3.1 Learned hologram synthesis

Our overall hologram generation network is illustrated in Fig. 2. To efficiently reduce inter-color-channel redundancy, our proposed hologram generation and compression network utilizes RGB (three channels) to represent the holographic data. In the initial stage, we use a basic U-Net to estimate the unknown phase distribution at the target plane, forming a complex-valued wave-field that is propagated to the spatial light modulator (SLM) plane using backward ASM wave propagation $\mathcal{P}_{(-d)}$. The three color channels, treated separately due to wavelength dependency, are processed jointly by our proposed network. This results in six components (real and imaginary components for each channel) being fed into an encoder, $g_a(\mathrm{Re}_s, \mathrm{Im}_s)$, which maps the input to a latent space representation, y, capturing essential information from the complex field to provide a suitable structure for efficient compression. For brevity, we provide the derivation of ASM in Appendix A.

As Fig. 3 shows, our encoder g_a and decoder g_s are

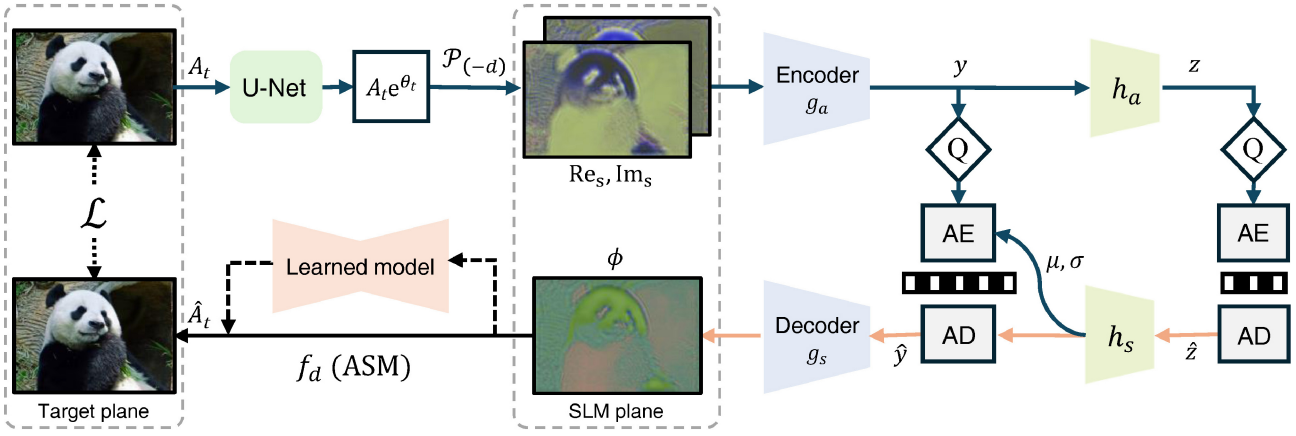


Fig. 2 Architecture of our proposed HoLIC network, which inputs the target amplitude A_t and outputs a phase-only hologram ϕ. The blue arrows (above) represent the encoding process, which is performed in the cloud or on a server. The orange arrows (below) indicate the decoding process, executed on the edge device. This division between encoding and decoding ensures efficient hologram data transmission and processing, with computationally intensive encoding handled by powerful resources, while the lighter decoding is optimized for near real-time performance on edge devices. Here, Q denotes the quantization process while AE and AD denote arithmetic encoding and decoding, respectively. f_d denotes forward wave propagation, which can use a plain ASM operator or a learned, camera-calibrated model, with the holographic image reconstructed at a distance D.

built with a similar structure to the slim version of ELIC [35]. We observe that since a small hologram region affects a large area of the target image, a large receptive field is crucial for effective hologram generation. To achieve this, we use large convolution kernels and residual blocks, which help to increase the receptive field. Additionally, we find that increasing the number of channels in the network improves POH generation performance and the compression ratio, particularly in high-bitrate regions. However, in lower-bitrate regions, the performance improvement is marginal, and using fewer channels reduces the network's complexity, resulting in a trade-off.

To balance performance and complexity, we explore two scales for the HoLIC network. The original version uses $(N, M) = (192, 320)$, while a smaller less complex version, HoLIC-small uses $(N, M) = (144, 144)$. Without applying further compression to this latent feature, for example, in Fig. 3, the latent variable y can be directly input into the decoder $g_s(y)$, which outputs a complex-valued wave field. For compatibility with the SLM, we constrain the output to a phase-only value ϕ by computing the angular component, which serves as the network's final output as a POH. The POH output, ϕ, can be reconstructed by applying the forward ASM wave propagation from the SLM plane to the target plane. In this POH generation scheme, we optimize our HoLIC model to generate a POH ϕ by minimizing the distortion metric $\mathcal{L}_D$ between the simulated amplitude $\hat{A}_t$ derived from ϕ and the target amplitude A_t. During the optimization including the learned camera-calibrated model, we introduce total variation (TV) regularization that locally smooths out phase values to improve display quality. In all, our hologram generation optimization process is defined as

$$\phi = \underset{\phi}{\operatorname{argmin}} \left(\mathcal{L}_D(|\hat{A}_t|, |A_t|) + \alpha \, \|\nabla \phi\|_2 \right) \quad (1)$$

where α weights the impact of the TV regularizer.

