

Learned Propagation Model with Complex Convolutions for Holographic Systems

Manu Gopakumar
Stanford University
manugopa@stanford.edu

Bryan Chiang
Stanford University
bhchiang@stanford.edu

Abstract

By directly modulating an emitted wavefront of light, holographic displays offer a unique degree of control over a viewer’s experience. However, physical prototypes have struggled to realize the impressive theoretical potential of such a display. This disparity in performance results from a mismatch between the phenomena modelled by theoretical simulations and the phenomena present in physical systems. Conventional theoretical analysis makes assumptions about the illumination of the display, the wavefront modulation by the display, and the propagation of light from the display which often fail to hold in physical prototypes. In this project we build upon recent work which has sought to learn more accurate models of propagation through the physical prototypes. We aim to improve these methods by utilizing the advancements of machine learning approaches while operating in the complex domain in which the physical phenomena occurs. Using the native support for complex values in the JAX ecosystem, we analyze the potential of learned models of this form and compare the performance of various complex activation functions.

1. Introduction

In recent years, advancements in virtual reality displays have greatly reduced visual discrepancies between the actual world and displayed scenes. Higher spatial resolution has allowed displays to more accurately depict fine details, and higher frame rates has produced smoother viewing experiences. However, recreating the visual focus cues found in the real world has remained challenging since this requires both spatial and angular control of the light emitted by a display.

Prior work has used different trade-offs to achieve the desired angular resolution. Various light field display implementations have tackled this issue by trading away spatial resolution for angular resolution [8]. Recent multifocal display prototypes have created visual focus cues by presenting

objects at different depths in subsequent frames presenting a trade-off between frames per scene and depth resolution [5]. Varifocal display prototypes have presented scenes at a single depth which is adjusted based on which part of the scene the viewer is looking at [1]. This avoids any of the trade-offs of the other approaches but requires very fast and accurate eye-tracking.

These limitations motivate yet another approach for reproducing visual focus cues, holography. By directly modulating the wavefront of light which contains both spatial and angular information, holographic displays aim to sidestep the difficulties faced by the discussed prior approaches. While the theoretical potential for holographic displays is great, physical implementations have struggled to approach their simulated potential. Computer-generated holography algorithms, which compute the modulation to be applied to the wavefront of light, are hampered by their reliance on simple simulation models of propagation that do not accurately predict the propagation in physical prototypes. This mismatch occurs because the images produced by holographic displays are very sensitive to aberrations which current physical setups have failed to eliminate completely. Early simulations of holographic displays make assumptions about the illumination of the display, the wavefront modulation by the display, and the propagation of light from the display which often fail to hold due to these aberrations.

This project investigates the potential of different deep learning configurations to more accurately model the propagation of light through a physical prototype. Specifically, we develop a framework for learning a transform in the complex domain from the ideal simulated wavefront to an aberrated wavefront which is consistent with captured data from a physical prototype. To produce the complex valued networks, we utilize JAX [3], a Python library which brings together native complex value support, automatic differentiation, and high-performance computing on accelerated hardware.

The performance of the proposed operation in the complex domain reveals benefits for retaining the complex rep-

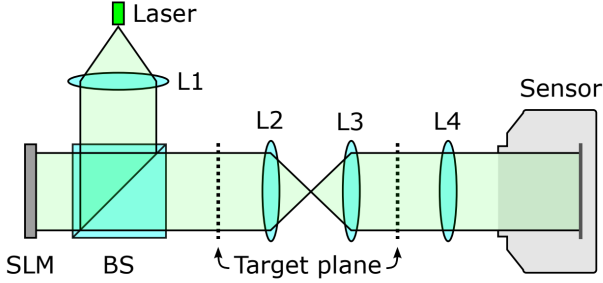


Figure 1. Diagram illustrating the components in a holographic display prototype [10]. This system was used to capture the phase+amplitudes pairs which we train our learned propagation model with. The SLM is illuminated by light from a laser source after it has been collimated by a lens (L1). A beam splitter (BS) redirects the wavefront to a reflective spatial light modulator (SLM) which modulates the incident wavefront and reflects the wavefront back through the BS and a 4f system that uses an aperture to filter out higher order modes in the Fourier space. Finally, the wavefront passes through an eyepiece lens (L4) before reaching the viewer.

resentation of the simulated wavefront by improving on the performance provided by prior learned architectures which operate solely on the amplitude of the simulated wavefront. However, these performance improvements do not hold across all complex architectures, and our work compares different ways of passing complex values through our network as well different activation functions that operate in the complex domain.

2. Related Work

Current implementations of holographic display prototypes use a coherent laser source to illuminate a spatial light modulator (SLM) which controllably delays the phase of the incident wavefront before it is relayed to a viewer. The original simulations of propagation through holographic displays assume spherical or planar illumination, perfectly controllable phase modulation at the SLM, and free space propagation from the SLM to the viewer [9]. Unfortunately, this simple propagation model fails to account for many of the effects present in a physical system. In a real system imperfections in the laser coherent source can result in illumination that does not perfectly match the assumed illumination. Additionally, the modulation at the SLM does not perfectly match the expected modulation. There is undiffracted light which is not modulated by the SLM, and there are phase non-linearities which produce deviations from the expected phase modulation. Finally, even the expectation of free space propagation from the SLM to the viewer is incorrect as there are often relay lenses and apertures between

the viewer and the SLM as illustrated in Figure 1.

Recent work has sought to model these unaccounted for effects by introducing additional learned terms to the simulated model [10]. These terms explicitly model the effect of content independent source and target field variations, Zernike aberrations along the optical propagation, phase non-linearities, and content-dependent undiffracted light. However, even with all of these added terms, this work still does not predict the propagation through the physical prototype very well. While these physically inspired parameters provide interpretability for the learned operations, they lack the flexibility to fully model the aberrations of the physical system.

Other recent work has used a more flexible model to learn the propagation of the light through the aberrated physical system [4]. Borrowing from recent machine learning advances, this work uses an image-to-image translation U-Net to learn the transformation from images predicted by an ideal angular spectrum method (ASM) propagation model and the images captured on a physical prototype. Since this U-Net performs an image-to-image translation, it operates solely on the amplitude of the wavefront at the target plane. However, without the phase information of the simulated wavefront, this approach also cannot learn the aberrations very well despite the additional flexibility of the U-Net architecture.

3. Method

The pipeline for our method is illustrated in Figure 2. In the proposed approach, a U-Net is directly applied to the complex ASM predicted wavefront in order to learn the aberrations of the physical system. Since this U-Net accepts the complex wavefront, it has more information than prior approaches which discarded the phase [4]. This should better model the physical aberrations since they are a function of the complex wavefront, not just the amplitude at the target plane.

3.1. Ideal Propagation Model

To implement this pipeline, we begin by modelling an ideal holographic systems using ASM. For this step, the wavefront at the SLM is modelled with a constant unit amplitude with the phase at each pixel modulated by an ideal SLM. If a phase of $\phi(x, y)$ is sent to the SLM, the wavefront at the SLM is modelled as $u_{SLM}(x, y)$ in Equation (1).

$$u_{SLM}(x, y) = \exp[\phi(x, y)] \quad (1)$$

In order to model propagation using the angular spectrum method, we first decompose the complex wavefront across the SLM plane into frequency components using a Fourier decomposition as described in Equation (2) where

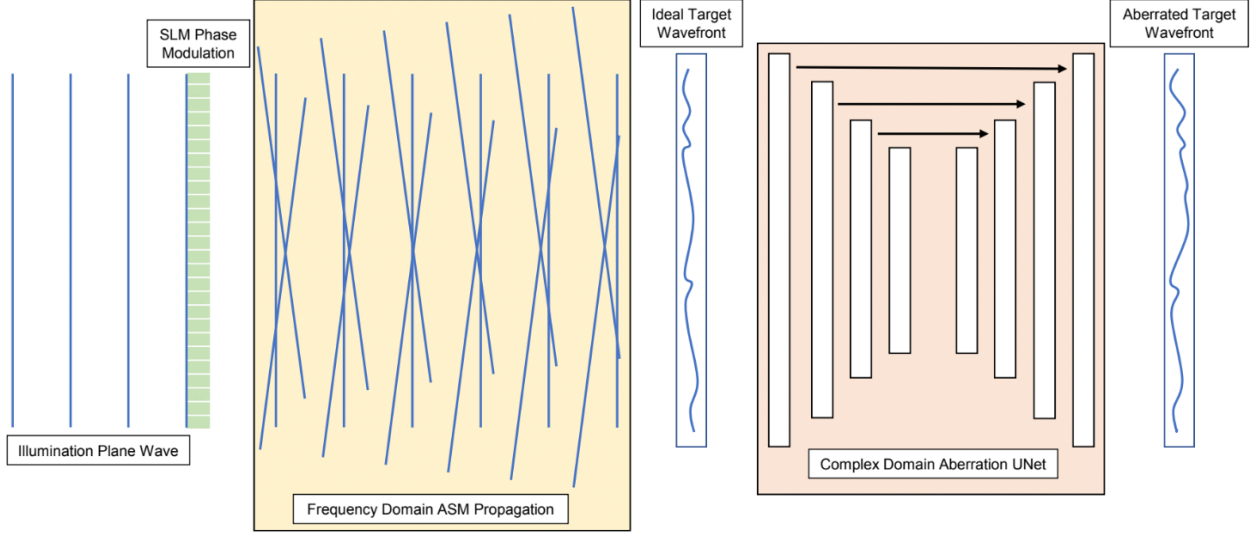


Figure 2. Diagram illustrating our proposed simulation pipeline. The pipeline begins by using ASM propagation to model the ideal target wavefront propagated from an ideal SLM illuminated by an ideal coherent source. Then a U-Net is applied to the ideal wavefront at the target plane to generate the aberrated wavefront.

$u_{SLM}(x, y)$ is the complex wavefront and $U_{SLM}(f_X, f_Y)$ is the resulting frequency decomposition [6].

$$U_{SLM}(f_X, f_Y) = \iint_{-\infty}^{\infty} u_{SLM}(x, y) \exp[-j2\pi(f_X x + f_Y y)] dx dy \quad (2)$$

These frequency components correspond to different plane waves propagating at different angles through the plane in space. The propagation of the complex wave through space can be viewed as the independent propagation of these component plane waves through space. The propagation of plane waves can be viewed as a phase shift, so the complex wave propagation is equivalent to different phase shifts for the different plane wave components. Together these phase shifts can be used to describe the change in the frequency components of a complex wave as it propagates over a distance z from the SLM to the target plane as expressed in Equation (3) where λ is the wavelength of light [6].

$$U_{target}(f_x, f_y) = U_{SLM}(f_x, f_y) \exp\left(j \frac{2\pi}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} z\right) \quad (3)$$

Finally at the target plane, the wavefront can be reformed

using the inverse Fourier transform as in Equation (4).

$$u_{target}(f_X, f_Y) = \iint_{-\infty}^{\infty} U_{target}(x, y) \exp[j2\pi(f_X x + f_Y y)] dx dy \quad (4)$$

3.2. Aberration Network

For our aberration network, we use the same U-Net architecture as prior work which learned the aberrations as a function of the amplitude at the target plane [4]. This allows us to focus on the benefits of utilizing the complex domain in our network. Applying a U-Net on the complex amplitude does require some modifications of conventional U-Net architectures. The `conv` and `convtranspose` layers of PyTorch do not support complex convolution, which motivates the use of JAX which has more developed support for complex valued operations. We also use Flax [7] which is a neural network library and ecosystem built on JAX that has been designed for flexibility to support research use cases.

Since the activation functions of conventional U-Net architectures are designed for real values, we implement various complex activation functions whose modulation functions are illustrated in Figure 3. For our activation functions, the first two of our non-linearities simply modulate based on the quadrant of the complex values. The complex ReLU passes all values in the first quadrant of the complex plane, and the real ReLU passes all values in the first and second quadrant of the complex plane. The modReLU activation function modulates a complex value, x , based solely

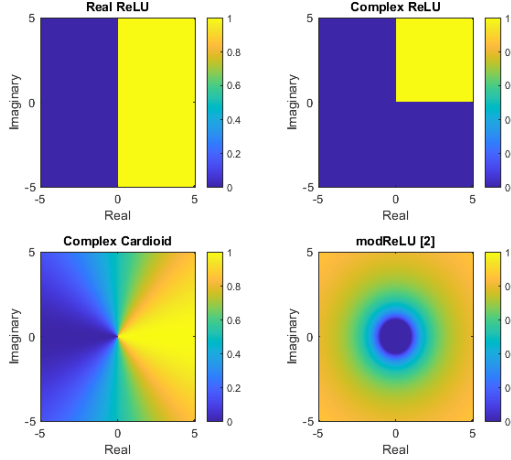


Figure 3. Visualizations of our complex activation functions. Here the modulation of the complex values is illustrated as a function of the location of the complex values on the complex plane.

on its amplitude using a learnable parameter, b , as in Equation (5) [2].

$$\text{modReLU}(x) = \frac{\max(0, |x| + b)}{|x|} * x \quad (5)$$

Finally, the complex cardioid function modulates an input complex value, x , as described in Equation (6).

$$\text{cardioid}(x) = \frac{0.5 * (1 + \text{real}(x))}{|x|} * x \quad (6)$$

We also experimented with different formats of inputs and outputs for the complex aberration network. The complex wavefronts are either represented by complex values directly or by the real and imaginary values of the wavefront stacked as separate channels. In the remainder of the report, the former case is referenced as ComplexCNNc, and the latter case is referenced as StackedCNNc. In our StackedCNNc configuration, the network had the additional flexibility of treating the real and imaginary components as independent channels for the intermediate convolution and activation layers. While this has more flexibility, it loses the intrinsic regularization provided by complex convolutions and activations which operate in the domain of the light wavefront. While ComplexCNNc has this benefit of operating in the complex domain throughout the whole network, we found that training ComplexCNNc with the predefined JAX optimizers fails to converge to a good loss value. As a result, we developed a modified Adam optimizer in JAX which takes steps in the negative conjugate gradient direction as described in Algorithm 1. This optimizer was crucial for achieving the ComplexCNNc results which are discussed in subsequent sections.

Algorithm 1: Modified Adam Optimizer Update Rule. Inputs are α (step size), $\beta_1, \beta_2 \in [0, 1]$ (exponential decay rates for moment estimates), $f(\theta)$ stochastic objective function, θ represents parameters, and θ_0 the initial parameters.

Result: θ_t Resulting parameters

```

 $m_0 \leftarrow 0$   $v_0 \leftarrow 1$   $t \leftarrow 0$ 
while  $\theta_t$  not converged do
   $t \leftarrow t + 1$ 
   $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ 
   $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ 
   $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ 
   $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ 
   $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ 
   $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / \sqrt{\hat{v}_t} + \varepsilon$ 
end

```

4. Results

To compare the performance of different network configurations, we train our learned propagation models on a dataset of 10,200 SLM phase and target amplitude pairs broken up into 8/1/1 train/validation/test split. The phase set was composed of intermediate phases sampled every 100th iteration produced with the iterative ASM SGD and camera-in-the-loop (CITL) optimization using SGD procedures proposed in prior work [10]. The SGD optimization procedure was run on 1600 images, and the CITL optimization using SGD procedure was run on 100 images. For the captured amplitudes, these phase patterns were displayed on the holographic system from a prior work [10]. The setup from this work uses a phase-only SLM with $1,920 \times 1,080$ pixels with a pixel size of $6.4 \times 6.4 \mu\text{m}$. The illumination is from a 517.7 nm laser source, and the target image is 13.33 cm away from the SLM.

We compared the performance of various network configurations on this dataset. We trained the ComplexCNNc networks using each of the complex activation functions described in the methods section. Table 1 quantitatively compares the performance of the networks trained with these configurations as well as the StackedCNNc network on our dataset. As a note, the Adam optimizer was developed shortly before the poster presentation, so the performance of the ComplexCNNc networks has improved since then as the networks have continued training. Based on the results in Table 1, the best ComplexCNNc and StackedCNNc network configurations are compared with prior works in Table 2. Finally, sample amplitudes generated from our pipeline with the StackedCNNc and best ComplexCNNc networks are displayed in Figure 4 and compared against the captured amplitudes and the ideal amplitudes.

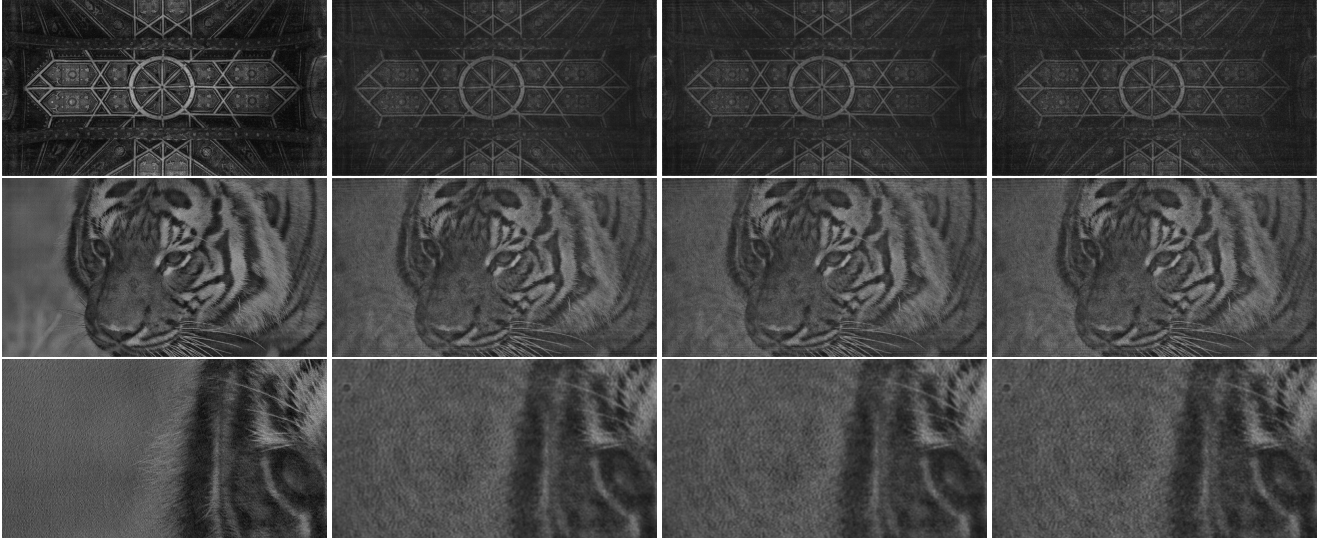


Figure 4. Amplitudes predicted using our StackedCNNc and best ComplexCNNc networks are compared against captured amplitudes and ideal ASM predicted amplitudes. The columns from left to right contain the ASM, ComplexCNNc, StackedCNNc, and captured amplitudes respectively. The last row contains zoomed crops from the images in the second row.

Network Type	Activation	Train PSNR	Test PSNR
ComplexCNNc	Complex ReLU	30.6 dB	30.4 dB
	Real ReLU	32.8 dB	32.4 dB
	Complex Cardioid	27.9 dB	27.1 dB
	ModReLU	26.2 dB	25.5 dB
StackedCNNc	ReLU	33.7 dB	32.9 dB

Table 1. Comparisons of performance of different complex aberration network configurations on dataset of 10,200 phase and capture pairs.

Method	Train PSNR	Test PSNR
ASM	20.6 dB	20.7 dB
Neural Holography [10]	29.3 dB	29.2 dB
HIL Holography [4]	29.7 dB	29.6 dB
ComplexCNNc	32.8 dB	32.4 dB
StackedCNNc	33.7 dB	32.9 dB

Table 2. Comparisons of performance of our best complex aberration network configurations with methods from prior works on dataset of 10,200 phase and capture pairs.

5. Analysis

The results from our experiments illustrate the benefit of applying the aberration U-Net on the complex wavefront at the target plane instead of just applying the network on the amplitude of the wavefront at the target plane. Using our modified Adam optimizer with the ComplexCNNc networks, we were able to exceed the performance of prior works. These give us the results in the fourth row of Ta-

ble 2 and confirms our hypothesis from the start of this project that combining the flexibility of machine learning approaches along with operation on the complex wavefront could yield better results than prior work. With its increased flexibility, our StackedCNNc networks improve the performance of our learned propagation model slightly more producing the results in the fifth row of Table 2.

Our experiments with different activation functions indicated that the complex activations that were similar to the standard ReLU function performed the best in the complex domain. ComplexReLU with ComplexCNNc performed slightly better than prior work on the captured dataset, and RealReLU with ComplexCNNc performed significantly better. The ComplexCNNc networks trained with ModReLU and complex cardioid activation functions predicted the target intensity better than the ideal ASM simulations but not as well as the prior methods.

The simulated amplitudes generated with our StackedCNNc and best ComplexCNNc networks in Figure 4 illustrate the benefits of the learned propagation model. The ideal ASM simulated amplitudes lack many of the phenomena present in the captured image. The scene in the first row highlights how the ideal ASM simulated amplitudes do not model the undiffracted light which results in the lowered contrast for the image captured in the physical system. Both our trained networks reproduce this phenomena. The scene in the second row, along with the zoomed in crop in the third row, highlight how the ideal ASM simulated amplitudes do not capture the ripple patterns present in the captured image. However, this ripple pattern near the eye of the tiger is

once again reproduced in both of our predicted amplitudes.

6. Discussion

In this project we analyzed the benefits that could be achieved by using an aberration neural network to learn the transform from ideal simulated wavefronts to aberrated wavefronts which are consistent with captured amplitudes. This approach merges the machine learning approaches of prior work which used image-to-image translation and the complex domain operation of prior work which learned physically-inspired transformations applied to the ideally simulated wavefronts.

Our results show that the added information from the full complex representation of the ideal wavefront at the target wavefront can be used to train aberration networks that perform better than prior approaches. By training with our modified Adam optimizer, our ComplexCNNc network which used complex valued channels and our RealReLU activation performed about 3db better than prior work. These performance gains can be slightly increased with our StackedCNNc network which treats the real and imaginary components as independent terms within the network.

In further work, we would be interested in analyzing if aberration networks with fewer channels could be used when operating with complex values based on the increased information stored in each complex value. We are also interested in seeing if the implicit regularization of the complex domain operations allows ComplexCNNc networks to generalize better than the StackedCNNc network for novel phase distributions. Finally we would want to see if the complex-valued networks perform better for the problem of predicting intensities at multiple target planes when the complex information at the target wavefronts is even more relevant.

References

- [1] K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke. Near-eye varifocal augmented reality display using see-through screens. *ACM Transactions on Graphics (TOG)*, 36(6):1–13, 2017.
- [2] M. Arjovsky, A. Shah, and Y. Bengio. Unitary evolution recurrent neural networks. In *International Conference on Machine Learning*, pages 1120–1128. PMLR, 2016.
- [3] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018.
- [4] P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Transactions on Graphics (TOG)*, 39(6):1–18, 2020.
- [5] J.-H. Chang, B. V. K. Vijaya Kumar, and A. C. Sankaranarayanan. Towards multifocal displays with dense focal stacks. *SIGGRAPH Asia / ACM Trans. Graphics*, 37(6):1–13, 2018.
- [6] J. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Series in Electrical and Computer Engineering: Communications and Signal Processing. McGraw-Hill, 1996.
- [7] J. