Optics Letters

3D-HoloNet: fast, unfiltered, 3D hologram generation with camera-calibrated network learning

Wenbin Zhou,† Feifan Qu,† Xiangyu Meng, Zhenyang Li, and Yifan Peng*

Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Rd, Hong Kong SAR, China
† *These authors contributed equally to this work.*
**evanpeng@hku.hk*

Received 16 October 2024; revised 7 January 2025; accepted 7 January 2025; posted 8 January 2025; published 5 February 2025

Computational holographic displays typically rely on time-consuming iterative computer-generated holographic (CGH) algorithms and bulky physical filters to attain high-quality reconstruction images. This trade-off between inference speed and image quality becomes more pronounced when aiming to realize 3D holographic imagery. This work presents *3D-HoloNet*, a deep neural network-empowered CGH algorithm for generating phase-only holograms (POHs) of 3D scenes, represented as RGB-D images, in real time. The proposed scheme incorporates a learned, camera-calibrated wave propagation model and a phase regularization prior into its optimization. This unique combination allows for accommodating practical, unfiltered holographic display setups that may be corrupted by various hardware imperfections. Results tested on an unfiltered holographic display reveal that the proposed *3D-HoloNet* can achieve 30 fps at full HD for one color channel using a consumer-level GPU while maintaining image quality comparable to iterative methods across multiple focused distances.

https://doi.org/10.1364/OL.544816

In recent years, augmented reality and virtual reality (AR and VR) have gradually become popular fields. However, existing products face issues such as visual-accommodation conflict (VAC), lack of focus cues, and bulky device form factors. Holographic imaging is one of the solutions proposed to alleviate these problems [1,2]. However, the use of bulky optics, in particular the extra physical filter to cut out partial unwanted light [3], in recent holographic near-eye display prototypes hinders their miniaturization. The challenge of achieving high image fidelity in 3D space and in real time without these systems remains a major obstacle to the widespread adoption of holographic technologies in practical, near-eye display platforms.

Recently, significant efforts have been directed toward advancing computer-generated holographic (CGH) algorithms using artificial intelligence to enhance experimental outcomes [4,5] and generate three-dimensional (3D) or multi-depth holograms [6]. A spatial light modulator (SLM) facilitates dynamic holography when illuminated by a laser. However, high diffraction orders (HDOs) and partial unwanted light are inherently present in all physical processes of optical image formation in holographic displays, especially when using algorithms such as double amplitude phase encoding (DPAC), to directly encode complex holograms into phase-only holograms (POHs) [7]. Although the filtering configuration significantly improves image quality by mitigating the HDOs, the additional optical components increase bulk.

To compactly handle the HDOs problem, a high-order gradient descent (HOGD) method was proposed to mitigate HDOs algorithmically [8]; however, this method is both time-consuming and computationally memory-consuming. To facilitate POH generation speed, the well-established HoloNet [4] introduced a convolutional neural network (CNN), and Zhong *et al.* [9] further developed a complex-valued CNN, though these approaches could only produce 2D POHs. Choi *et al.* [5] advanced the neural network-based CGH algorithm to 3D hologram generation with unprecedented image quality; however, its iterative procedure increases runtime. To the best of our knowledge, it remains a challenge to optimally compromise among algorithm runtime, the quality of 3D holographic imaging, and unfiltered system form factor simultaneously.

In this Letter, we propose the *3D-HoloNet*, a neural network-empowered CGH algorithm that efficiently synthesizes high-quality multi-depth holograms without the need for any filtering system, making it the first non-iterative method to achieve high-fidelity unfiltered 3D image reconstruction. We demonstrate that 3D-HoloNet achieves superior 3D image quality in real time, with our experiments on an unfiltered prototype showcasing excellent results in the green channel and paving the way for full-color 3D displays with high image quality and a compact form factor.