3.2 Jointly optimizing compression and generation

To jointly optimize hologram generation and compression, we incorporate well-established quantization and entropy coding. Specifically, the latent feature y is quantized using a quantization

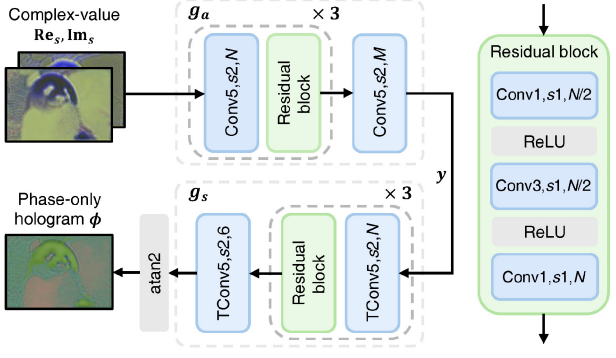


Fig. 3 Encoder and decoder modules. TConv denotes transposed convolution, while s represents the stride value. N and M refer to the numbers of output channels with the final layer of g_s producing 6 channels corresponding to the real and imaginary components of the complex values, respectively.

operator, allowing for compression of the holographic data, following approaches similar to those used in image compression learning frameworks. To further enhance compression, entropy coding is applied, enabling efficient transmission of the quantized latent feature $\hat{y}$. To perform entropy coding effectively, the probability distribution of the quantized feature $\hat{y}$ must be modeled. Prior work by Balle et al. [30] introduced a hyper-prior model that uses side information z to capture spatial dependencies between the elements of latent feature y. This is achieved through an auxiliary auto-encoder, with analysis and synthesis transforms h_a and h_s, respectively. The hyper-prior helps improve the entropy model, allowing for the use of methods like arithmetic coding to compress the quantized data in a nearly lossless manner.

To ensure that the entire system is differentiable end-to-end, we employ the straight-through estimator (STE) [36, 37], which provides an approximation for gradient flow through the non-differentiable operations like quantization. This enables bitrate estimation within the network. Overall, the system is trained using rate-distortion loss, defined as

$$\mathcal{L} = \mathbb{E}_{A_t \sim p(A_t)}\big[-\log_2 p_{\hat{y}}(\hat{y}) - \log_2 p_{\hat{z}}(\hat{z}) + \lambda \mathcal{L}_{\mathrm{D}}(A_t, \hat{A}_t)\big] \tag{2}$$

This loss balances the trade-off between the compression rate (through $p_{\hat{y}}$ and $p_{\hat{z}}$) and reconstruction quality (through the distortion term $\mathcal{L}_{\mathrm{D}}$. We encode each $y - \mu$ instead of directly quantizing y, and restore the quantized symbol as $\hat{y} = \mathrm{STE}(y-\mu)+\mu$, which benefits the entropy model by providing smoother gradients during optimization. In terms of network architecture, we adopt the design of h_a and h_s from ELIC [35] but exclude the space-channel context model to give a fast and lightweight structure.

3.3 Calibrated wave propagation model

The propagation function f_d in Fig. 2 can be implemented with basic ASM propagation, as detailed in Appendix A, for simulation. However, this will lead to a mismatch between simulated images and those observed on a physical display. To address this, we adopt a modified version of the camera-calibrated, learned wave propagation model introduced by Choi et al. [6], which uses a hybrid model combining wave propagation with neural network corrections. This model integrates two U-Net architectures around an ASM propagator. The ASM propagator serves as the ideal wave propagation model, while the two U-Nets are designed to predict the residual discrepancies between the analytical model and the real optical display. Specifically, the first U-Net accounts for non-ideal SLM behavior like crosstalk. After the ASM propagation step, the second U-Net resolves additional optical aberrations in the system.

To train this model, 2000 POHs generated by SGD and DPAC are sequentially displayed on the SLM and captured at the target plane. The loss function minimizes the discrepancy between the predicted and captured intensities using the l_1 norm, ensuring the model effectively learns the mismatch between the simulated and physical system, particularly in the unfiltered setup. Once trained, the camera-calibrated model can accurately predict the reconstructed POH on the physical display for new input scenes by applying these learned corrections. Since this model already accounts for real-world discrepancies during training, its parameters are frozen and used for joint training with our HoLIC model. This eliminates the need for the HoLIC model to rely on additional feedback from the physical display while generating the POH.

4 Implementation and results

4.1 Configuration details

Our model was trained using the DIV2K dataset [38], which consists of 800 full HD images, with vertical and horizontal flipping for data augmentation. The images were pre-processed to yield a region of interest of 1600×880 pixels, followed by zero-padding to restore the full HD resolution, as described in the work by Peng et al. [5].

The training process was divided into two stages to speed convergence. In the first stage, the model was pre-trained for POH generation over 40 epochs using a learning rate of 10^{-4}, without applying quantization to the latent information y, and the holograms are generated only considering the distortion loss. In the second stage, the entire network was trained for compression, where the latent information y was quantized to $\hat{y}$, and the network was optimized using the hyper-prior model h_a and h_s, with a joint rate-distortion loss. This stage was trained for 60 epochs,

with the first 40 epochs trained using a learning rate of 10^{-4}, followed by 20 epochs trained at a learning rate 10^{-5}. To optimize the distortion loss, training minimized the mean squared error. We trained the camera-calibrated wave propagation model for 50 epochs using a learning rate of 5×10^{-4}.

To control the trade-off between bitrate and quality, we trained the compression model using different λ values, where higher λ results in a higher bitrate and quality. The chosen λ values for training HoLIC-small and HoLIC were {0.0005, 0.002, 0.007, 0.02, 0.08} and {0.0005, 0.002, 0.0045, 0.0125, 0.045}, respectively. Lossless entropy encoding and decoding were implemented using the range-based asymmetric numeral systems (rANS) [39]. Our network's compression performance was evaluated against the state-of-the-art compression standard H.266/VVC [40]. We used the standard VVC test model VTM 23.4 [41], using intra-mode with RGB444 as input color format. The quantization parameters (QPs) were set to {25, 30, 35, 40} for comparison to other neural network-based POH compression methods using similar bit ranges. Evaluation of reconstruction performance was conducted using 100 test images from the DIV2K dataset [38]. In all experiments, the pixel pitch of the SLM phase was set to 8 μm. Investigated wavelengths for red, green, and blue channels were 639, 524.9, and 445.8 nm, respectively, with a propagation distance of 10 cm. All algorithms and neural network models were trained and tested using a single NVIDIA RTX 4090 GPU, while the conventional codec was tested with a CPU. When testing the generation speed and decoding speed, we used GPU synchronization to measure accurate inference times.

