

PHOTONICS Research

Physics-aware cross-domain fusion aids learning-driven computer-generated holography

GANZHANGQIN YUAN,^{1,†} MI ZHOU,^{1,†}  FEI LIU,²  MU KU CHEN,³  KUI JIANG,⁴ YIFAN PENG,^{5,6} 
AND ZIHAN GENG^{1,*} 

¹Institute of Data and Information, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518071, China

²School of Optoelectronic Engineering, Xidian University, Xi'an 710071, China

³Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China

⁴School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

⁵Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China

⁶Department of Computer Science, The University of Hong Kong, Hong Kong SAR, China

[†]These authors contributed equally to this work.

*Corresponding author: geng.zihan@sz.tsinghua.edu.cn

Received 19 April 2024; revised 17 August 2024; accepted 17 August 2024; posted 19 August 2024 (Doc. ID 527405);
published 13 November 2024

The rapid advancement of computer-generated holography has bridged deep learning with traditional optical principles in recent years. However, a critical challenge in this evolution is the efficient and accurate conversion from the amplitude to phase domain for high-quality phase-only hologram (POH) generation. Existing computational models often struggle to address the inherent complexities of optical phenomena, compromising the conversion process. In this study, we present the cross-domain fusion network (CDFN), an architecture designed to tackle the complexities involved in POH generation. The CDFN employs a multi-stage (MS) mechanism to progressively learn the translation from amplitude to phase domain, complemented by the deep supervision (DS) strategy of middle features to enhance task-relevant feature learning from the initial stages. Additionally, we propose an infinite phase mapper (IPM), a phase-mapping function that circumvents the limitations of conventional activation functions and encapsulates the physical essence of holography. Through simulations, our proposed method successfully reconstructs high-quality 2K color images from the DIV2K dataset, achieving an average PSNR of 31.68 dB and SSIM of 0.944. Furthermore, we realize high-quality color image reconstruction in optical experiments. The experimental results highlight the computational intelligence and optical fidelity achieved by our proposed physics-aware cross-domain fusion. © 2024 Chinese Laser Press

<https://doi.org/10.1364/PRJ.527405>

1. INTRODUCTION

Holographic technology, with its unique ability to reconstruct images, stands at the forefront of advancing optical displays, offering realism and immersive qualities [1–10]. The quest for dynamic, high-quality computer-generated holography (CGH) has long been driven by the dual research in optical engineering and computational science [11–15]. Despite remarkable progress [16,17], the challenge of efficiently generating phase-only holograms (POHs) remains, primarily due to the intricate nature of modulating light with spatial light modulators (SLMs) to achieve desired visual effects without compromising on quality or computational demand [17].

To date, the typical methods for generating POHs can be divided into iterative and non-iterative ones. Iterative methods, including the Gerchberg–Saxton (GS) algorithm [18–20], non-convex optimization algorithms [21,22], and the stochastic gradient descent method (SGD) [23,24], facilitate accurate

hologram generation but fall short in terms of efficiency, hindering real-time applications. In contrast, non-iterative methods such as double phase-amplitude coding (DPAC) and its variants [22,25–28], enable quicker hologram generation in fewer steps, but this comes from a significant quality decline. Therefore, a compromise between fidelity and efficiency is inevitable.

Recent advancements in computational techniques, particularly deep learning, have provided promising solutions to long-standing problems in CGH [29–35]. These approaches, leveraging the vast computational power and flexibility of neural networks, have shown potential in transcending traditional barriers, enabling more control over the hologram generation process. By automating the intricate phase modulation required for POHs, deep learning models have facilitated significant strides towards achieving real-time, high-fidelity holographic displays.

During the exploration of the combination of neural networks with CGH, the UNet architecture [36] has emerged as the processing backbone in most cases due to its effective feature-extraction capabilities and encoder-decoder design [29–31,34,35,37,38]. However, adapting UNet to the specific requirements of CGH presents a significant challenge. The fundamental issue stems from the fact that the UNet's architecture may not be optimally designed for the intricate cross-domain transformations inherent in CGH. These transformations involve converting color images from the amplitude domain to phase-only holograms in the phase domain, a critical process that demands precise handling to ensure fidelity and accuracy in hologram reconstruction [39]. Furthermore, the traditional activation functions, borrowed from 2D image processing, map output values to a fixed interval through nonlinear variations. This oversimplification of phase values overlooks their intrinsic periodic nature, leading to potential inaccuracies in the reconstructed holograms.

In recognition of these challenges, our study introduces an approach that combines the computational power of neural networks with the rigorous requirements of optical physics. The cross-domain fusion network (CDFN) is designed to facilitate the efficient and accurate conversion from the amplitude domain to the phase domain. By dissecting the cross-domain conversion process into multiple stages and incorporating the physics-aware mapping function, CDFN significantly enhances the quality and efficiency of hologram generation. This innovation addresses the technical hurdles inherent in CGH and aligns with the pursuit of optical excellence. Figure 1 presents a visual comparison between conventional approaches and our approach. The results of various experiments demonstrate the capability of our method to reconstruct high-quality 2K color images both in simulation and real-world experiments.

2. METHOD

A. Hologram Generation Workflow

Figure 2 delineates the workflow of our proposed method. The algorithm begins by inputting the amplitude image into the first sub-network. The next step involves combining the amplitude image with the initial phase image output from the first sub-network to represent the complex hologram at the object plane. After this integration, the algorithm utilizes the angular spectrum method (ASM) to simulate the forward propagation to compute the complex hologram at the SLM plane, which simulates the physical process of light propagation, bridging the digital world with the real-world holograms. The complex hologram at the SLM plane is then fed into the second sub-network. This network is trained to translate the complex hologram into the final POH, which can be loaded into the SLM.

B. Multi-stage Network Architecture

The two sub-networks in our algorithm share a similar structure, which is a derivative of the UNet used in HoloNet [30]. But considering the cross-domain conversion problem in CGH, we break down the cross-domain conversion process into multiple stages by constructing more intermediate feature maps and thus designing a multi-stage network architecture.

The multi-stage architecture is motivated by the need for mapping from the amplitude domain to a different phase domain in POH generation. Traditional UNet architecture directly forwards high-resolution feature maps from the encoder to the decoder. While this approach is efficient in many applications, it is less effective in the CGH context due to the differing nature of amplitude values in the encoder and phase values in the decoder [39].

To tackle this issue, our network consists of more intermediate feature maps representing multiple stages, which serve as a buffer and enable a more gradual conversion. These stages

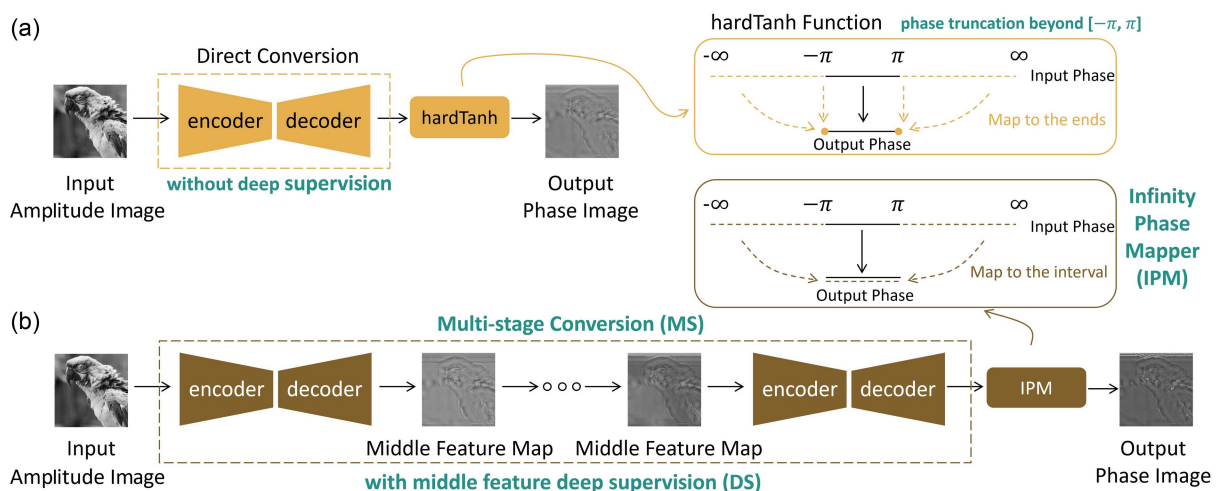


Fig. 1. (a) Workflow of conventional methods based on the vanilla UNet architecture. (b) Our proposed cross-domain fusion network highlights the role of our multi-stage conversion architecture and the infinity phase mapper (IPM) in accomplishing the cross-domain transformation task. The multi-stage conversion architecture employs multiple feature maps to facilitate a gradual transformation between two domains. Concurrently, the infinity phase mapper is designed to accommodate the periodic nature of phase values, ensuring the preservation of the physical consistency.

resolve the discrepancies between the encoder inputs (amplitude values) and the final decoder outputs (phase values). As depicted in Fig. 2, the feature maps representing the various conversion stages are color-coded diversely. This design guarantees a smoother and more cohesive conversion from the amplitude domain to the phase domain, minimizing sudden changes and improving the results.

To prove the effectiveness of the multi-stage design, we try to exclude the impact from different numbers of network parameters by keeping a similar number of parameters with HoloNet [30]. Specifically, we downsize the top-down feature maps' channel numbers from [32, 64, 128, 256] to [16, 32, 64, 128]. As shown in Table 2 presented later, our algorithm's network has fewer parameters than HoloNet.

C. Infinity Phase Mapper

In CGH, it is necessary for the network to predict phase values when generating POH. To get a phase value in the range of $[-\pi, \pi]$, existing algorithms append a traditional activation function, e.g., a hardTanh function in HoloNet [30], after the final output layer of the network to restrict the value in the required range. However, this approach will introduce inconsistencies in the predicted phase values. Ideally, the phase angle, represented by the network's raw output value, could span an infinity range $[-\infty, \infty]$. A well-defined mapping function should consider the periodicity of the phase angle and map the value in $[-\infty, \infty]$ to a corresponding value in $[-\pi, \pi]$. But most activation functions simply do a truncating and map values beyond $[-\pi, \pi]$ to the endpoints $-\pi$ and π , resulting

in a misrepresented mapping of phase values that significantly impacts the results.

To address this problem, we propose a physics-aware mapping function. The primary objective of this function is to ensure a more accurate mapping of output values, carefully mapping each value within the desired $[-\pi, \pi]$ range. The specific formulation of this function is detailed below:

$$\phi_o = \mathcal{M}(\phi_i) = \text{atan2}(\sin \phi_i, \cos \phi_i),$$

$$\text{atan2}(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } x > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{if } y \geq 0, x < 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & \text{if } y < 0, x < 0 \\ +\frac{\pi}{2} & \text{if } y > 0, x = 0 \\ -\frac{\pi}{2} & \text{if } y < 0, x = 0 \\ \text{undefined} & \text{if } y = 0, x = 0 \end{cases} \quad (1)$$

where the phase mapping function, denoted as \mathcal{M} , is defined by the input phase value ϕ_i and the output phase value ϕ_o . The function $\text{atan2}(\cdot)$ is a modified version of the \arctan function that considers all four quadrants when computing the angle from the positive x -axis to the point (x, y) .

This specific phase mapping function, named the infinity phase mapper, translates any phase value in the infinite range into its corresponding value in $[-\pi, \pi]$. This approach avoids the limitations of implementing hard truncation. The graphical representation of the infinity phase mapper is shown in Fig. 3.

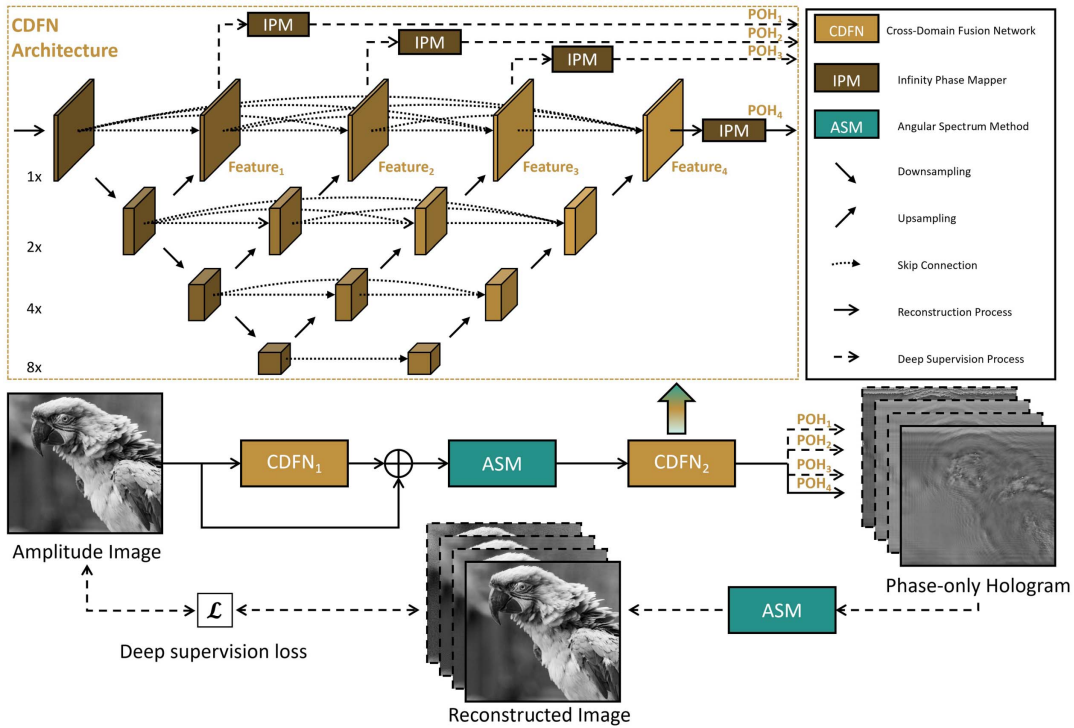


Fig. 2. Hologram generation workflow of our proposed method and the corresponding network architecture. The feature maps representing the various conversion stages are color-coded diversely. The first sub-network CDFN₁ is used to predict the initial phase and the second sub-network CDFN₂ is used to predict POHs.

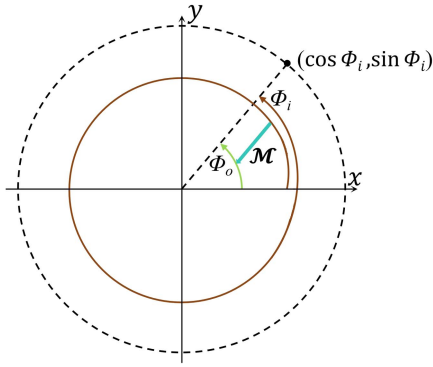


Fig. 3. Schematic of infinity phase mapper (IPM) \mathcal{M} , which maps infinite phase values ϕ_i to their corresponding points ϕ_o within the $[-\pi, \pi]$ interval.

D. Deep Supervision in the Learning Process

To improve our method's performance, we have incorporated a deep supervision strategy into the training process. When training the network, we supervise the loss between the reconstructed image from the POH and the target image. Additionally, we supervise the losses between the reconstructed images from top multi-stage feature maps and the target image. By applying supervision to the intermediate feature maps in the initial stages of the network, the network is encouraged to learn more task-relevant features early in the model. This is critical because features learned in the initial stage form the foundation for subsequent stages. Deep supervision ensures that these foundational features are sensitive to cross-domain tasks, enhancing the overall performance of the network for cross-domain transformations. The loss function employed under this deep supervision paradigm can be formulated as follows:

$$\mathcal{L}_{\text{deep}}(I, \hat{I}_i) = \sum_{i=1}^D \alpha_i \|\hat{I}_i - I\|^2, \quad (2)$$

where the target amplitude image is denoted as I , and the amplitude image reconstructed at the i th level is represented as \hat{I}_i . D is the same as the number of the network depth, which is 4 in the implementation. The weighting coefficient α_i is used to balance the losses at different levels, with i indexing the intermediate feature images, which is 0.25, 0.5, 0.75, and 1.0 in the implementation.

Moreover, because of the importance of the final POH, special attention is given during the training. An extra perceptual loss \mathcal{L}_p [40] is computed between the reconstructed image from the final POH and the target image. Therefore, the overall loss function is formulated as follows:

$$\mathcal{L}_{\text{overall}}(I, \hat{I}_i) = \mathcal{L}_p(I, \hat{I}_D) + \mathcal{L}_{\text{deep}}(I, \hat{I}_i). \quad (3)$$

This overall loss, taking into account both the deep supervision loss from multi-stage feature maps and the perceptual loss related to the final POH, guarantees the algorithm's effectiveness in producing high-quality POHs.

3. EXPERIMENT

The experiments are performed using a 24 GB NVIDIA RTX 3090 GPU. The DIV2K dataset [41] is utilized for both training and testing. The training set consists of 800 images and the testing set consists of 100. During the training process, data augmentation techniques, such as image flipping and rotation, are employed to facilitate the learning process, similar to HoloNet [30]. The training is limited to a maximum of 20 epochs and the learning rate is set as 0.001. For the ASM model used in our algorithm, specific parameters are defined as follows: the diffraction distance is set as 0.2 m, the SLM's pixel pitch is set as $6.4 \mu\text{m}$, and the wavelengths for the red, green, and blue lights are 638, 520, and 450 nm, respectively.

A. Numerical Simulation

As shown in Fig. 4, we first compare the numerical simulation results of our proposed CDFN with other methods such as non-iterative method DPAC [26], iterative method SGD [16], and learning-based methods HoloNet [30] and CCNN [38] in the DIV2K dataset. For consistency, we first resize RGB images to 1600×880 pixels and then add zero padding to 1920×1072 pixels. Figure 4 shows enlarged sections to better highlight the differences between these methods. Our results demonstrate CDFN's ability to capture fine details and ensure structural integrity, proving its effectiveness and superiority.

Table 1 presents the quantitative comparison of various methods. Among all methods, SGD achieves the best evaluation results in PSNR, while CDFN excels in SSIM. Among the learning-based methods, under the same training conditions, our proposed CDFN achieves the highest performance, surpassing the current state-of-the-art method CCNN. This demonstrates that our method has a certain competitive advantage over current CGH approaches.

