Filter-Free 3D HoloNet with Hardware-Aware Calibration

Feifan Qu*† , Wenbin Zhou*†, Xiangyu Meng*, Zhenyang Li*, Hyunmin Ban* and Yifan (Evan) Peng*

***The University of Hong Kong, Hong Kong SAR, China**

†Denotes shared first-author contribution

Abstract

Computational holography faces challenges in balancing speed and quality, especially for 3D content. This work presents 3D-HoloNet, a deep learning framework that generates phase-only holograms from RGB-D scenes in real time. The method integrates a learned wave propagation model calibrated to physical displays and a phase regularization strategy, enabling robust performance under hardware imperfections. Experiments show the system achieves 30 fps at full-HD resolution (single-color channel) on consumer GPUs while matching iterative methods in reconstruction quality across multiple focal planes. By eliminating iterative optimization and physical filtering, 3D-HoloNet addresses the critical speed-quality trade-off in unfiltered holographic displays.

Author Keywords

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

1. Introduction

In recent years, augmented reality (AR) and virtual reality (VR) have emerged as rapidly growing fields. Despite their popularity, existing devices are often plagued by challenges such as visual-accommodation conflict (VAC), insufficient focus cues, and bulky form factors. Holographic imaging has been proposed as a potential solution to address these issues [1]. However, the reliance on bulky optical components—particularly the use of additional physical filters to suppress unwanted light [3] in current holographic near-eye display prototypes poses significant barriers to miniaturization. Achieving high-fidelity 3D imaging in real time without relying on such systems remains a critical hurdle for the practical adoption of holographic technologies in compact, near-eye display platforms.

In recent years, considerable progress has been made in leveraging artificial intelligence to advance computer-generated holography (CGH) algorithms, improving experimental outcomes [4,5] and enabling the generation of three-dimensional (3D) or multi-depth holograms [6]. Dynamic holography is made possible by the spatial light modulator (SLM) when illuminated by a laser. However, the formation of holographic images is inherently accompanied by high diffraction orders (HDOs) and unwanted light, which are particularly prominent in physical optical systems. These issues become more pronounced with algorithms like double-phase amplitude encoding (DPAC), which directly convert complex holograms into phase-only holograms (POHs) [7]. While filtering configurations can effectively reduce HDOs and enhance image quality, the inclusion of additional optical components inevitably adds to the system's bulk.

To address the issue of high diffraction orders (HDOs) in a more compact manner, the high-order gradient descent (HOGD) method was introduced as an algorithmic solution [8]. However, this approach is both computationally intensive and memory demanding. To accelerate the generation of phase-only holograms (POHs), the widely recognized HoloNet [4] incorporated a convolutional neural network (CNN), while Zhong et al. [9] extended this concept by proposing a complex-valued CNN. Despite these advancements, both methods are limited to producing 2D POHs. Choi et al. [5] further advanced neural network-based CGH algorithms to support 3D hologram generation, achieving unprecedented image quality. However, their iterative procedure significantly increases runtime [15]. To the best of our knowledge, achieving an optimal balance among algorithm efficiency, 3D holographic image quality, and the form factor of unfiltered systems remains an unresolved challenge.