4.2 Hologram generation performance in simulation

Various methods have been proposed solely to generate POH; HoloNet and the stochastic gradient descent (SGD) method are well-known. While HoloNet directly generates holograms, the basic SGD method is an iterative approach. We compare POH generation simulation results in Table 1. In our experiments, the SGD method was tested using 1000 iterations to ensure a thorough evaluation of its performance. While HoloNet is recognized for fast POH generation, our HoLIC model demonstrates even faster generation times per color channel, while achieving significantly higher PSNR performance. Although the SGD method produces the highest quality results, its iterative nature results in a slower speed, limiting its practicality for real-time applications.

Table 1 Qualitative comparison of diverse POH generation methods. The top three results are highlighted. Note that for all baseline methods, run-time is reported for single-channel optimization, while (·) also reports it for three-channel joint optimization

Method	PSNR (dB) ↑	SSIM ↑	Time (ms) ↓	
SGD (iterative)	**35.88**	**0.938**	3771	
HoloNet	29.70	0.846	**34**	
DPRC (PRN)	30.49	0.884	78	
NHVC-G	**34.23**	**0.917**	115	
HoLIC-small	31.82	0.879	**20**	**(59)**
HoLIC	**33.58**	**0.908**	**22**	**(67)**

In addition to hologram generation, recent methods such as DPRC [17] and NHVC [18] have investigated optimization schemes that jointly generate holograms and compress them. These works have reported the performance of their POH generation processes independently allowing for a direct comparison to our method. Our HoLIC-small and HoLIC models exhibit results 1.33 and 3.09 dB better in terms of reconstruction quality, respectively, when compared to the PRN (phase retrieval network) of the DPRC model, and are faster. We note that while our model's hologram generation performance is comparable to NHVC's generation model, it has slightly lower PSNR and SSIM metrics, as NHVC is optimized for single-channel operation. Despite this, our model maintains competitive performance overall. Notably, our POH generation speed is much faster than that of NHVC, bringing our HoLIC model much closer to real-time generation, as needed in practical applications. Additionally, the difference in performance between HoLIC and HoLIC-small highlights the importance of the number of feature channels in POH generation. The increased number of feature channels in the HoLIC model directly contributes to improved reconstruction quality.

4.3 Rate-distortion performance in simulation

Rate-distortion (RD) curves are a standard method to illustrate the trade-off between image quality and bits per pixel (bpp), where fewer bpp typically results in reduced quality. Since one pixel corresponds to 24 bits (8 bits per channel for RGB), achieving fewer bpp is advantageous as it represents more effective

compression. At any given quality level, higher PSNR and SSIM metrics for the same number of bpp indicate better image quality. As such, an RD curve positioned closer to the top-left of the figure reflects superior performance.

The RD-curve comparison in Fig. 4 shows that our model consistently outperforms other methods in terms of compression efficiency. Specifically, we demonstrate bitrates between 0.065 bpp (compression ratio of 1:369) and 1.186 bpp (compression ratio of 1:20), with corresponding PSNR values ranging from 25.57 to 32.44 dB. Our proposed network achieves lower bitrates while maintaining higher image quality, especially when compared to existing neural network-based compression models and standard video codecs like VVC. Our approach demonstrates superior RD performance across a range of bitrates. Additionally, the RD curves show, our two networks exhibit similar performance in the low-bitrate region, with both HoLIC and HoLIC-small showing comparable compression ratios. However, at higher bitrates, the HoLIC network outperforms HoLIC-small. This can be attributed to the fact that, at higher bitrates, the upper bound of performance is largely determined by the POH generation quality, where the increased feature channels of the HoLIC model allow it to converge to a higher reconstruction quality.

To further assess compression effectiveness, we calculate the BD-rate [42], which quantifies the average rate deviation between two compression methods for the same quality. The BD-rate results in Table 2 indicate that our proposed models achieve substantial bitrate savings compared to the baseline DPRC and NHVC models. Specifically, HoLIC-small achieves a −72.56% reduction in BD-rate, while HoLIC achieves a reduction of −74.32%. Here DPRC serves as the baseline with a BD-rate of 0%. In comparison to NHVC, HoLIC achieves a BD-rate value of 44.17%, setting a new benchmark for compression efficiency in hologram generation.

Table 2 Rate distortion and model decoding complexity of varying hologram compression models: the BD-rate is calculated relative to DPRC from PSNR and bpp curves. Decoding time is reported for a single channel, with three-channel joint optimization times for our HoLIC model denoted (·). The number of decoder parameters for our model is based on three channels but a single channel for the other methods. P = number of decoding parameters, D = decoding time

Method	P	BD-rate (%) ↓	D (ms) ↓	
DPRC	0.54 M	0.00	215	
NHVC	5.04 M	−49.38	179	
HoLIC-small	3.21 M	**−72.56**	45	(134)
HoLIC	10.30 M	**−74.32**	83	(250)

In terms of decoding time, our proposed models show a substantial improvement over the competing methods. HoLIC-small has a decoding time of just 45 ms per channel (134 ms for all three channels), which is significantly faster than DPRC (215 ms) and NHVC (179 ms). Even HoLIC, with 83 ms per channel (250 ms for three channels), maintains a competitive decoding time while achieving better compression efficiency. We note that in our experiments, incorporating additional residual blocks and attention modules increases decoder-side complexity by approximately 3.6 million parameters and adds approximately 30 ms to the decoding time,

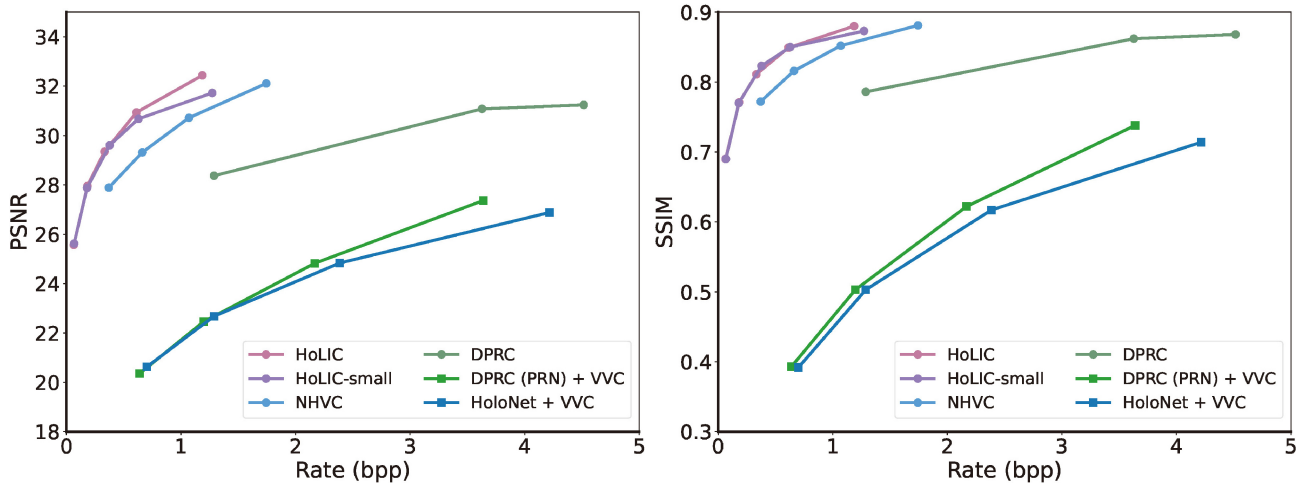


Fig. 4 RD-curve comparison of POH image compression methods. Our HoLIC model was compared to various neural network-based hologram compression methods, and methods combining neural network-based hologram generation with the state-of-the-art VVC video codec, evaluated using PSNR and SSIM for reconstruction quality, as well as bpp for compression efficiency.

with a corresponding BD-rate improvement of around 5%. While the total parameter counts for our models are 9.4 million for HoLIC-small and 19.7 million for HoLIC, only a portion of these parameters, 3.21 million and 10.3 million respectively, are required on the decoder side for deployment on edge devices: HoLIC-small is lightweight, and suitable for edge device deployment. In contrast, HoLIC provides higher quality at higher bitrates but with increased complexity. For comparison, NHVC requires 5.04 million parameters while DPRC requires only 0.54 million parameters. However, it is important to note that both DPRC and NHVC are optimized for single-color channels, whereas HoLIC and HoLIC-small are designed for three-color channels, which inherently increases their complexity but also leads to superior performance in both BD-rate and decoding time, making them more suitable for real-time applications.

In terms of visual quality, Fig. 5 provides a comparison where our HoLIC model demonstrates significantly better performance than DPRC at similar bpp, preserving more high-frequency details. Additionally, our HoLIC model achieves considerable bitrate savings compared to NHVC while maintaining similar image quality, particularly in terms of sharpness and detail retention. Compared to the HoloNet + VVC scheme, our model delivers much better visual quality at lower bitrates, with fewer compression artifacts. The visual differences are especially noticeable in zoomed-in areas, where HoLIC effectively avoids blurring and distortion, leading to clearer and more detailed reconstruction quality.

Fig. 5 Visual and quantitative comparison of POH compression methods across three different scenes. Each row shows different methods with PSNR (dB) and bpp indicated for each. The boxes show close-ups, emphasizing visual details.

4.4 Validation on unfiltered holographic display

A bench-top holographic display prototype was constructed to validate the proposed methodology, as depicted in Fig. 6. Appendix B gives a schematic outline of typical holographic near-eye displays. The synthesized holograms were displayed on a phase-only SLM (HOLOEYE Pluto 2), with a resolution of 1080×1920 pixels and a pixel pitch of 8 μm. In our experiment, the propagation distance was set to 10 cm based on empirical studies from Padmanaban et al. [43] and Peng et al. [5]. Illumination was provided by an RGB laser (Fisba ReadyBeam), whose beam was directed through a sequence of optical elements, including a collimating lens, a neutral density filter, a linear polarizer, and a beam splitter. The shaped beam was incident onto the SLM. The modulated wavefronts were reflected, propagated, and transmitted through an eyepiece and a lens before being recorded by a FLIR Grasshopper 3 sensor with a resolution of 1200×1920 pixels in RAW16 format.

Notably, our bench-top display prototype was configured without a conventional optical filter to explore challenging scenarios. The eyepiece was a Nikon 50 mm SLR lens, while the camera lens was a Canon 35 mm SLR lens. The wave propagation model was trained without the inclusion of optical filters. Calibration procedures were carried out for both the SLM and the camera before image acquisition. The SLM's voltage response was calibrated using vortex patterns, with manual adjustment. A homography-based affine transformation [6] was applied to reduce jitter in the captured images, ensuring dataset stability through image alignment. At this proof-of-concept stage, only the green laser (524.9 nm) was functioning. The full-color display could be represented by temporal multiplexing at 120 Hz of RGB channels, a goal that could be reasonably attained with some engineering effort in the near future.

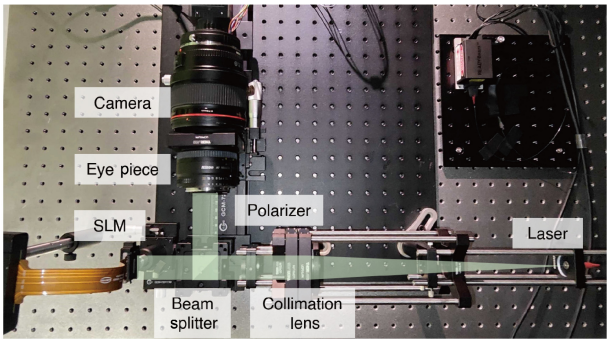


Fig. 6 Photograph of our holographic display prototype. Note that our experiment does not involve any optical filtering parts that are often placed in-between the SLM and the eye-piece.