Table 2 compares the computational efficiency of SGD, HoloNet, and our CDFN method, including metrics like computational time, parameter count, and floating-point operations (Flops). CDFN and HoloNet, as learning-based methods, show a significant speed advantage over the traditional SGD algorithm. Despite CDFN's use of a greater number of intermediate feature maps—which increases its computational demand—it optimizes performance by reducing convolution channel numbers. Consequently, CDFN boasts fewer parameters (210,000) than HoloNet (287,000), underscoring that its enhanced performance stems from an effective network design rather than mere parameter count.

Our CDFN method represents a significant advancement in CGH technology, striking a balance between reconstruction accuracy and computational efficiency. This is evidenced by its performance in PSNR and SSIM evaluations. CDFN distinguishes itself by employing multi-stage feature maps. This strategic approach allows CDFN to outperform the state-of-the-art methods. Consequently, CDFN has established itself as a sophisticated and effective CGH method.

B. Ablation Study

In our ablation study, we test the diverse strategies integral to our proposed method, beginning with an examination of the pure multi-stage (MS) architecture, illustrated in Fig. 5. This

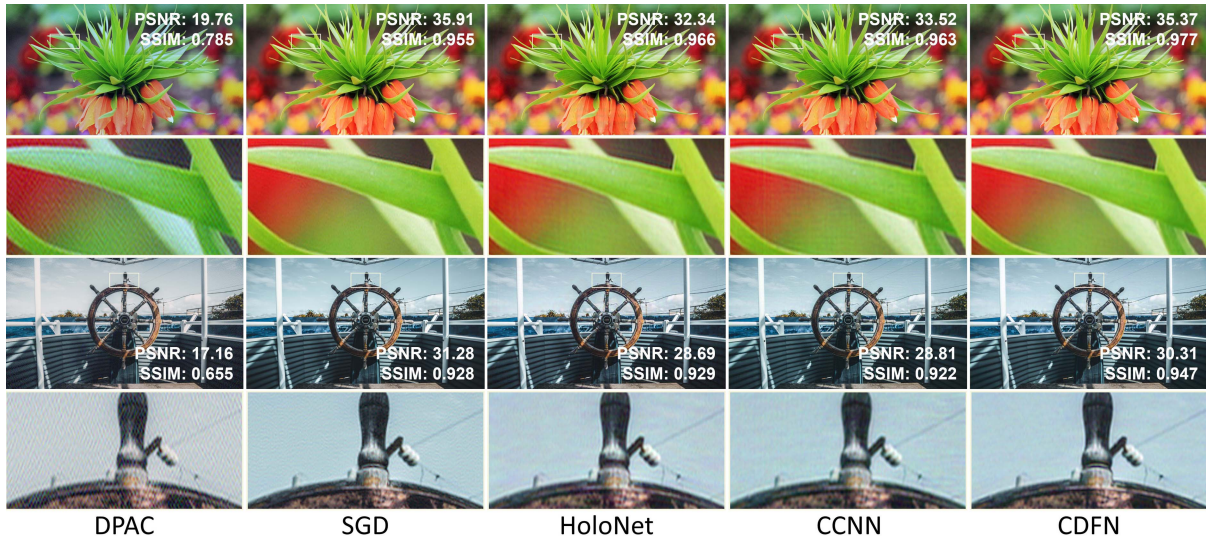


Fig. 4. Comparison of numerically reconstructed color images. From left to right: results of double phase-amplitude coding (DPAC), stochastic gradient descent method (SGD), HoloNet, CCNN, and our proposed CDFN, respectively (PSNR in dB).

Table 1. Quantitative Results of Different Methods Tested on the DIV2K Testing Dataset, which Consists of 100 Images (Color Channels)^a

	DPAC	SGD	HoloNet	CCNN	CDFN (Ours)
PSNR/SSIM	19.97/0.689	32.69/0.942	29.87/0.926	30.72/0.927	31.68/0.944

^aEvaluation metrics include PSNR (dB) and SSIM.

Table 2. Efficiency Comparison for Three CGH Methods Tested on the DIV2K Testing Dataset, which Consists of 100 Images with a Resolution of 1920 × 1072 Pixels^a

	SGD	HoloNet	CDFN (Ours)
Time (s)	16.851	0.010	0.012
Parameter quantity	—	2.87×10^5	2.10×10^5
FLOPs	—	3.29×10^{11}	3.39×10^{11}

^aEvaluation metrics include methods' average inference time, number of parameters, and the floating point of operations (FLOPs). The evaluation is performed using a 24 GB NVIDIA RTX 3090 GPU.

foundational aspect of our approach, even in its most basic form, showcases high reconstruction accuracy, outperforming HoloNet.

Quantitative evidence, as detailed in Table 3, underscores this performance advantage. The MS architecture attained a PSNR of 30.88 dB and an SSIM of 0.932, surpassing the corresponding metrics of HoloNet which are 30.15 and 0.926, respectively.

Our analysis extends to evaluating enhancements within the multi-stage architecture through the incorporation of an infinity phase mapper (MS w IPM) and deep supervision (MS w DS), individually and in combination (MS w IPM&DS). The addition of the infinity phase mapper alone elevates the reconstruction accuracy, achieving a PSNR of 32.13 dB and an SSIM of 0.948. This improvement underscores

the crucial contribution of the phase mapping function to the algorithm's performance.

Further integration of deep supervision into the architecture (MS w DS) marks another leap in reconstruction quality, with PSNR increasing to 31.65 dB and SSIM to 0.940. This indicates that deep supervision can improve the quality of reconstruction.

The peak performance is realized when both the phase mapping function and deep supervision are applied together (MS w IPM&DS), culminating in a PSNR of 32.26 dB and an SSIM of 0.948. These outstanding results not only highlight the individual strengths of each strategy but also the superior reconstruction accuracy achieved through their combined implementation, showcasing a significant synergistic effect.

