Focal Surface Holographic Light Transport using Learned Spatially Adaptive Convolutions

Chuanjun Zheng
University College London
United Kingdom

Yicheng Zhan
University College London
United Kingdom

Liang Shi
Massachusetts Institute of Technology
USA

Ozan Cakmakci
Google
USA

Kaan Akşit*
University College London
United Kingdom

ABSTRACT

Computer-Generated Holography (CGH) is a set of algorithmic methods for identifying holograms that reconstruct Three-Dimensional scenes in holographic displays. CGH algorithms decompose 3D scenes into multiplanes at different depth levels and rely on simulations of light that propagated from a source plane to a targeted plane. Thus, for n planes, CGH typically optimizes holograms using n plane-to-plane light transport simulations, leading to major time and computational demands. Our work replaces multiple planes with a focal surface and introduces a learned light transport model that could propagate a light field from a source plane to the focal surface in a single inference. Our model leverages spatially adaptive convolution to achieve depth-varying propagation demanded by targeted focal surfaces. The proposed model reduces the hologram optimization process up to 1.5x, which contributes to hologram dataset generation and the training of future learned CGH models.

CCS CONCEPTS

• **Hardware → Emerging optical and photonic technologies**; • **Human-centered computing → Displays and imagers**; • **Computing methodologies → Computer graphics**.

KEYWORDS

Computer-Generated Holography, Light Transport, Optimization, Spatially Adaptive Convolutions, Convolutional Neural Networks

ACM Reference Format:
Chuanjun Zheng, Yicheng Zhan, Liang Shi, Ozan Cakmakci, and Kaan Akşit. 2024. Focal Surface Holographic Light Transport using Learned Spatially Adaptive Convolutions. In *SIGGRAPH Asia 2024 Technical Communications (SA Technical Communications '24), December 3–6, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3681758.3697989

1 INTRODUCTION

Computer-Generated Holography (CGH) is a family of algorithmic methods used to generate holographic interference patterns. Identifying these interference patterns using learned [Shi et al. 2022] and optimization [Kavaklı et al. 2023a] CGH methods require conventional simulations of light propagation from plane-to-plane [Matsushima and Shimobaba 2009; Shen and Wang 2006]. Recently, learned proxy methods [Choi et al. 2021; Kavaklı et al. 2022] have been proposed to replace conventional light propagation methods [Matsushima and Shimobaba 2009; Shen and Wang 2006]. As these learned proxy methods for light propagation are trained using camera-in-the-loop strategies, they are able to capture imperfections of optical hardware, closing the gap between theoretical simulations and actual hardware. Either learned or conventional, simulating light propagation among multiple planes in a 3D volume is computationally demanding, as a 3D volume is represented with multiple planes and each plane requires a separate calculation of light propagation to reconstruct the target image.

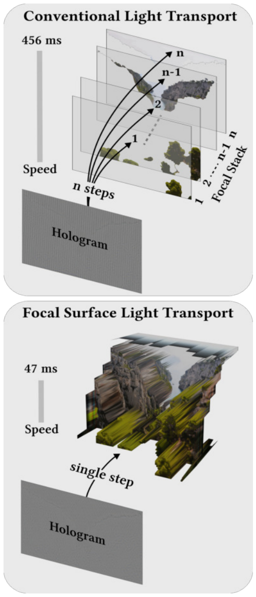

Figure 1: Conventional Light Transport VS. Proposed Focal Surface Light Transport.(Source image: Tobi 87, Link: Wikimedia Commons)

Our work introduces a learned focal surface light propagation model that could help free light simulations from plane dependence. Specifically, our model can propagate a phase-only hologram represented with a plane to a targeted focal surface, see Fig. 1. In our model, we extract Spatially Varying (SV) depth features of a focal surface by learning a set of SV kernels. In addition, our model combines these SV learned kernels with Spatially Invariant (SI) kernels using a Spatially Adaptive Convolution (SAC). Thus, effectively capturing SV and SI features of light propagation over a focal surface. Our work makes the following contributions:

- **Learned focal surface light transport model.** By uniquely leveraging SAC for CGH, we introduce a new learned light transport model. Our model identifies a mapping from a phase-only hologram represented over a plane to a targeted focal surface.
- **Focal surface-based hologram optimization.** To evaluate its practicality, we utilize our model for a 3D phase-only hologram optimization application. Compared with conventional light propagation based hologram optimization methods [Kavaklı et al. 2023a,b], our approach accelerates the optimization process up to 1.5x, leading to speed up benefits in hologram dataset generation and training future learned CGH models.
- **Experimental Validation.** We evaluate our method in simulation for various propagation distances and validate the result using a bench-top on-axis holographic display prototype.

*denotes corresponding author

SA Technical Communications '24, December 3–6, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1140-4/24/12
https://doi.org/10.1145/3681758.3697989

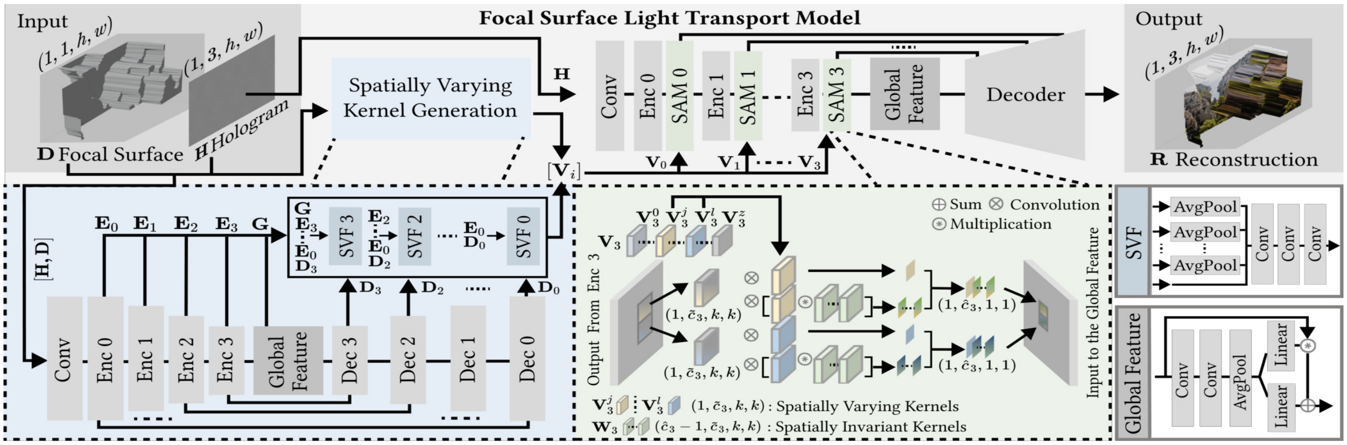

Figure 2: Our proposed learned focal surface light transport model. The process starts with an input hologram H and a focal surface D to generate spatially varying kernels $[\mathrm{V_i}]$, where $i = 0, 1, 2, 3$ indicates the index of scales. Those kernels are utilized in the Spatially Adaptive Module (SAM) to achieve focal surface light transport. In the SAM, $\mathrm{V}_3^0, \mathrm{V}_3^j, \mathrm{V}_3^l, \mathrm{V}_3^z$ represent kernels used at different spatial locations, where 0, j, l, and z indicate specific positions. (Source image: Tobi 87, Link: Wikimedia Commons)

2 FOCAL SURFACE LIGHT TRANSPORT

We introduce the SAC, a modified convolution structure for encoding SV features. Leveraging the SAC, our work enables the learned focal surface light transport network.

2.1 Spatially Adaptive Convolution

Standard Convolution. Given an input feature $\tilde{\mathbf{I}} \in \mathbb{R}^{\tilde{c}\times\tilde{h}\times\tilde{w}}$ in a Convolutional Neural Network (CNN), where $\tilde{c}$, $\tilde{h}$, and $\tilde{w}$ represent the number of channels, height, and width of the input $\tilde{\mathbf{I}}$ (in our case, $\tilde{c} = 3$, $\tilde{w} = 1080$, $\tilde{h} = 1920$), the discrete convolution based on a SI kernel $\mathbf{W} \in \mathbb{R}^{\hat{c}\times\tilde{c}\times k\times k}$ is defined as:

$$\underbrace{\mathbf{I}[c, x, y]}_{\text{output}} = \sum_{c',x',y'} \underbrace{\mathbf{W}[c, c', x', y']}_{\text{SI Kernel}} \underbrace{\tilde{\mathbf{I}}[c', x + x', y + y']}_{\text{input}}, \tag{1}$$

where $\tilde{c}$ and $\hat{c}$ indicate the number of input and output channels. The indices satisfy $1 \leq c' \leq \tilde{c}$ and $1 \leq c \leq \hat{c}$. The pair (x', y') belongs to the set $\Omega(k)$, which specifies a $k \times k$ convolutional window. The summation operation acts on all input channels, which implies that each input channel contributes to every output channel. According to Eq. (1), this operation is characterized by a kernel that is spatially shared and content-independent. Learning-based light transport models could use Eq. (1) as a basic operation. However, it is challenging for this method to project a hologram onto a focal surface. As each pixel on the hologram plane may correspond to a different depth on the focal surface, which makes the SI kernel a sub-optimal choice to capture SV features [Xu et al. 2020; Zheng et al. 2021], including focusing or out-of-focus effects due to depth variance. A typical solution is to employ a large number of parameters for feature encoding, resulting in an increased memory footprint. Alternatively, we could consider using SV convolution [Xu et al. 2020; Zheng et al. 2021]. The SV kernel $\mathbf{V} \in \mathbb{R}^{\hat{c}\times h\times w\times\tilde{c}\times k\times k}$ incorporates two new dimensions h, w into SI kernel, where h and w indicate height, and width of the output feature. However, relying solely on SV kernels may increase model capacity due to extra parameters, particularly when h and w are large. These alternative designs all demand extra network capacity.

Spatially Adaptive Convolution Operation. To address these problems, we utilize the SAC based on [Xu et al. 2020]. Our method reduces the network parameters by multiplying the SV kernel with the standard SI kernel. Initially, the SV kernel $\mathbf{V} \in \mathbb{R}^{1\times h\times w\times\tilde{c}\times k\times k}$ is introduced, the output channel is set to 1 to reduce the number of parameters. The Spatially Adaptive (SA) kernel $\mathbf{A} \in \mathbb{R}^{\hat{c}\times h\times w\times\tilde{c}\times k\times k}$ is computed by multiplying the $\mathbf{W}$ and $\mathbf{V}$, which defined as:

$$\mathbf{A}[c, x, y, c', x', y'] = \mathbf{V}[1, x, y, c', x', y'] * \mathbf{W}[c, c', x', y'], \tag{2}$$

where $1 \leq c \leq \hat{c}$, $1 \leq c' \leq \tilde{c}$,$1 \leq x \leq h$ and $1 \leq y \leq w$. Eq. 2 enhances the output channel capacity in $\mathbf{V}$ while maintaining spatially variant. Both $\mathbf{V}$ and $\mathbf{W}$ can be either pre-defined or learned, making the network content-adaptive. By using $\mathbf{A}$, the SAC is defined as:

$$\mathbf{I}[c, x, y] = \sum_{c',x',y'} \underbrace{\mathbf{A}[c, x, y, c', x', y']}_{\text{SA Kernel}} \tilde{\mathbf{I}}[c', x + x', y + y']. \tag{3}$$

SAC retains both the dimensional coherence of the SI kernel in CNN and is spatially variant at the same time. Note that when $\mathbf{W}$ becomes an all-one tensor, Eq. 3 is equivalent to the SV convolution in CNN.

2.2 Learned Focal Surface Light Transport

We first generate SV kernels to encode depth-varying features of the focal surface, which are later used in SAC for focal surface light transport. For the schematic figure of our system, please see Fig. 2.

Spatially Varying Kernel Generation. As shown in Fig. 2, the SV kernel generation module takes the hologram $\mathbf{H} \in \mathbb{R}^{1\times 3\times h\times w}$ and focal surface $\mathbf{D} \in \mathbb{R}^{1\times 1\times h\times w}$ as inputs. We adopted the architecture in RSGUNet [Huang et al. 2018] for SV kernel generation module. The output of each decoder layer is integrated with feature maps from different layers in the encoders. Then combined features will be fed into Spatially Varying Feature (SVF) module to learn a set of SV kernels $[\mathbf{V}_i]$, where $\mathbf{V}_i \in \mathbb{R}^{n\times\tilde{c}_i\times k\times k}$, $i = 0, 1, 2, 3$ refers to different scale levels, $\tilde{c}_i$ denotes the input channel, k is the kernel size, and $n = \frac{h}{2^i} \times \frac{w}{2^i}$ is the number of kernels. The SVF module contains convolution layers and average pooling layers. To mitigate artifacts, we modify the global feature module in [Huang et al. 2018] to an attention block and apply it at the bottleneck of the U-Net.

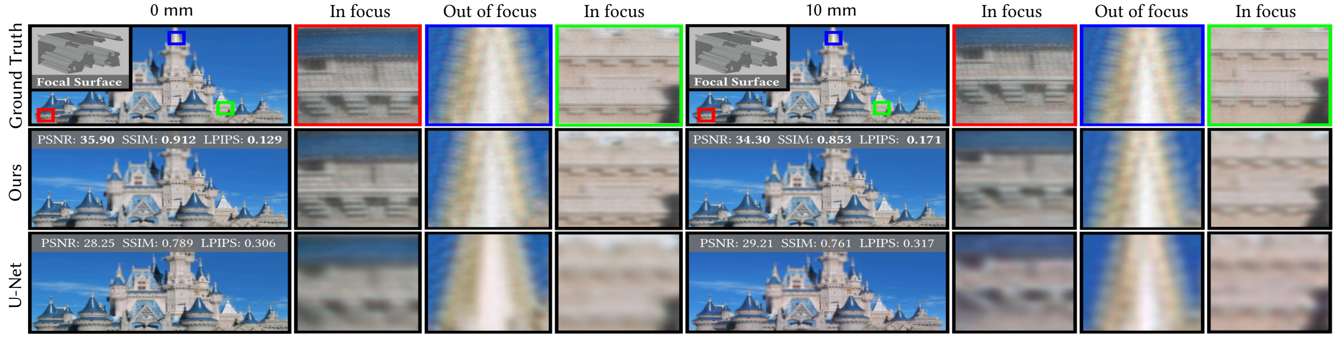
0 mm
In focus
Out of focus
In focus
10 mm
In focus
Out of focus
In focus
Ground Truth
Ours
U-Net
Focal Surface
Focal Surface
PSNR: 35.90 SSIM: 0.912 LPIPS: 0.129
PSNR: 34.30 SSIM: 0.853 LPIPS: 0.171
PSNR: 28.25 SSIM: 0.789 LPIPS: 0.306
PSNR: 29.21 SSIM: 0.761 LPIPS: 0.317

Figure 3: Visual comparison of simulating light transported onto a focal surface (specified in the first row of each case) at 0 mm and 10 mm propagation distances. The ground truth is obtained via ASM [Matsushima and Shimobaba 2009]. Both focused and defocused regions indicate poor performance of the U-Net model. (Source image: Matt H. Wade, Link: Wikimedia Commons)

Focal Surface Light Transport. We leverage the generated SV kernels to build our light transport module based on RSGUNet [Huang et al. 2018]. The module takes the hologram $\mathbf{H}$ as input without requiring depth, as the depth feature of the focal surface is inherently encoded within the learned SV kernels. To integrate the SV features into the encoder, we propose a Spatially Adaptive Module (SAM) based on SAC. As shown in Fig 2, we first replace the SI kernel $\mathbf{W}$ to an all-ones tensor in Eq. (2), which ignores the SI kernels and only considers the SV kernels to capture the original SV information. In parallel, we introduce the SI kernels back to Eq. (2) as a learning parameter and multiply with the SV kernels for better diverse feature extraction. These features from the two operations will be concatenated to form the output of SAM. Finally, the global feature module and the decoder will process the output to generate the reconstruction at the given focal surface denoted as $\mathbf{R}$.

Loss function. We employ the L_2 norm to quantify the discrepancy between the reconstruction $\mathbf{R}$ and the target image $\mathbf{R}'$. Both $\mathbf{R}$ and $\mathbf{R}'$ are focal surface depended image reconstructions. Since $\mathbf{R}$ contains both focus and defocus regions [Kavaklı et al. 2023a], we utilize a binary mask $\mathbf{M}$ that highlights only the focus parts of the image. The loss function for the reconstruction on a single focal surface $\mathcal{L}_D$ is defined as:

$$\mathcal{L}_D = \alpha_0 \mathbf{M} \|\mathbf{R} - \mathbf{R}'\|_2^2 + \alpha_1 (1 - \mathbf{M}) \|\mathbf{R} - \mathbf{R}'\|_2^2, \tag{4}$$

where α_0 and α_1 represent weights ($\alpha_0 = 1$ and $\alpha_1 = 0.5$).

2.3 Optimizing Holograms with Focal Surfaces

Recently, learning-based methods have been proposed to solve 3D hologram generation tasks [Choi et al. 2021; Shi et al. 2022]. However, the ideal 3D hologram for the holographic display has not yet been precisely defined [Kim et al. 2024]. Optimization-based hologram generation methods [Kavaklı et al. 2023a,b] could potentially help identify the ideal 3D hologram and generate hologram datasets for learning-based approaches. Typically, optimization methods are based on the multiplane representation, where a full-color hologram is synthesized by making use of the phase patterns of the three color primaries. Following previous work [Kavaklı et al. 2023b], each single-color phase pattern is obtained by:

$$\hat{\mathbf{H}}_p \leftarrow \arg\min_{\mathbf{H_p}} \sum_{p=1}^{3} \mathcal{L}\left(\left|e^{i\mathbf{H}_p} \otimes \mathbf{K}_p\right|^2, s\mathbf{R}_p\right), \tag{1}$$

where p denotes the index of a color primary, $\mathbf{H}_p$ is the SLM phase, $\hat{\mathbf{H}}_p$ is the optimized SLM phase, $\mathbf{K}_p$ is the wavelength-dependent light transport kernel [Matsushima and Shimobaba 2009], $\mathbf{R}_p$ is the target image intensity, s is an intensity scaling factor ($s = 1$ by default), $\otimes$ denotes convolution. We substitute the conventional light transport model with our focal-surface-based model:

$$\hat{\mathbf{H}} \leftarrow \arg\min_{\mathbf{H}} \mathcal{L}\left(F(\mathbf{H}, \mathbf{D}), s\mathbf{R}\right). \tag{5}$$

In this case, the hologram optimization problem is simplified. Our approach simultaneously optimizes hologram in three color primaries and maintains phase-only at the same time.

3 EVALUATION AND DISCUSSION

We generate the focal surface light transport dataset based on previous work [Kavaklı et al. 2023a,b] at the resolution 1920 × 1080. See Section 1 of the supplementary material for more details. We use Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999, \alpha_{decay} = 0.5\ after\ 50\ epochs$). The model is trained for 500 epochs, with an initial Learning Rate (LR) of 2×10^{-4}. All experiments are conducted on a single NVIDIA V100 16G GPU.

Evaluation. To assess the image quality, we utilize metrics including Peak Signal-to-noise Ratio (PSNR), Structural Similarity (SSIM), and Perceptual Similarity Metric (LPIPS) [Zhang et al. 2018]. First, we assess the quality of light simulation on a focal surface. As shown in Tbl. 1, our model outperforms U-Net [Ronneberger et al. 2015] across all metrics. Fig. 3 shows that our model preserves more high-frequency content than U-Net, providing finer details and sharper edges, closer to the ground truth. Second, we utilize our model for a 3D phase-only hologram optimization application under $0mm$ propagation distance. Optimizing holograms with six target planes using Angular Spectrum Method (ASM) [Matsushima and Shimobaba 2009] is denoted as ASM 6, while Ours 4 and Ours 6 represent optimizing holograms using our model with four and six focal surfaces, respectively. All holograms are reconstructed

Table 1: Evaluation of various light transport models on our dataset. The speed is tested by simulating an all-in-focus, full-color 3D image with six depth planes. Note that higher PSNR/SSIM and lower Params/Speed indicate better performance, denoted by ↑ and ↓ in the tables.

Methods	PSNR (dB) ↑ 0 mm/10 mm	SSIM ↑ 0 mm/10 mm	Stage	Params ↓ (M)	Speed ↓ (s)
ASM (GT) [Matsushima and Shimobaba 2009]	-	-	Two	-	0.4559
U-Net [Ronneberger et al. 2015]	29.662/30.112	0.8015/0.7760	Single	7.7760	0.0565
Ours	**36.016/34.279**	**0.9128/0.8470**	Single	**7.4446**	**0.0471**

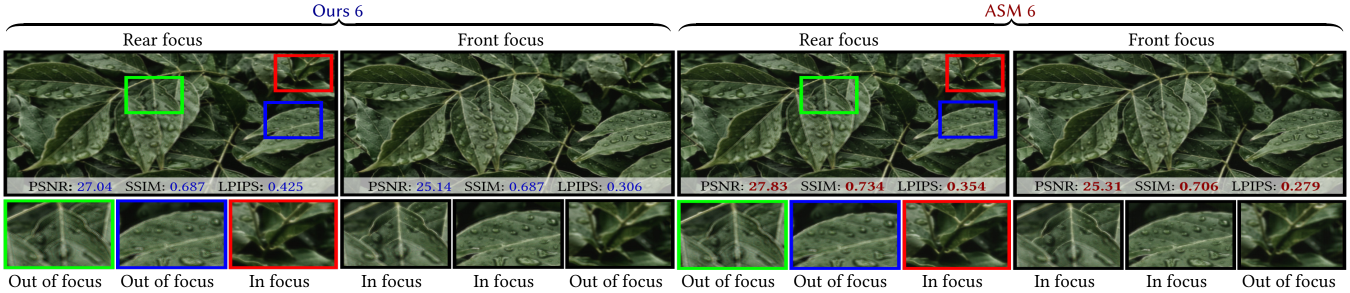

Figure 4: Visual comparison on simulated holograms optimized using ASM 6 and Ours 6 under 0 mm propagation distance. All holograms are reconstructed using ASM for evaluation. (Source image : Jaimie Phillips, Link: Wikimedia Commons)

Table 2: Comparison of image quality for the scene in Fig. 4 among ASM 6, Ours 6, and Ours 4 across different iterations at 0 mm propagation distance. Note that higher PSNR/SSIM ↑ and lower LPIPS/Speed ↓ indicate better performance.

ASM 6/Ours 6 / Ours 4	Iteration		
	50	100	200
Speed (s) ↓	42.580/30.182/**20.869**	84.626/61.460/**39.792**	171.49/119.02/**77.878**
PSNR (dB) ↑	27.377/**27.501**/26.088	**27.795**/27.598/26.905	**27.801**/27.625/26.928
SSIM ↑	**0.7100**/0.6868/0.6142	**0.7193**/0.6933/0.6753	**0.7195**/0.6890/0.6767
LPIPS ↓	**0.3971**/0.4747/0.5431	**0.3894**/0.4687/0.4707	**0.3889**/0.4787/0.4689

using ASM for performance assessment. As shown in Fig. 4 and Tbl. 2, Ours 6 achieves comparable results with about 70% of the optimization time compared to ASM 6. Actual captures of Ours 6 and ASM 6 in Fig. 5 demonstrate the capability of our model for generating 3D holograms. For more details on the display prototype and comparisons, see Sections 2 and 3 in supplementary material.

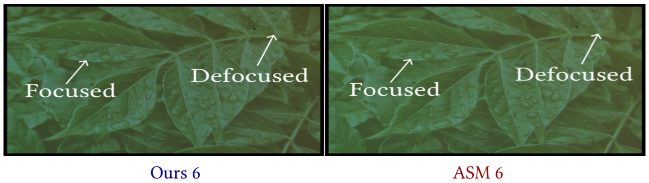

Figure 5: Comparing experimental captures of ASM 6 and Ours 6 under 0 mm propagation distances. (Source image : Jaimie Phillips, Link: Wikimedia Commons)

Computational Complexity Analysis. First, we assess the computational complexity of simulating a full-color, all-in-focus 3D image across six depth planes. As shown in Tbl. 1, conventional ASM-based model [Matsushima and Shimobaba 2009] requires eighteen forward passes to simulate a full-color, all-in-focus 3D image with six depth planes. In contrast, our model simulates the three color-primary images simultaneously onto a focal surface with a single forward pass, reducing simulation time by 10x and achieving better image quality with fewer parameters compared to U-Net [Ronneberger et al. 2015]. Second, we evaluate hologram optimization. In Tbl. 2, using four focal surfaces (Ours 4) to approximate six planes for focus and defocus guidance, speeding up optimization by up to 2x. Increasing the number of focal surfaces to six (Ours 6) achieves comparable results with about a 1.5x speedup.

Limitations and Future Works. As shown in Fig. 3, the performance of our model degrades at a long propagation distance (10 *mm*) compared to zero distance (0 *mm*). See Section 3 in the supplementary material for more comparisons. Future improvements could include using a factorized larger kernel for long-distance propagation. In addition, our model focuses on depth-varying propagation within a 3D volume, more investigation is needed for depth-varying propagation of the entire volume using conditional networks.

ACKNOWLEDGMENTS

The authors thank reviewers for their valuable feedback; Louise L. Xie for her feedback on the manuscript; Ziyang Chen and Doğa Yılmaz for their discussions and assistance. We also thank Northeastern University for computing resources for early experiments.

REFERENCES

Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, and Gordon Wetzstein. 2021. Neural 3d holography: Learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–12.

Jie Huang, Pengfei Zhu, Mingrui Geng, Jiewen Ran, Xingguang Zhou, Chen Xing, Pengfei Wan, and Xiangyang Ji. 2018. Range scaling global u-net for perceptual image enhancement on mobile devices. In *Proceedings of the European conference on computer vision (ECCV) workshops.* 0–0.

Koray Kavaklı, Yuta Itoh, Hakan Urey, and Kaan Akşit. 2023a. Realistic defocus blur for multiplane computer-generated holography. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR).* IEEE, 418–426.

Koray Kavaklı, Liang Shi, Hakan Urey, Wojciech Matusik, and Kaan Akşit. 2023b. Multi-color Holograms Improve Brightness in Holographic Displays. In *SIGGRAPH Asia 2023 Conference Papers.* 1–11.

Koray Kavaklı, Hakan Urey, and Kaan Akşit. 2022. Learned holographic light transport. *Applied Optics* 61, 5 (2022), B50–B55.

Dongyeon Kim, Seung-Woo Nam, Suyeon Choi, Jong-Mo Seo, Gordon Wetzstein, and Yoonchan Jeong. 2024. Holographic Parallax Improves 3D Perceptual Realism. *arXiv preprint arXiv:2404.11810* (2024).

Kyoji Matsushima and Tomoyoshi Shimobaba. 2009. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Optics express* 17, 22 (2009), 19662–19673.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015.* Springer, 234–241.

Fabin Shen and Anbo Wang. 2006. Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula. *Applied optics* 45, 6 (2006), 1102–1110.

Liang Shi, Beichen Li, and Wojciech Matusik. 2022. End-to-end learning of 3d phase-only holograms for holographic display. *Light: Science & Applications* 11, 1 (2022), 247.

Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. 2020. Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation. In *European Conference on Computer Vision.* Springer, 1–19.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition.*

Chuanjun Zheng, Daming Shi, and Yukun Liu. 2021. Windowing decomposition convolutional neural network for image enhancement. In *Proceedings of the 29th ACM International Conference on Multimedia.* 424–432.