Figure 7 presents holograms generated with varying compression rates. The top row demonstrates compression ratios ranging from 1:67 to 1:237, while in the second row compression ranges from 1:31 to 1:118. The number of bpp for each image is indicated. Despite the high compression ratios, the optical captures maintain high fidelity, showing the effectiveness of the compression while preserving the quality of the holographic reconstructions. In this experiment, we optimized for single-color channels using half of the feature channels from HoLIC-small. Jointly optimizing the RGB channels would result in higher compression ratios, further improving efficiency without compromising visual fidelity.

4.5 Discussion

In Fig. 8, we compare latent entropy maps of RGB joint training and single-channel optimized models, both evaluated using the same HoLIC-small architecture. The entropy distribution in the latent maps reflects the complexity of the target image, where regions with more complicated textures require more bits for encoding. While single-channel optimization with more feature channels usually improves POH generation quality, its impact on compression efficiency is limited. We observe that despite single-channel optimization using three times as many parameters as the RGB joint approach, it results in higher entropy (brighter values across the map), indicating less efficient compression. RGB joint compressive POH synthesis, on the other hand, more effectively exploits spatial redundancy across the complex-valued color channels, leading to superior compression performance with fewer parameters, without compromising reconstruction quality. Thus, by jointly optimizing the RGB channels, our model can achieve better bit-rate savings, validating the advantages of leveraging inter-color redundancy.

One of the primary objectives of this work was to facilitate real-time compression and synthesis of POHs for application to VR/AR displays. Our framework can achieve a decoding speed of 40 ms per color channel, which is usable for real-time holography, particularly considering that the encoding would be

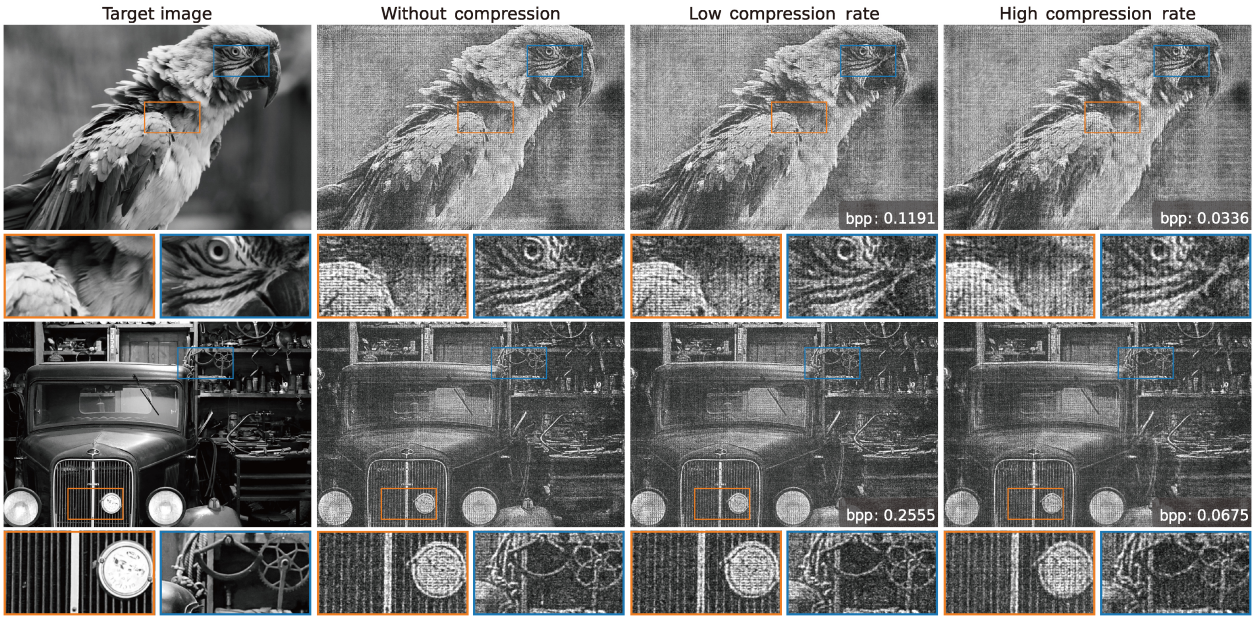


Fig. 7 Experimental results using our HoLIC-small model tested on a bench-top holographic display with only the green laser source operating. The "low" and "high" compression rates correspond to λ values of 0.003 and 0.0005, respectively. The number of bpp for each image are indicated for both low and high compression rates. Note that we have converted the green laser images to grayscale for presentation.

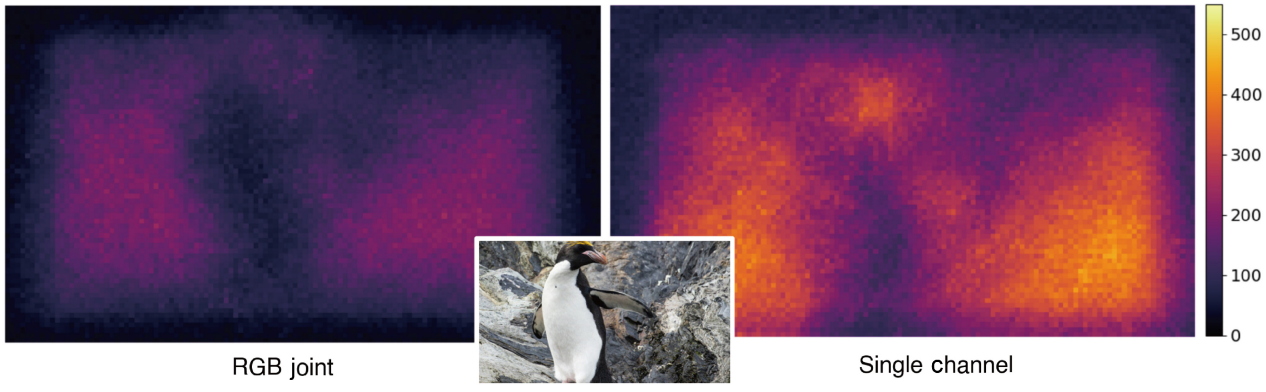


Fig. 8 Latent entropy maps for RGB joint training and single-channel, optimized at similar PSNR quality, showing the spatial distribution of entropy (in bits) after compressing the holographic data, calculated using $-\log_2 p_{\hat{y}}(\hat{y})$ from the likelihoods of the latent variables. They highlight regions with higher information density, requiring more bits for encoding, indicating regions of complexity in the data. The testing image to be encoded is shown in the center, for reference.

done server-side, while decoding would be performed on edge devices. Note that the decoding speed of our model is primarily influenced by the entropy decoding process, which is inherently sequential. While this poses a limitation on decoding speed, further optimization in this area, such as hardware acceleration, could lead to faster performance in future. Additionally, the POH generation process requires approximately 20 ms per color channel, leading to a total generation time that closely matches real-time requirements for visual media.

While these results are highly favorable, there is still potential for optimization. Incorporating techniques such as mask decay [44] or implicit neural representations [45] into compression could further reduce processing time. In our optical experiments, we have demonstrated a significant step toward practical implementation of unfiltered holography setups. Nonetheless, it remains challenging for the learned camera-calibrated model to fully account for mismatches in unfiltered environments. Further investigations in this area could lead to higher-fidelity displays for holographic systems. Although currently limited by readily-available hardware configurations, full-color optical reconstruction is planned as an important direction for future work. Integrating

our method with the hologram video compression approach proposed by Ban et al. [18] could further enhance compression performance, enabling efficient streaming and rendering of holographic videos. In addition, expanding our model to support 3D holography by training with diverse 3D datasets could significantly broaden its utility, particularly in advanced VR/AR systems [46].

5 Conclusions

In this work, we have presented a neural network-based approach for efficiently generating and compressing POHs with RGB input, using a pre-trained camera-calibrated wave propagation model for a real-world display system. This approach, crucial for practical holography, compensates for hardware imperfections and ensures high display fidelity without requiring optical filters. The compact, filter-free design offers a significant advantage, especially in VR/AR applications where form factor is a key issue. With a decoding time of around 40 ms per color channel, our approach is suitable for real-time holography deployed on edge devices, combining high performance with efficiency. Looking ahead, our approach provides a foundation for further advances in 3D holography, which could bring significant advances to VR/AR displays by providing enhanced depth perception and immersion. By addressing both computational efficiency and practical design constraints, our work lays the groundwork for the next generation of holographic display technologies, bringing them closer to real-world deployment.

Appendix A Fundamentals of wave propagation

In holographic displays, a coherent wavefront illuminates the SLM, where each pixel of the SLM modulates the incoming wavefront according to a phase-only hologram $\phi(x, y)$. The modulated wave $Ae^{j\phi}$ is then propagated from the SLM plane to a target plane using the well-established angular spectrum method (ASM) [29]. Wave propagation from the SLM plane to the target plane can be efficiently modeled using the Fourier transform. This process can be expressed as

$$f^d(U, \phi) = \mathcal{F}^{-1}\left(\mathcal{F}\left(Ae^{j\phi}\right) \cdot H(f_x, f_y)\right) \quad (3)$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the Fourier transform and its inverse, respectively. The transfer function $H(f_x, f_y)$ is given by

$$H(f_x, f_y) = \begin{cases} e^{j2\pi d\sqrt{1/\lambda^2 - f_x^2 - f_y^2}}, & \text{if } \sqrt{f_x^2 + f_y^2} < 1/\lambda \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where λ represents the wavelength of the light, f_x and f_y are spatial frequencies, and d denotes the distance between the SLM and the target plane. The transfer function $H(f_x, f_y)$ with band-limited mask restricts the propagation to a finite spatial frequency range determined by λ, as detailed in Ref. [29].

Appendix B Scheme of holographic displays

A system schematic widely used for holographic near-eye displays [4–6, 21] is given in Fig. 9. A laser emits coherent light, which is then collimated by a lens before reaching the SLM. At the SLM, the phase of the light wave is modulated pixel-wise. As the wavefront continues to propagate further, interferes in space, creating an intensity pattern that is perceived by the user through an eyepiece. The hologram to display on the SLM, usually phase-only, can be generated using various CGH algorithms.

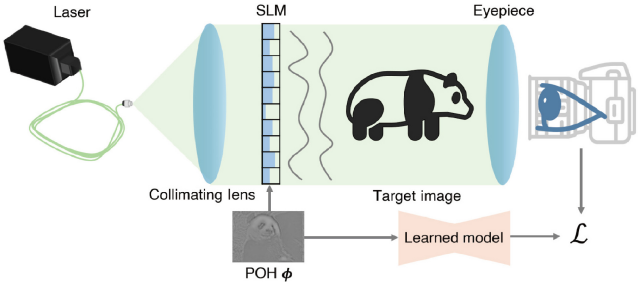


Fig. 9 Schematic of a common holographic near-eye display configuration, showing how the user can perceive the reconstructed images in 3D space and how the intensity of the holographic images is captured to train the camera-calibrated wave propagation model. The key to CGH research is how to synthesize a phase-only hologram that can be displayed on the SLM.

Appendix C Learned wave propagation model

To train our learned wave propagation model, we display a sequence of phase-only holograms on the SLM and, with our prototype's focus-tunable camera, record the resulting images at eight different focal planes. To maximize the dataset's variability, we have compiled a library of phase-only holograms derived from both 2D images at various depths and 3D scenes, generated by a mix of direct and iterative CGH techniques: DPAC [23], SGD [5], and HOGD [9]. In

all, the training set comprises 5860 unique phase patterns and 46,880 captured images.

Our model follows the learned wave-propagation framework reported by Choi et al. [6]. As illustrated in Fig. 10, the input phase is first corrected by a UNet, i.e., SLM CNN, that compensates for SLM crosstalk and phase non-uniformity, outputting the complex field at the SLM plane. We then propagate this field to eight target planes via ASM. Each of the eight resulting fields is passed through a second UNet, i.e., a target CNN, to capture residual optical aberrations. During training, the final outputs are supervised by the captured dataset introduced above.

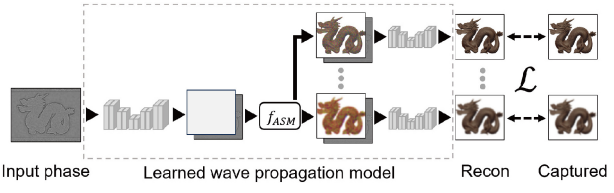


Fig. 10 Training the camera-calibrated, learned wave propagation model. The dual-UNet architecture with ASM propagation is adopted.

Both UNets accept two input channels (real and imaginary) and produce two output channels. The SLM CNN consists of 8-strided convolution downsampling layers followed by 8-transposed convolution upsampling layers. Its first layer produces 32 feature maps, which double at each downsampling stage up to 512 channels. The target CNN uses 5 downsampling and 5 upsampling layers, starting at 8 channels and reaching 128 at its deepest level. In both networks we apply instance normalization, leaky ReLU activations in the encoder blocks, and ReLU activations in the decoder blocks.

Availability of data and materials

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

Author contributions

Hyunmin Ban and Yifan Peng conceived the idea. Hyunmin Ban conducted the multi-color hologram compression and ran the simulation. Hyunmin Ban, Wenbin Zhou, and Xiangyu Meng performed the experiments on the holographic display. Yifan Peng supervised the project. All authors discussed the data and wrote the manuscript.

Acknowledgements

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

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

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* Vol. 7, No. 11, Article No. 1563, 2020.

[2] Kim, J.; Gopakumar, M.; Choi, S.; Peng, Y.; Lopes, W.; Wetzstein, G. Holographic glasses for virtual reality. In: Proceedings of the ACM SIGGRAPH Conference Proceedings, 2022.

[3] Jang, C.; Bang, K.; Chae, M.; Lee, B.; Lanman, D. Waveguide holography for 3D augmented reality glasses. *Nature Communications* Vol. 15, Article No. 66, 2024.

[4] Gopakumar, M.; Lee, G. Y.; Choi, S.; Chao, B.; Peng, Y.; Kim, J.; Wetzstein, G. Full-color 3D holographic augmented-reality displays with metasurface waveguides. *Nature* Vol. 629, 791–797, 2024.

[5] Peng, Y.; Choi, S.; Padmanaban, N.; Wetzstein, G. Neural holography with camera-in-the-loop training. *ACM Transactions on Graphics* Vol. 39, No. 6, Article No. 185, 2020.

[6] 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* Vol. 40, No. 6, Article No. 240, 2021.

[7] 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* Vol. 70, 114–130, 2019.

[8] Chakravarthula, P.; Tseng, E.; Srivastava, T.; Fuchs, H.; Heide, F. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Transactions on Graphics* Vol. 39, No. 6, Article No. 186, 2020.

[9] 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* Vol. 46, No. 23, 5822–5825, 2021.

[10] Sullivan, G. J.; Ohm, J. R.; Han, W. J.; Wiegand, T. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology* Vol. 22, No. 12, 1649–1668, 2012.

[11] Wallace, G. K. The JPEG still picture compression standard. *Communications of the ACM* Vol. 34, No. 4, 30–44, 1991.

[12] Oh, K. J.; Ban, H.; Choi, S.; Ko, H.; Kim, H. Y. HEVC extension for phase hologram compression. *Optics Express* Vol. 31, No. 6, 9146–9164, 2023.

[13] Blinder, D.; Schretter, C.; Schelkens, P. Global motion compensation for compressing holographic videos. *Optics Express* Vol. 26, No. 20, 25524–25533, 2018.

[14] Zhou, M.; Zhang, H.; Jiao, S.; Chakravarthula, P.; Geng, Z. End-to-end compression-aware computer-generated holography. *Optics Express* Vol. 31, No. 26, 43908–43919, 2023.

[15] Jiao, S.; Jin, Z.; Chang, C.; Zhou, C.; Zou, W.; Li, X. Compression of phase-only holograms with JPEG standard and deep learning. *Applied Sciences* Vol. 8, No. 8, Article No. 1258, 2018.

[16] Shi, L.; Webb, R.; Xiao, L.; Kim, C.; Jang, C. Neural compression for hologram images and videos. *Optics Letters* Vol. 47, No. 22, Article No. 6013, 2022.

[17] 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* Vol. 41, No. 4, Article No. 110, 2022.

[18] Ban, H.; Choi, S.; Cha, J. Y.; Kim, Y.; Kim, H. Y. NHVC: Neural holographic video compression with scalable architecture. In: Proceedings of the IEEE Conference Virtual Reality and 3D User Interfaces, 969–978, 2024.

[19] Jia, J.; Dong, Z.; Ling, Y.; Li, Y.; Su, Y. Deep learning-based approach for efficient generation and transmission of high-definition computer-generated holography. In: Proceedings of the Advances in Display Technologies XIV, 35–41, 2024.

[20] Kuo, G.; Waller, L.; Ng, R.; Maimone, A. High resolution étendue expansion for holographic displays. *ACM Transactions on Graphics* Vol. 39, No. 4, Article No. 66, 2020.

[21] Choi, S.; Gopakumar, M.; Peng, Y.; Kim, J.; O'Toole, M.; Wetzstein, G. Time-multiplexed neural holography: A flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. In: Proceedings of the ACM SIGGRAPH Conference, Article No. 32, 2022.

[22] Hsueh, C. K.; Sawchuk, A. A. Computer-generated double-phase holograms. *Applied Optics* Vol. 17, No. 24, 3874–3883, 1978.

[23] Maimone, A.; Georgiou, A.; Kollin, J. S. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics* Vol. 36, No. 4, Article No. 85, 2017.

[24] Liu, K.; Wu, J.; He, Z.; Cao, L. 4K-DMDNet: Diffraction model-driven network for 4K computer-generated holography. *Opto-Electronic Advances* Vol. 6, No. 5, Article No. 220135, 2023.

[25] Zhou, W.; Qu, F.; Meng, X.; Li, Z.; Peng, Y. 3D-HoloNet: Fast, unfiltered, 3D hologram generation with camera-calibrated network learning. *Optics Letters* Vol. 50, No. 4, 1188–1191, 2025.

[26] Gerchberg, R. W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* Vol. 35, 237–246, 1972.

[27] Peng, Y.; Dun, X.; Sun, Q.; Heidrich, W. Mix-and-match holography. *ACM Transactions on Graphics* Vol. 36, No. 6, Article No. 191, 2017.

[28] Chakravarthula, P.; Peng, Y.; Kollin, J.; Fuchs, H.; Heide, F. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics* Vol. 38, No. 6, Article No. 213, 2019.

[29] Matsushima, K.; Shimobaba, T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Optics Express* Vol. 17, No. 22, 19662–19673, 2009.

[30] Balle, J.; Minnen, D.; Singh, S.; Hwang, S. J.; Johnston, N. Variational image compression with a scale hyperprior. In: Proceedings of the International Conference on Learning Representations, 2018.

[31] Rissanen, J.; Langdon, G. Universal modeling and coding. *IEEE Transactions on Information Theory* Vol. 27, No. 1, 12–23, 1981.

[32] Minnen, D.; Balle, J.; Toderici, G. D. Joint autoregressive and hierarchical priors for learned image compression. *arXiv preprint* arXiv:1809.02736, 2018.

[33] Minnen, D.; Singh, S. Channel-wise autoregressive entropy models for learned image compression. In: Proceedings of the IEEE International Conference on Image Processing, 3339–3343, 2020.

[34] Liu, J.; Sun, H.; Katto, J. Learned image compression with mixed transformer-CNN architectures. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14388–14397, 2023.

[35] 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. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5708–5717, 2022.

[36] Bengio, Y.; Léonard, N.; Courville, A. Estimating or

propagating gradients through stochastic neurons for conditional computation. *arXiv preprint* arXiv:1308.3432, 2013.

[37] Theis, L.; Shi, W.; Cunningham, A.; Huszar, F. Lossy image compression with compressive autoencoders. *arXiv preprint* arXiv:1703.00395, 2017.

[38] Agustsson, E.; Timofte, R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1122–1131, 2017.

[39] Duda, J. Asymmetric numeral systems: Entropy coding combining speed of Huffman coding with compression rate of arithmetic coding. *arXiv preprint* arXiv:1311.2540, 2013.

[40] Bross, B.; Wang, Y. K.; Ye, Y.; Liu, S.; Chen, J.; Sullivan, G. J.; Ohm, J. R. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology* Vol. 31, No. 10, 3736–3764, 2021.

[41] JVET. VTM-23.4: Versatile video coding test model (VTM). 2024. Available at `https://jvet.hhi.fraunhofer.de/`

[42] Bjontegaard, G. Calculation of average PSNR differences between RD-curves. ITU-T SG16 Q. 2001.

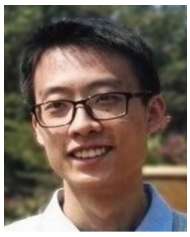
[43] Padmanaban, N.; Peng, Y.; Wetzstein, G. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics* Vol. 38, No. 6, Article No. 214, 2019.

[44] Wang, G. H.; Li, J.; Li, B.; Lu, Y. EVC: Towards real-time neural image compression with mask decay. *arXiv preprint* arXiv:2302.05071, 2023.

[45] Kim, H.; Bauer, M.; Theis, L.; Schwarz, J. R.; Dupont, E. C3: High-performance and low-complexity neural compression from a single image or video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9347–9358, 2024.

[46] Wang, M.; Lyu, X. Q.; Li, Y. J.; Zhang, F. L. VR content creation and exploration with deep learning: A survey. *Computational Visual Media* Vol. 6, No. 1, 3–28, 2020.

Hyunmin Ban received his master and bachelor degrees in computer science and engineering from Kyung Hee University. He is working towards his Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong. His research interests include 3D imaging acquisition, holographic displays, learning-based video compression, and computational imaging.



Wenbin Zhou received his bachelor degree in physics from the University of Science and Technology of China and his master degree in computer science from the University of Hong Kong. He is working towards his Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong. His research interests include holography, optics, computer graphics, and virtual and augmented reality.

Xiangyu Meng received his bachelor degree in mathematics and his master degree in computer science from Zhejiang University. He is working toward his Ph.D. degree in the Department of Electrical and Electronic Engineering at the University of Hong Kong. His research interests include holography displays, computational optics, and computer graphics.

Yifan (Evan) Peng is an assistant professor in the Departments of Electrical & Electronic Engineering and Computer Science, the University of Hong Kong. He was a postdoctoral fellow at Stanford University. He received his Ph.D. degree in computer science from the University of British Columbia, and his master and bachelor degrees in optical science & engineering, from Zhejiang University. His research focuses on incorporating optical and computational techniques to enable new imaging modalities.

To submit a manuscript, please go to `https://jcvm.org`.