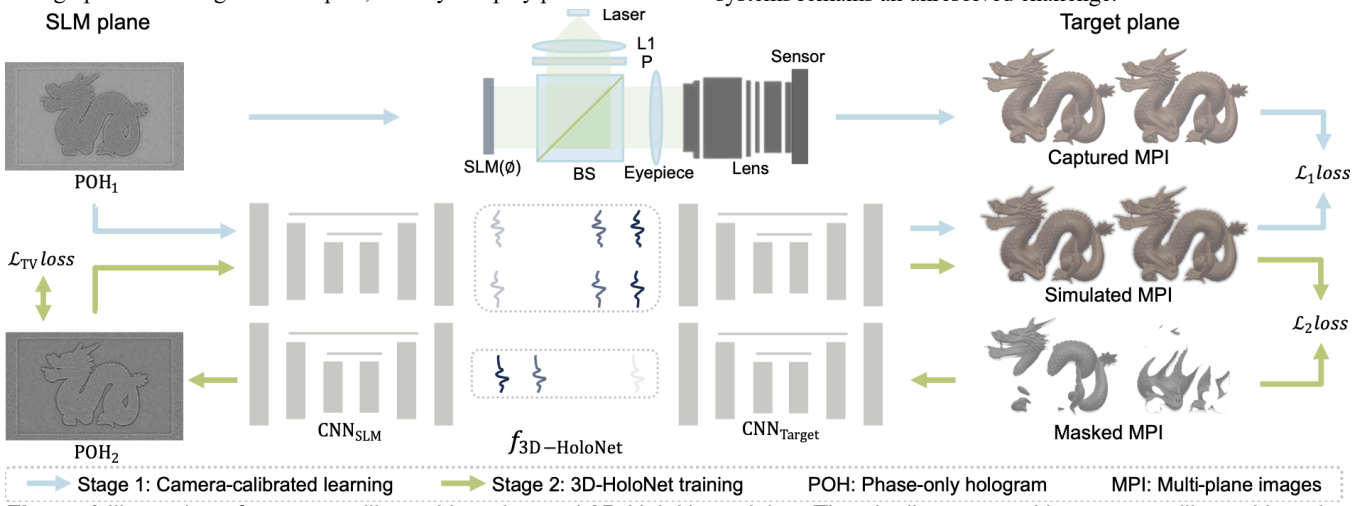

Figure 1 Illustration of camera-calibrated learning and 3D-HoloNet training. The pipeline starts with camera-calibrated learning to establish a forward propagation model that replicates the unfiltered display hardware. This is achieved using $\mathcal{L}_1$between captured and simulated multi-plane images. The learned forward model then serves as a foundation for training the 3D-HoloNet, enabling back-propagation 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.

ISSN 0097-996X/25/5602-1143-$1.00 © 2025

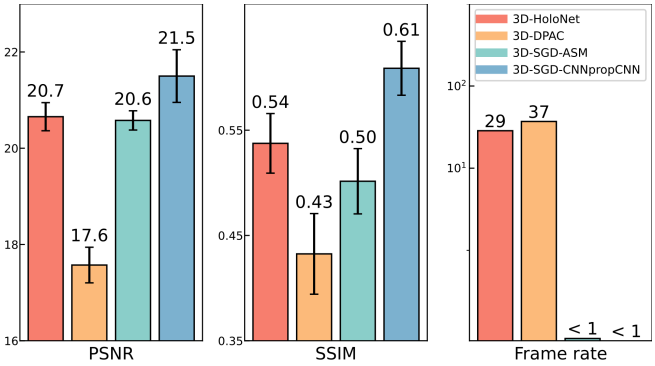

Figure 2 Comparison of PSNR (dB)↑, SSIM↑, and frame rate↑ of captured results from varying 3D CGH algorithms, including our implementation of the 3D version of DPAC method (3D-DPAC) [7, 12], the SGD solver using the vanilla ASM (3D-SGD-ASM) [5], and the SGD solver using a learnable wave propagation model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

We introduce 3D-HoloNet, a neural network-driven CGH algorithm capable of efficiently generating high-quality multi-depth holograms without requiring any filtering system, making it the first non-iterative approach to achieve high-fidelity 3D image reconstruction in unfiltered setups. Our experiments with an unfiltered prototype demonstrate that 3D-HoloNet delivers outstanding 3D image quality in real-time, particularly in the green channel, laying the foundation for full-color 3D displays with superior image quality and a compact design.

2. Methodology

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:

$$\hat{a}^{j}_{\text{target}} = f^{j}_{\text{forward}}(\phi), \hat{\phi} = f^{j}_{\text{backward}}(a_{\text{target}})$$

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

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:

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

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 worthy noting that by setting 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 Figure 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 8 separate ASM operations, each corresponding to a distinct propagation distance, aligned with 8 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 8 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 shown in Figure 1, we develop a prototype unfiltered holographic display and train a forward model tailored to the specific hardware configuration. This model accurately simulates the input-output relationship, serving as a differentiable and

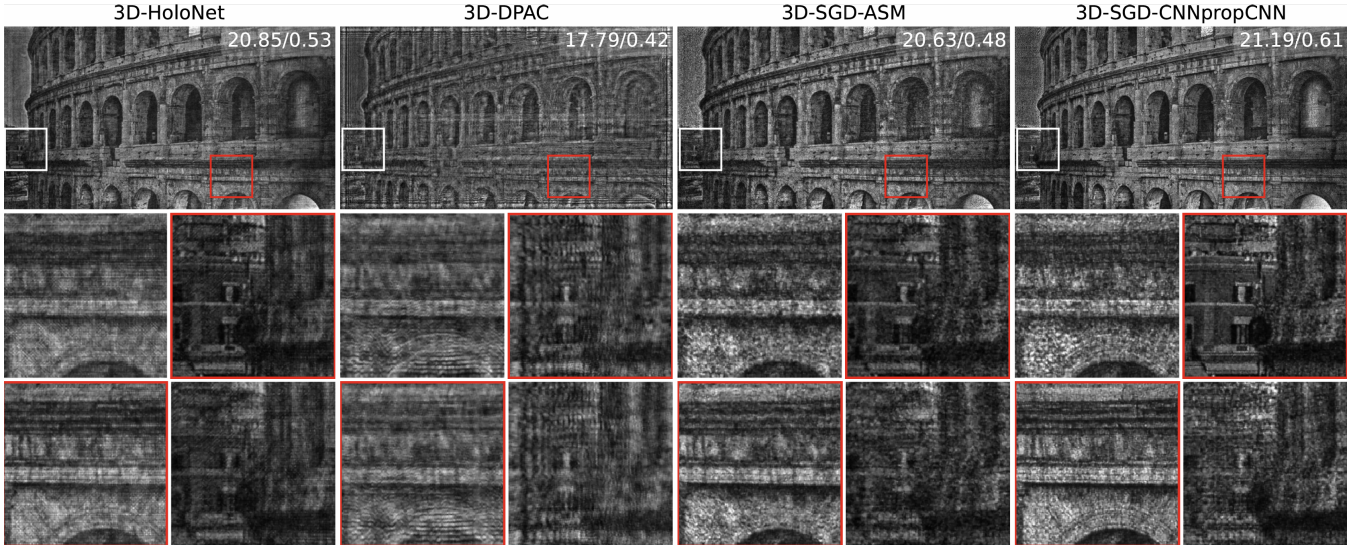

Figure 3 Experimental results on the unfiltered holography setup with PSNR (dB) / SSIM metrics of various CGH algorithms. For fair comparison, the mean amplitude of all results is scaled to match that of the target image. We convert all displayed results to grayscale for visualization purpose. Red boxes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

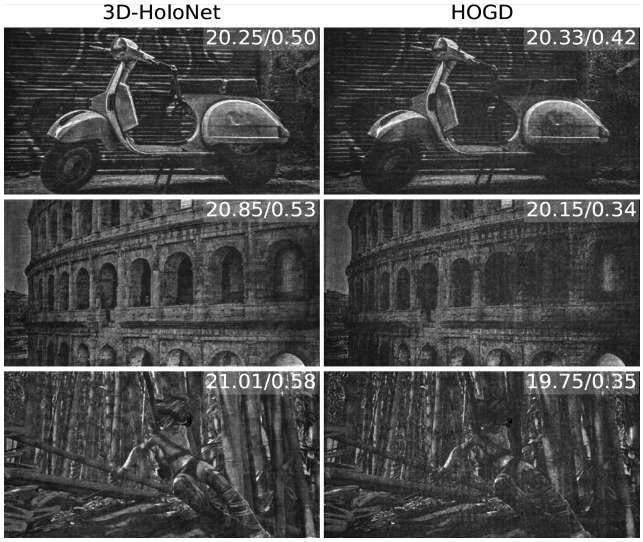

Figure 4 Captured unfiltered holographic display results of 3D-HoloNet and HOGD. The multi-plane results of 3D-HoloNet are superimposed to single plane for visual comparison. The ability of 3D-HoloNet to handling HDOs is proven better than HOGD.

parameterized proxy for the physical hardware. In our experiments, phase-only holograms (POHs) generated by different methods, including SGD, DPAC, and HOGD, are displayed on the SLM and captured by a camera set to focus on eight target planes.

The forward model is trained on 3,000 pairs of phase-only holograms (POHs) and their corresponding multi-plane captured images using the $\mathcal{L}_1$loss function and the Adam optimizer. Once trained, the model can predict captured multi-depth images from a given POH input. Similar to the ASM-based SGD (SGD-ASM), the well-trained model enables high-quality image reconstruction through iterative optimization of POHs. Although combining SGD with the forward model is computationally intensive and time-consuming, it delivers superior image quality on the calibrated holographic display. This supervisory role is essential for guiding the training process of the proposed 3D-HoloNet.

3D-HoloNet shares similarities with the forward model but introduces a significant difference: it employs a single-distance ASM propagator, as opposed to the conventional eight, strategically placed between two U-Nets. The network takes an RGB-D image as input, which is first converted into a masked multi-plane target amplitude. The first U-Net transforms this input into a complex field, which is then propagated to the SLM plane using the single-distance ASM propagator. This enables far-distance propagation, a task that CNNs typically struggle with [13]. After propagation, the second U-Net converts the resulting complex field into phase-only holograms (POHs). This approach represents a more efficient, inverse process of the forward model, reconstructing POHs by back-propagating from the target amplitude.

In our framework, a set of eight masks is computed for eight target planes, distributed evenly in dioptric space with distances of 0.000, 0.084, 0.141, 0.243, 0.317, 0.416, 0.532, and 0.611 diopters from the camera. This configuration is informed by the human visual system's ability to perceive inter-plane spacing of up to 0.31 diopters, as investigated in prior research [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:

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

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

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 amplitude [5]. The ϕ represents POH and λ represents the weight of TV loss, which is used to suppress the variance between neighboring pixels in the phase.

3. Results

Figure 2 illustrates the experimental performance of various CGH algorithms without filtering, evaluated using peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and frame rate, averaged over a small set of images. The inference time is precisely measured with torch.cuda.Event and synchronized GPU timing on an NVIDIA 4090 GPU to ensure accuracy. The well-known DPAC algorithm, while straightforward and fast, exhibits the lowest reconstruction quality due to inherent high-frequency amplitude copies that necessitate additional physical filtering. Furthermore, 3D-DPAC requires multiple ASM propagation operations to compute the complex field at the SLM plane, leading to significantly longer run-times compared to 2D-DPAC. Iterative approaches like SGD-ASM [4] perform better in unfiltered systems by utilizing reconstruction feedback during optimization. Specifically, the 3D-SGD-CNNpropCNN method surpasses the standard 3D-SGD-ASM by tailoring the process to specific hardware, though it requires more iterations to optimize POHs. The performance gap between these two methods in our implementation is smaller than that reported by Peng et al. [4]. This discrepancy is likely due to our unfiltered setup, which retains more high-frequency components that are difficult for the CNN to learn, thereby reducing the effectiveness of the forward model.

The 3D-HoloNet achieves an impressive balance between high-quality reconstruction and fast runtime, representing a significant improvement over previous algorithms that often fail to excel in both aspects simultaneously. It delivers the highest reconstruction quality among direct methods and even outperforms the iterative SGD-ASM method in output fidelity. In terms of speed, 3D-HoloNet is significantly faster than iterative approaches, achieving real-time performance by processing 29 frames per second. This makes it highly suitable for real-time applications while maintaining excellent reconstruction quality.

Figure 3 presents the multi-plane captured results of unfiltered 3D visuals. The 3D-DPAC method struggles to produce coherent reconstructions in an unfiltered holographic setup. While the PSNR and SSIM metrics of the two SGD-based methods are slightly higher, they fail to mitigate the prominent speckle artifacts in the absence of a filter. These artifacts not only detract from the aesthetic appeal but also reduce visual clarity. In contrast, 3D-HoloNet achieves reconstructed images with significantly fewer speckle artifacts, resulting in a cleaner and more visually comfortable appearance. This makes 3D-HoloNet particularly well-suited for applications requiring high-quality holographic imaging.

Figure 4 compares our captured results with HOGD, a competitive unfiltered POH optimization algorithm [8]. Our approach delivers superior image quality while significantly reducing runtime and

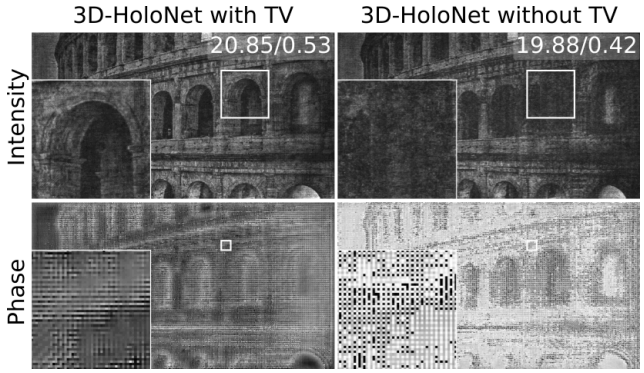

Figure 5 Ablation on the TV loss in experiments. To facilitate comparison, the experimental results are aggregated into a single focal plane accompanied by a zoomed-in section. Architectural details are more accurately reconstructed with the TV loss. Additionally, the phase exhibits greater structure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

computational resource requirements. For the 3D-HoloNet results, we aggregate the multi-plane captured data into a single focus plane to enable comparison with HOGD, as HOGD's high CUDA memory demands make 3D implementation on a consumer-grade GPU infeasible.