To further explore the impact of the IPM on reconstruction quality, we analyze the distribution of phase values in POHs generated under different experimental conditions, as shown in Fig. 6. Notably, the use of the hardTanh activation function, as implemented in HoloNet and CDFN w/o IPM, introduces a discontinuity around π . This discontinuity arises from hardTanh's insensitivity to phase periodicity. However, in the CDFN w IPM, this discontinuity is mitigated, leading to enhanced reconstruction quality. The improved metrics with the implementation of IPM, as shown in Table 3, support this finding and reinforce our hypothesis that the IPM enhances the accuracy of POH generation by considering phase periodicity. These results also illustrate the unique properties of holographic

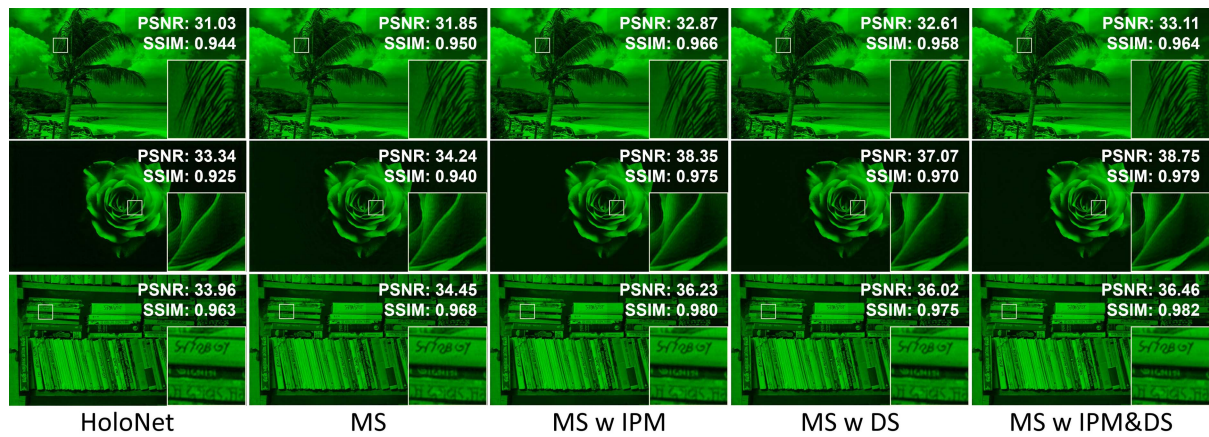


Fig. 5. Comparison of simulated reconstruction images in our ablation study (green channel). From left to right: reconstruction results of HoloNet, multi-stage architecture (MS), multi-stage architecture with infinity phase mapper (MS w IPM), multi-stage architecture with deep supervision (MS w DS), multi-stage architecture with infinity phase mapper and deep supervision (MS w IPM&DS) (PSNR in dB).

Table 3. Ablation Study with Average PSNR (dB) and SSIM Metrics on the DIV2K Testing Dataset, which Consists of 100 Images (Green Channel)

	HoloNet	MS	MS w IPM	MS w DS	MS w IPM&DS
PSNR/SSIM	30.15/0.926	30.88/0.932	32.13/0.948	31.65/0.940	32.26/0.948

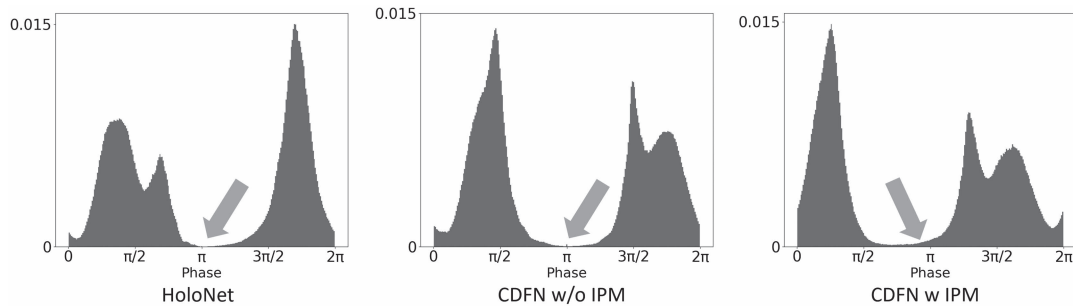


Fig. 6. Normalized phase value distribution of a randomly picked POH. From left to right: results of HoloNet, CDFN without infinity phase mapper (CDFN w/o IPM), CDFN with infinity phase mapper (CDFN w IPM). Note that we shift the value in $[-\pi, \pi]$ to $[0, 2\pi]$ in this figure for the easier observation of values around π . The arrows highlight that the IPM can generate a continuous phase value distribution around the π value, where the conventional active function generates a gap.

data compared to traditional 2D image data, emphasizing the need for specialized consideration.

C. Optical Experiment

Our algorithm's effectiveness is further validated through optical reconstruction experiments. In the experiments, we load the computed 1920×1072 pixels POHs into a 1920×1080 SLM and capture images using a Sony A7 Mark III camera. The optical display system setup is illustrated in Fig. 7. The laser source used in this setup is the FISBA READY Beam, which emits light at wavelengths of 638, 520, and 450 nm for the red, green, and blue channels, respectively. After passing through a collimating lens, the beam is split into two paths by a beam splitter. The incident beam is then modulated by a HOLOEYE LETO-3-CFS-127 SLM, which has a resolution

of 1920×1080 pixels and a pixel pitch of $6.4 \mu\text{m}$. The modulated beam is reflected, and an aperture is used to filter out undesired high diffraction orders. The target plane is situated at a distance of 0.2 m from the SLM, and the image plane is determined by the thin lens equation.

Figure 8 illustrates the optical reconstruction results obtained with green light. To enhance clarity, magnified patches of the reconstructed images are also presented. In this experiment, we evaluate the performance of our CDFN in comparison to other methods like GS [18], DPAC [26], SGD [23], and HoloNet [30]. These optical experiments consistently match our simulation expectations. DPAC's reconstructions struggle with accurately depicting fine details, as seen in the parrot's feathers and the butterfly's intricacy. The SGD method's

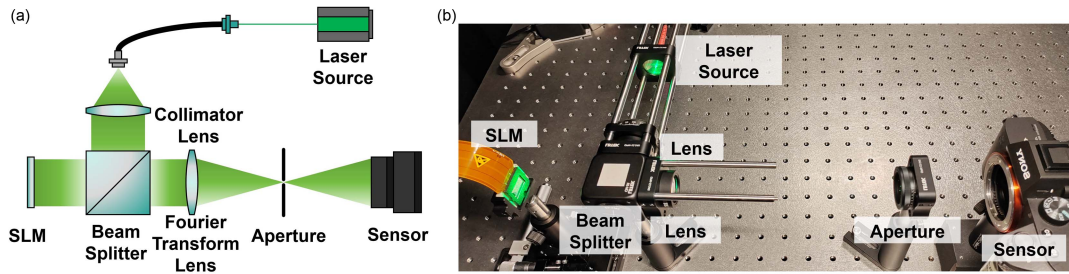


Fig. 7. Holographic display setup. (a) Schematic diagram of the optical display system setup. (b) Photograph of the optical display system setup.

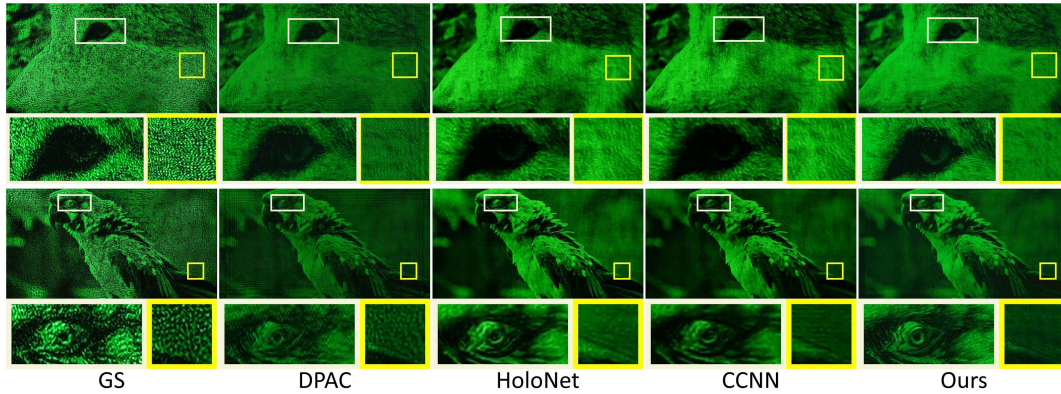


Fig. 8. Optical reconstruction images in the green channel. From left to right: results of Gerchberg–Saxton algorithm (GS), double phase-amplitude coding (DPAC), HoloNet, CCNN, and CDFN, respectively.

reconstructions are marred by significant speckle noise, obfuscating delicate features. In contrast, our CDFN method overcomes these issues, presenting reconstructions with enhanced clarity.

For full-color imaging, as shown in Fig. 9, we sequentially loaded the SLM with POHs for each color channel, synchronously activating the corresponding laser color. The detailed results, showcased in Fig. 10, affirm our method's capability to reconstruct full-color images with high fidelity. The experimental outcomes not only underscore the practical effectiveness of our approach but also its potential for advancing optical reconstruction technologies.

To evaluate our method's generalization ability across different data types, we test the model trained on the color DIV2K dataset using binary USAF images. Figure 11 shows the results from both simulation and optical experiments. The PSNR and SSIM metric results for the three images are as follows: 18.37 dB and 0.604, 17.03 dB and 0.643, and 20.71 dB and 0.781, respectively. While these results confirm the applicability of our method from color to binary images, they also highlight certain challenges. Notably, some speckle noise can be observed in the reconstructions. This occurs because binary images, with their high contrast between black and white regions, introduce significant high-frequency components that are

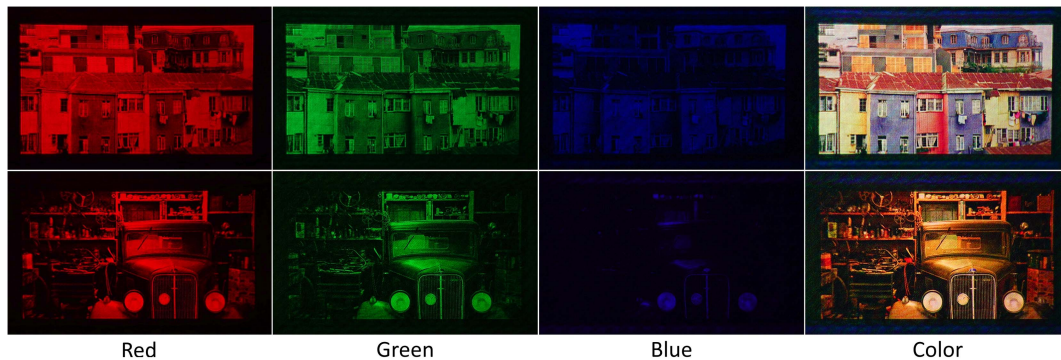


Fig. 9. Optical reconstruction images from 1920×1072 pixels POHs on a 1920×1080 SLM in red, green, blue, and color channels. The images are directly captured by a camera. The color image is obtained by synchronizing the three-color laser source and sequentially loading different POHs. From left to right: results in red, green, blue, and color channels, respectively.

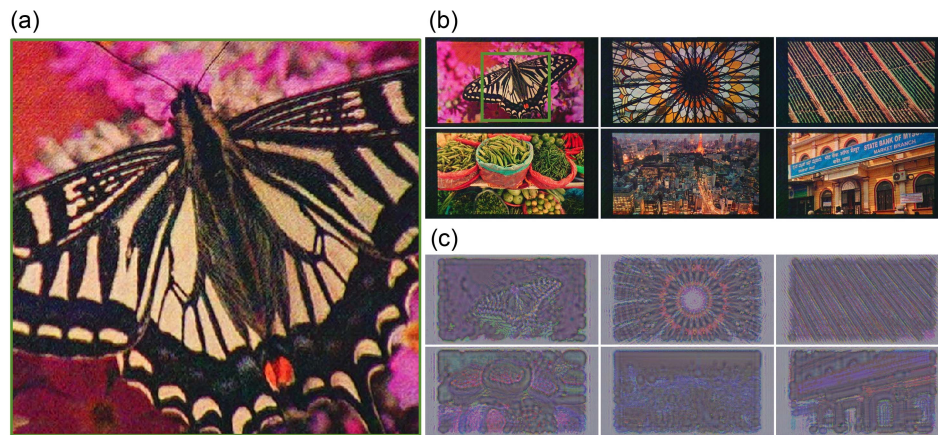


Fig. 10. Optical reconstruction images of our method in color channels directly captured by a camera and the corresponding phase-only holograms. (a) Cropped-zoomed patch of the first color image in (b) for visualization. (b) Captured color images. (c) The corresponding phase-only holograms of (b).

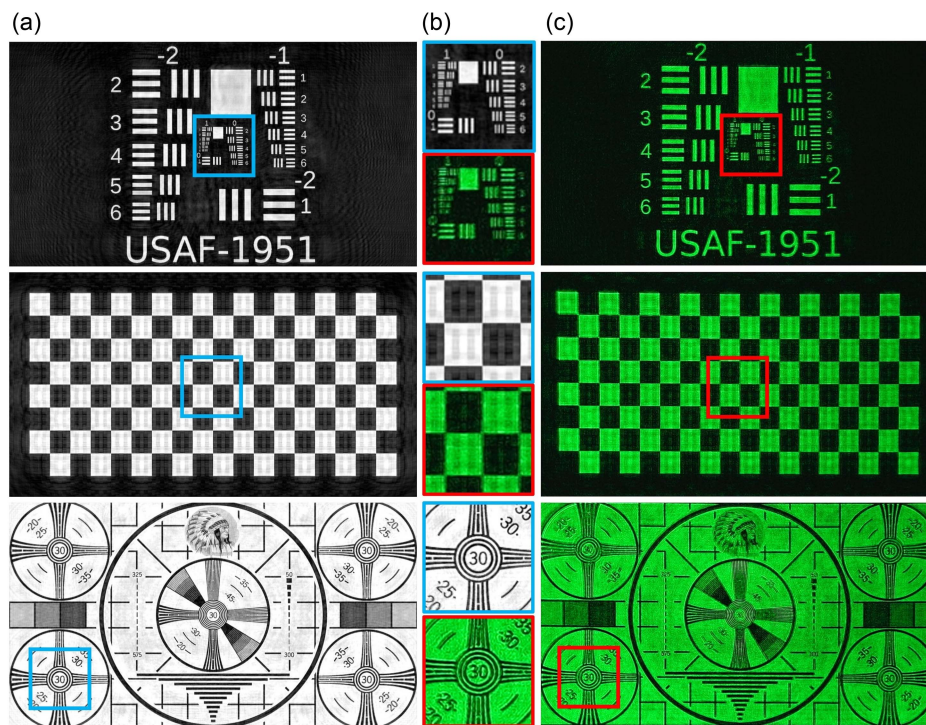


Fig. 11. Results of our method applied to binary images. (a) Simulated images. (b) Zoom-in patches. (c) Experimental images. Note that target binary images are not in our training set.

difficult for a network trained on color images to accurately reconstruct. Despite this, our method demonstrates robustness, as identifiable areas are evident in the reconstruction results. These findings have prompted us to explore further refinements to our method to enhance its generalizability across different data types.

4. CONCLUSION

In conclusion, our study represents an advancement in the field of computer-generated holography by introducing the

cross-domain fusion network (CDFN) as a solution for generating high-quality and efficient phase-only holograms. By addressing the complexities of cross-domain conversion, CDFN stands as a testament to the potential of integrating computational methods with traditional optical principles. The inclusion of the infinite phase mapper, which incorporates an understanding of optical physics, ensures that the generated holograms maintain a high level of fidelity to the original optical phenomena. This work signifies not only technological progress but also a deeper integration between the realms of

computation and physical reality, promising a future where holographic displays can achieve unprecedented levels of realism.

The proposed physics-aware CDFN encounters scalability limitations at higher resolutions and requires substantial computational resources, specifically over 24 GB of GPU memory. Additionally, the infinity phase mapper, while innovative, may not completely model all elements of holographic displays. Future efforts will focus on exploring computational models for greater efficiency [42], supporting higher resolutions [43], and refining physical assumptions to broaden the CDFN's applicability in CGH. This aims to not only improve current metrics but also expand the network's utility in broader holographic imaging and display challenges. Finally, while this work does not delve into the details of 3D CGH [44], we note that the proposed method could potentially extend to these domains but leave this question for future investigations.

Funding. National Natural Science Foundation of China (62305184); Basic and Applied Basic Research Foundation of Guangdong Province (2023A1515012932); Science, Technology and Innovation Commission of Shenzhen Municipality (WDZC20220818100259004); Research Grants Council of the Hong Kong Special Administrative Region, China (C5031-22G, CityU11300123, CityU11310522); Guangdong Provincial Department of Science and Technology (2020B1515120073); City University of Hong Kong (9610628); Research Grants Council of Hong Kong (ECS 27212822).

Disclosures. The authors declare no conflicts of interest.

Data Availability. The authors will supply the relevant data and code in response to reasonable requests.

REFERENCES

1. F. Yaraş, H. Kang, and L. Onural, "State of the art in holographic displays: a survey," *J. Disp. Technol.* **6**, 443–454 (2010).
2. J. Hong, Y. Kim, H.-J. Choi, *et al.*, "Three-dimensional display technologies of recent interest: principles, status, and issues," *Appl. Opt.* **50**, H87–H115 (2011).
3. L. Huang, X. Chen, H. Mühlenbernd, *et al.*, "Three-dimensional optical holography using a plasmonic metasurface," *Nat. Commun.* **4**, 2808 (2013).
4. J. An, K. Won, Y. Kim, *et al.*, "Slim-panel holographic video display," *Nat. Commun.* **11**, 5568 (2020).
5. S. K. Saha, D. Wang, V. H. Nguyen, *et al.*, "Scalable submicrometer additive manufacturing," *Science* **366**, 105–109 (2019).
6. D. Yang, W. Seo, H. Yu, *et al.*, "Diffraction-engineered holography: beyond the depth representation limit of holographic displays," *Nat. Commun.* **13**, 6012 (2022).
7. M. Chae, K. Bang, D. Yoo, *et al.*, "Étendue expansion in holographic near eye displays through sparse eye-box generation using lens array eyepiece," *ACM Trans. Graph.* **42**, 58 (2023).
8. L. Huang, H. Chen, T. Liu, *et al.*, "Self-supervised learning of hologram reconstruction using physics consistency," *Nat. Mach. Intell.* **5**, 895–907 (2023).
9. S. Choi, J. Kim, Y. Peng, *et al.*, "Optimizing image quality for holographic near-eye displays with Michelson holography," *Optica* **8**, 143–146 (2021).
10. Y. Peng, S. Choi, J. Kim, *et al.*, "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," *Sci. Adv.* **7**, eabg5040 (2021).
11. Z. He, X. Sui, G. Jin, *et al.*, "Progress in virtual reality and augmented reality based on holographic display," *Appl. Opt.* **58**, A74–A81 (2019).
12. Y. Zhao, L. Cao, H. Zhang, *et al.*, "Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method," *Opt. Express* **23**, 25440–25449 (2015).
13. M. Zhou, H. Zhang, S. Jiao, *et al.*, "End-to-end compression-aware computer-generated holography," *Opt. Express* **31**, 43908–43919 (2023).
14. M. Zhou, S. Jiao, P. Chakravarthula, *et al.*, "Point spread function-inspired deformable convolutional network for holographic displays," *Proc. SPIE* **13104**, 131042M (2024).
15. X. Sui, W. Wu, M. Pivnenko, *et al.*, "Polarimetric calibrated robust dual-SLM complex-amplitude computer-generated holography," *Opt. Lett.* **48**, 3625–3628 (2023).
16. G. A. Koulouris, K. Akşit, M. Stengel, *et al.*, "Near-eye display and tracking technologies for virtual and augmented reality," in *Computer Graphics Forum* (2019), Vol. **38**, pp. 493–519.
17. D. Pi, J. Liu, and Y. Wang, "Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display," *Light Sci. Appl.* **11**, 231 (2022).
18. R. W. Gerchberg, "A practical algorithm for the determination of plane from image and diffraction pictures," *Optik* **35**, 237–246 (1972).
19. S. Tian, L. Chen, and H. Zhang, "Optimized Fresnel phase hologram for ringing artifacts removal in lensless holographic projection," *Appl. Opt.* **61**, B17–B24 (2022).
20. H. Zheng, C. Zhou, X. Shui, *et al.*, "Computer-generated full-color phase-only hologram using a multiplane iterative algorithm with dynamic compensation," *Appl. Opt.* **61**, B262–B270 (2022).
21. J. Zhang, N. Pégard, J. Zhong, *et al.*, "3D computer-generated holography by non-convex optimization," *Optica* **4**, 1306–1313 (2017).
22. Z. He, X. Sui, G. Jin, *et al.*, "Optimal quantization for amplitude and phase in computer-generated holography," *Opt. Express* **29**, 119–133 (2021).
23. P. Chakravarthula, Y. Peng, J. Kollin, *et al.*, "Wirtinger holography for near-eye displays," *ACM Trans. Graph.* **38**, 213 (2019).
24. C. Chen, B. Lee, N.-N. Li, *et al.*, "Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function," *Opt. Express* **29**, 15089–15103 (2021).
25. Y. Qi, C. Chang, and J. Xia, "Speckleless holographic display by complex modulation based on double-phase method," *Opt. Express* **24**, 30368–30378 (2016).
26. X. Sui, Z. He, G. Jin, *et al.*, "Band-limited double-phase method for enhancing image sharpness in complex modulated computer-generated holograms," *Opt. Express* **29**, 2597–2612 (2021).
27. K. Liu, Z. He, and L. Cao, "Double amplitude freedom Gerchberg-Saxton algorithm for generation of phase-only hologram with speckle suppression," *Appl. Phys. Lett.* **120**, 061103 (2022).
28. K. Liu, Z. He, and L. Cao, "Pattern-adaptive error diffusion algorithm for improved phase-only hologram generation," *Chin. Opt. Lett.* **19**, 050501 (2021).
29. R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Appl. Opt.* **57**, 3859–3863 (2018).
30. Y. Peng, S. Choi, N. Padmanaban, *et al.*, "Neural holography with camera-in-the-loop training," *ACM Trans. Graph.* **39**, 185 (2020).
31. R. Horisaki, Y. Nishizaki, K. Kitaguchi, *et al.*, "Three-dimensional deeply generated holography," *Appl. Opt.* **60**, A323–A328 (2021).
32. L. Shi, B. Li, C. Kim, *et al.*, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature* **591**, 234–239 (2021).
33. G. Yuan, M. Zhou, Y. Peng, *et al.*, "Error-compensation network for ringing artifact reduction in holographic displays," *Opt. Lett.* **49**, 3210–3213 (2024).
34. Z. Dong, C. Xu, Y. Ling, *et al.*, "Fourier-inspired neural module for real-time and high-fidelity computer-generated holography," *Opt. Lett.* **48**, 759–762 (2023).
35. K. Liu, J. Wu, Z. He, *et al.*, "4K-DMDNet: diffraction model-driven network for 4K computer-generated holography," *Opto-Electron. Adv.* **6**, 220135 (2023).
36. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (2015), pp. 234–241.



37. J. Wu, K. Liu, X. Sui, *et al.*, "High-speed computer-generated holography using an autoencoder-based deep neural network," *Opt. Lett.* **46**, 2908–2911 (2021).
38. C. Zhong, X. Sang, B. Yan, *et al.*, "Real-time high-quality computer-generated hologram using complex-valued convolutional neural network," *IEEE Trans. Vis. Comput. Graph.* **30**, 3709–3718 (2023).
39. M. H. Eybposh, N. W. Caira, M. Atisa, *et al.*, "Deepcgh: 3D computer-generated holography using deep learning," *Opt. Express* **28**, 26636–26650 (2020).
40. J. Johnson, A. Alahi, and F.-F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proceedings of European Conference of Computer Vision* (2016), pp. 694–711.
41. E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: dataset and study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 126–135.
42. Z. Dong, J. Jia, Y. Li, *et al.*, "Divide-conquer-and-merge: memory- and time-efficient holographic displays," in *IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (2024), pp. 493–501.
43. S. Lee, S.-W. Nam, J. Lee, *et al.*, "Holosr: deep learning-based super-resolution for real-time high-resolution computer-generated holograms," *Opt. Express* **32**, 11107–11122 (2024).
44. S. Choi, M. Gopakumar, Y. Peng, *et al.*, "Neural 3D holography: learning accurate wave propagation models for 3D holographic virtual and augmented reality displays," *ACM Trans. Graph.* **40**, 240 (2021).