We note that the key challenge in diffraction-based hologram computation lies in computing a hologram based on the intensity distribution of a given object. In our work, we illustrate that image reconstruction and phase generation are reversible processes, reflecting the duality nature of forward and backward wave propagation in holography and can be expressed as follows:

$$\hat{a}^j_{\text{target}} = f^j_{\text{forward}}(\phi),\ \hat{\phi} = f^j_{\text{backward}}(a_{\text{target}}), \tag{1}$$

where a_{target} represents the target 3D contents of all planes, $\hat{a}^j_{\text{target}}$ is the reconstructed image, and $\hat{\phi}$ denotes the POH displayed on the phase-only SLM.

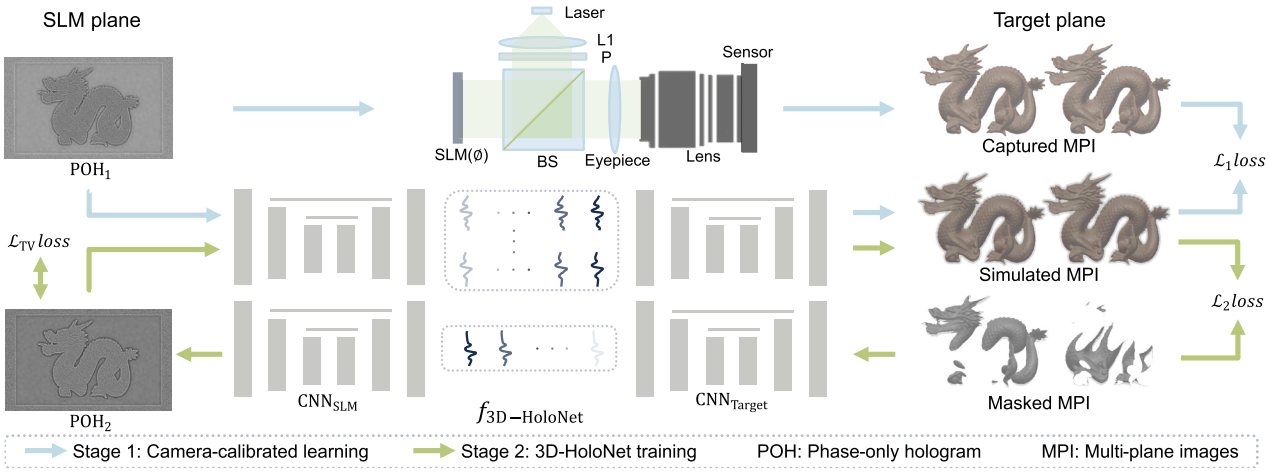

Fig. 1. Illustration of camera-calibrated learning and 3D-HoloNet training. The pipeline starts with camera-calibrated learning to establish a forward propagation model that replicates the unfiltered display hardware. This is achieved using $\mathcal{L}_1$ between captured and simulated multi-plane images. The learned forward model then serves as a foundation for training the 3D-HoloNet, enabling backpropagation operation, a capability not possible with the hardware alone. The total variance (TV) is applied on POHs to smooth it, and $\mathcal{L}_2$ is used between the masked multi-plane images and the simulated output from the forward model.

State-of-the-art neural network-based systems predominantly employ the vanilla angular spectrum method (ASM) as a forward model to supervise image reconstruction, which is popular for effectively computing free-space plane-to-plane wave propagation, mathematically represented as follows:

$$f(u,z) = \iint F(a \cdot e^{i\phi(x,y)}) e^{i2\pi(f_x x + f_y y + \sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2} z)} \mathrm{d}f_x \mathrm{d}f_y. \quad (2)$$

We observe that using the ASM as supervision can lead to a mismatch between simulation and the physical wave propagation due to the lack of HDOs and imperfect hardware. Since the activation maps in CNN-based models are derived from local convolutions, the POHs often tend to converge to checkerboard-like patterns [10], requiring a bulky optical filtering module to remove unwanted high-frequency signals, similar to the case with DPAC. It is worth noting that by setting the propagation distance to a negative value, the ASM can also be utilized to compute backward wave propagation [11]. With this insight, we design our network to be as similar as possible to the forward model.