Figure 5 presents the results of an ablation study on the inclusion of TV loss. Incorporating TV loss effectively reduces high-frequency noise, resulting in a more structured phase. It penalizes large variations between neighboring pixels in the POH, thereby mitigating abrupt black-and-white pixel transitions that are challenging for SLMs to reproduce due to electronic crosstalk. The intensity images enhanced with TV loss appear visually sharper in the captured data and better reconstruct fine details, demonstrating improved performance across multiple evaluation metrics.

4. Discussions

In conclusion, the proposed 3D-HoloNet demonstrates the potential of utilizing a unified model to simultaneously represent forward and backward wave propagation. Its inverse network efficiently generates 3D holograms in real time, achieving reconstruction quality comparable to iterative methods while maintaining a speed like the 3D-DPAC approach. By integrating a camera-calibrated forward model and TV loss into the optimization process, 3D-HoloNet effectively addresses high diffraction orders (HDOs), outperforming the baseline iterative SGD-ASM method. With further neural network compression and GPU optimization [2], its runtime can be further reduced. This proof-of-concept has been validated using a green light source, and it can be extended to full-color imaging by training separate models for different wavelengths or jointly optimizing all channels in one model. Additionally, time-multiplexing techniques for RGB visualization could address HDOs, speckle artifacts, and other noise, enhancing its practical applicability.

5. Acknowledgements

This work was partially supported by the NSFC Excellent Young Scientist Fund (62322217) and the Research Grants Council of Hong Kong (ECS 27212822, GRF 17208023).

6. References

1. Jang C, Bang K, Chae M, et al. Waveguide holography for 3D augmented reality glasses[J]. Nature Communications, 2024, 15(1): 66.
2. Polino A, Pascanu R, Alistarh D. Model compression via distillation and quantization[J]. arXiv preprint arXiv:1802.05668, 2018.
3. Kuo G, Schiffers F, Lanman D, et al. Multisource holography[J]. ACM Transactions on Graphics (Tog), 2023, 42(6): 1-14.
4. Peng Y, Choi S, Padmanaban N, et al. Neural holography with camera-in-the-loop training[J]. ACM Transactions on Graphics (TOG), 2020, 39(6): 1-14.
5. Choi S, Gopakumar M, Peng Y, et al. Neural 3D holography: learning accurate wave propagation models for 3D holographic virtual and augmented reality displays[J]. ACM Transactions on Graphics (TOG), 2021, 40(6): 1-12.
6. Sui X, He Z, Chu D, et al. Non-convex optimization for inverse problem solving in computer-generated holography[J]. Light: Science & Applications, 2024, 13(1): 158.
7. Shi L, Li B, Kim C, et al. Towards real-time photorealistic 3D holography with deep neural networks[J]. Nature, 2021, 591(7849): 234-239.
8. Gopakumar M, Kim J, Choi S, et al. Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays[J]. Optics letters, 2021, 46(23): 5822-5825.
9. Zhong C, Sang X, Yan B, et al. Real-time high-quality computer-generated hologram using complex-valued convolutional neural network[J]. IEEE Transactions on Visualization and Computer Graphics, 2023.
10. Sugawara Y, Shiota S, Kiya H. Checkerboard artifacts free convolutional neural networks[J]. APSIPA Transactions on Signal and Information Processing, 2019, 8: e9.
11. Zhou W, Meng X, Qu F, et al. 59-4: Towards Real-time 3D Computer-Generated Holography with Inverse Neural Network for Near-eye Displays[C]//SID Symposium Digest of Technical Papers. 2024, 55(1): 817-820.
12. Maimone A, Georgiou A, Kollin J S. Holographic near-eye displays for virtual and augmented reality[J]. ACM Transactions on Graphics (Tog), 2017, 36(4): 1-16.
13. Yu T, Zhang S, Chen W, et al. Phase dual-resolution networks for a computer-generated hologram[J]. Optics Express, 2022, 30(2): 2378-2389.
14. Campbell F W. The depth of field of the human eye[J]. Optica Acta: International Journal of Optics, 1957, 4(4): 157-164.
15. Zhou W, Qu F, Meng X, et al. 3D-HoloNet: fast, unfiltered, 3D hologram generation with camera-calibrated network learning[J]. Optics Letters, 2025, 50(4): 1188-1191.