The unified model architecture, as illustrated in Fig. 1, consists of three sequentially connected components: two U-Nets with an ideal ASM propagator positioned between them [4]. During the forward pass, the ASM propagator is composed of eight separate ASM operations, each corresponding to a distinct propagation distance, aligned with eight predefined depth planes. Initially, the first U-Net processes the POHs at the SLM plane, transforming them into a complex field. This complex field is then propagated through the ASM propagators to eight depth planes. The resulting complex fields at these target planes are finally proceeded with the second U-Net, generating the corresponding amplitude distributions at each plane.

As illustrated in Fig. 1, we prototype an unfiltered holographic display and train a forward model based on the specific hardware configuration (details are provided in Supplement 1). This model simulates the input–output relationship, effectively functioning as a differentiable, parameterized proxy of the actual hardware. In our experiments, POHs generated by various methods, including SGD, DPAC, and HOGD, are displayed on the SLM and captured by a camera controlled to focus on eight target planes.

The forward model is trained using 3000 pairs of POHs and the corresponding multi-plane captured images, utilizing the $\mathcal{L}_1$ loss function and the Adam optimizer. Additional details can be found at Supplement 1. After training, the learned wave propagation model can predict the captured multi-depth images from a given POH input. Similar to the ASM-based SGD (SGD-ASM), the well-trained model allows for high-quality image reconstruction through iterative optimization of POHs. While the combination of SGD and the forward model is computationally demanding and time-intensive, it achieves higher image quality on the calibrated holographic display. This supervisory role is critical in guiding the training of the proposed 3D-HoloNet.

3D-HoloNet closely resembles the forward model but introduces a key distinction: it uses a single-distance ASM propagator, unlike the typical eight, positioned between two U-Nets. It processes an RGB-D image as input, which is first transformed into a masked multi-plane target amplitude. The first U-Net converts the input into a complex field, and then the single-distance ASM propagator transfers it to the SLM plane, enabling far-distance propagation task that is notably difficult for CNNs [13]. After propagation, the second U-Net converts the resulting complex field into POHs. This is essentially the inverse and more efficient process of the forward model, where the POH is reconstructed by propagating backward from the input target amplitude.

Specifically, we divide the RGB-D image into eight planes based on its depth by quantizing each color pixel to the nearest plane, thereby obtaining multi-plane images. The pixel for each plane can be expressed as $a^j(x,y) \in \mathbb{R}^{M \times N}$, where $j \in \{1,\ldots,8\}$ and M and N represent the height and width of the plane, respectively:

$$a^j(x,y) = \begin{cases} a_{\text{target}}(x,y), & \text{if } j = \arg\min_k |z^k - D(x,y)|, \\ 0, & \text{otherwise}, \end{cases} \quad (3)$$

where z^k indicates the depth of the k-th target plane and $D(x,y)$ represents the depth at pixel (x,y). The pixel value is set to

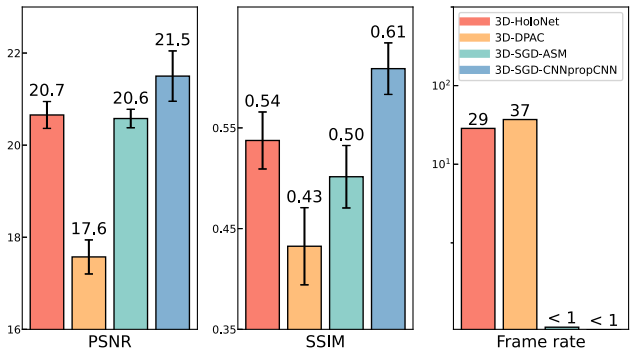

Fig. 2. Comparison of PSNR (dB) ↑, SSIM ↑, and frame rate (fps) ↑ of captured results from varying 3D CGH algorithms, including our implementation of the 3D version of DPAC method (3D-DPAC) [7,12], the SGD solver using the vanilla ASM (3D-SGD-ASM) [5], and the SGD solver using a learnable wave propagation model (3D-SGD-CNNpropCNN) [5]. Testing scenes are shown in Fig. 3 and Supplement 1.

zero on all other seven planes. In our framework, a stack of eight masks is computed for eight target planes, equally spaced in dioptric space, with distances from the camera set to 0.000, 0.084, 0.141, 0.243, 0.317, 0.416, 0.532, and 0.611 *diopters* (m^{-1}). This design choice is based on the visual clarity of the human visual system in perceiving a maximum of 0.31 diopter inter-plane spacing, as explored in the prior work [5,14].

The total loss function is thereby a combination of the pre-trained forward model $f_{forward}$, $\mathcal{L}_2$ loss, and total variation (TV) loss, which could be formulated as follows:

$$\mathcal{L} = \sum_{j=1}^{J} \left\| \hat{s} \cdot f^{j}_{\text{forward}}\left(f_{\text{3D-HoloNet}}(a_{\text{target}})\right) - a^{j}_{\text{target}} \right\|^2 + \lambda \left\| \nabla \phi \right\|_2 \quad \textbf{(4)}$$

$$\text{where } \hat{s} = \arg\min_s \| s \cdot a_{\text{recon}} - a_{\text{target}} \|_2^2.$$

Herein, $\hat{s} \in \mathbb{R}^1$ is a scaling factor for laser intensity that accounts for potential differences in value ranges between the captured and target amplitudes [5]. ϕ represents POH and λ represents the weight of TV loss, which is used to suppress the variance between neighboring pixels in the phase. Additional details for 3D-HoloNet can be found at Supplement 1.

Figure 2 presents the experimental performance of existing CGH algorithms without filtering, evaluated in peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and frame rate, averaged over a small set of images. The inference time is measured using *torch.cuda.Event* with synchronized GPU timing on an NVIDIA 4090 GPU to ensure accuracy. The well-established DPAC algorithm, although simple and fast, shows the worst reconstruction quality due to the inherent high-frequency amplitude copies that necessitate an extra physically filtering process. Moreover, the 3D-DPAC requires multiple ASM propagation operations to calculate the complex field at the SLM plane, significantly increasing the runtime compared with the 2D-DPAC. Iterative methods, such as SGD-ASM [4], offer advantages in unfiltered systems by incorporating reconstruction feedback during the optimization process. In particular, the 3D-SGD-CNNpropCNN method outperforms the vanilla 3D-SGD-ASM as is tailored to the specific hardware, although requiring more iterations to optimize POHs. The performance gap between these two methods in our implementation is smaller compared to that reported by Peng *et al.* [4]. This is likely because our setup, without any filters, retains more high-frequency components, which is challenging for the CNN to learn and therefore lower the performance of the forward model.

3D-HoloNet presents a remarkable balance between high-quality reconstruction and fast runtime, making it a significant advancement over prior algorithms that typically struggle to achieve both simultaneously. It offers the highest reconstruction quality among direct methods and even surpasses the iterative method SGD-ASM in terms of output fidelity. Regarding speed, 3D-HoloNet is much faster than the iterative methods, achieving real-time performance. This enables it to handle 29 frames per second, making it well-suited for real-time applications without compromising the quality of reconstruction.

Figure 3 showcases the multi-plane captured results with unfiltered 3D visuals. The 3D-DPAC method struggles to reconstruct coherent images in an unfiltered holographic system.

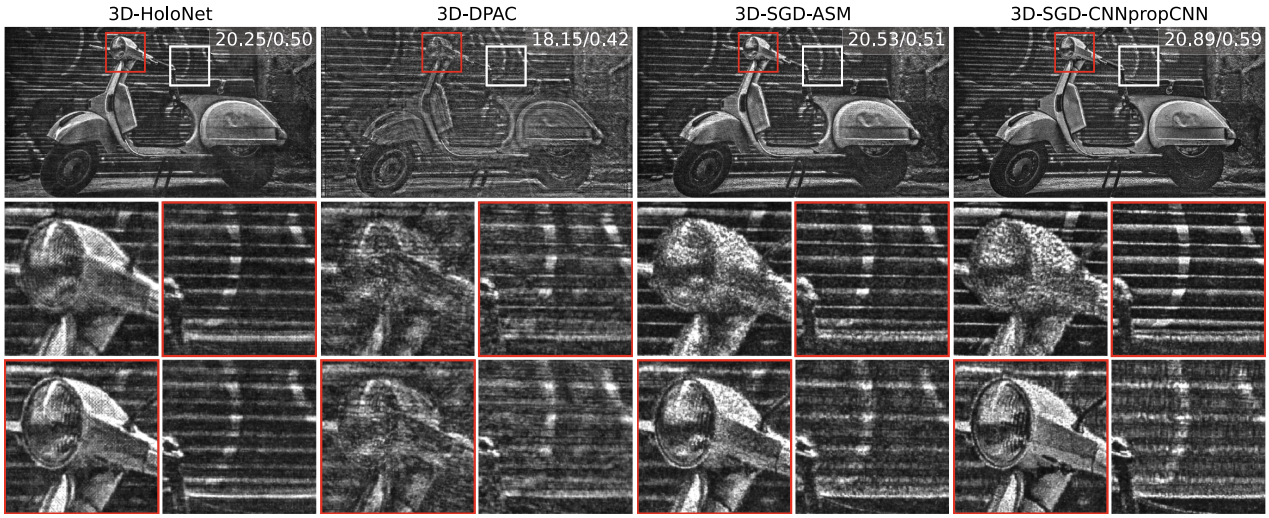

Fig. 3. Experimental results on the unfiltered holography setup with PSNR (dB)/SSIM metrics of various CGH algorithms. For fair comparison, the mean amplitude of all results is scaled to match that of the target image. We convert all displayed results to gray scale for visualization purposes. Red boxes highlight the in-focus regions. Specifically, the camera is focused at the near plane (0.55 m from the camera) for the first and the third rows, while at the far plane (1.74 m from the camera) for the second row.

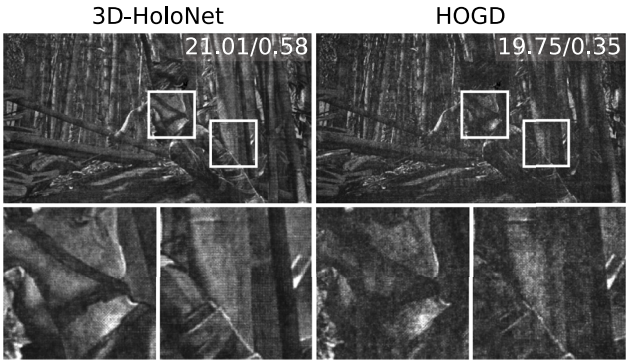

Fig. 4. Captured unfiltered holographic display results of 3D-HoloNet and HOGD. The multi-plane results of 3D-HoloNet are superimposed to a single plane for visual comparison. The ability of 3D-HoloNet to handle HDOs is proven better than HOGD both qualitatively and quantitatively. Full test images are provided in Supplement 1.

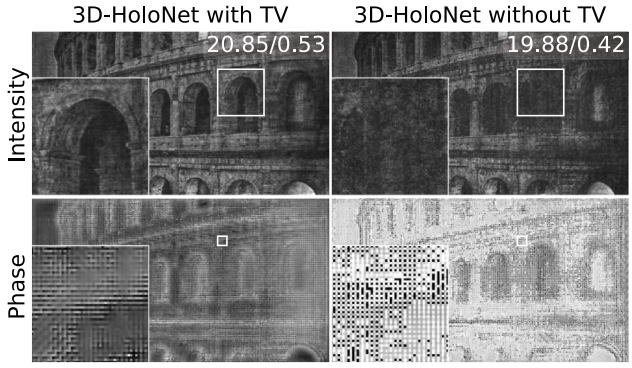

Fig. 5. Ablation on the TV loss in experiments. To facilitate comparison, the experimental results are aggregated into a single focal plane accompanied by a zoomed-in section. Architectural details are more accurately reconstructed with the TV loss. Additionally, the phase exhibits greater structure in large scale, and the high frequency of neighboring pixels is reduced, as demonstrated in the zoomed-in phase view.

Although PSNR and SSIM metrics of the two SGD-based methods are slightly higher, they fail to address the noticeable speckle artifacts without filter. These artifacts are aesthetically displeasing and can diminish visual clarity. In contrast, 3D-HoloNet leads to reconstructed images with fewer speckle artifacts. This cleaner appearance can enhance visual comfort, making 3D-HoloNet more suitable for applications that require high-quality holographic representations.

Figure 4 compares our captured results with HOGD, which is a competitive unfiltered POH optimization algorithm [8]. Ours achieves superior image quality while requiring significantly less runtime and computation resources. For the 3D-HoloNet results, we aggregate the multi-plane captured data into a single focus plane to facilitate comparison with HOGD, as HOGD is too CUDA memory-intensive to implement in 3D on a consumer-grade GPU.

Figure 5 illustrates the results of an ablation study on TV loss. Incorporating TV loss effectively reduces high-frequency noise and results in a more structured phase. It penalizes large variance between neighboring pixels in the POH, reducing the abrupt black and white pixel transitions, which is challenging for SLM to reproduce due to the electronic cross talk. The intensity images enhanced with TV loss appear visually clearer in the captured data and reconstruct finer details, demonstrating improved performance across various metrics.

In summary, the proposed 3D-HoloNet demonstrates the feasibility of using a single unified model to simultaneously represent both forward and backward wave propagation. Notably, its inverse network model can generate 3D holograms in real time with comparable reconstruction quality while operating similar speed as the 3D-DPAC method. Furthermore, by incorporating the camera-calibrated forward model and a TV loss into the optimization, 3D-HoloNet effectively handles HDOs, surpassing the baseline iterative SGD-ASM. With a reasonable amount of engineering effort in neural network compression and GPU processing optimization [15], the runtime can be further decreased. At this proof-of-concept stage, we have only experimentally verified the proposed pipeline using the green light source. Although it can be easily extended to full color by training three 3D-HoloNets with different wavelengths, future work could enhance this by jointly optimizing all three channels within a single model. Further employing time-multiplexing techniques for RGB visualization, the HDO problem, speckle artifacts, and alternative noise can be effectively mitigated.

Funding. University Grants Committee of Hong Kong (ECS 27212822, GRF 17208023); NSFC Excellent Young Scientists Fund (62322217).

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

Disclosures. The authors declare no conflicts of interest.

Data availability. All data and code needed to evaluate the conclusions will be made publicly available at Ref. [16].

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. C. Jang, K. Bang, M. Chae, *et al.*, Nat. Commun. **15**, 66 (2024).
2. M. Gopakumar, G.-Y. Lee, S. Choi, *et al.*, Nature **629**, 791 (2024).
3. G. Kuo, F. Schiffers, D. Lanman, *et al.*, ACM Trans. Graph. **42**, 203 (2023).
4. Y. Peng, S. Choi, N. Padmanaban, *et al.*, ACM Trans. Graph. **39**, 185 (2020).
5. S. Choi, M. Gopakumar, Y. Peng, *et al.*, ACM Trans. Graph. **40**, 240 (2021).
6. X. Sui, Z. He, D. Chu, *et al.*, Light: Sci. Appl. **13**, 158 (2024).
7. L. Shi, B. Li, C. Kim, *et al.*, Nature **591**, 234 (2021).
8. M. Gopakumar, J. Kim, S. Choi, *et al.*, Opt. Lett. **46**, 5822 (2021).
9. C. Zhong, X. Sang, B. Yan, *et al.*, IEEE Trans. Vis. Comput. Graph. **30**, 3709 (2023).
10. Y. Sugawara, S. Shiota, and H. Kiya, APSIPA Trans. Signal Inf. Process. **8**, e9 (2019).
11. W. Zhou, X. Meng, F. Qu, *et al.*, Dig. Tech. Pap. - Soc. Inf. Disp. Int. Symp. **55**, 817 (2024).
12. A. Maimone, A. Georgiou, and J. S. Kollin, ACM Trans. Graph. **36**, 85 (2017).
13. T. Yu, S. Zhang, W. Chen, *et al.*, Opt. Express **30**, 2378 (2022).
14. F. W. Campbell, Opt. Acta **4**, 157 (1957).
15. A. Polino, R. Pascanu, and D. Alistarh, "Model compression via distillation and quantization," arXiv (2018).
16. F. Qu and Z. W. Bean, "3D-HoloNet: fast, unfiltered, 3D hologram generation with camera-calibrated network learning," GitHub 2025 https://github.com/zhou-wb/3D-HoloNet.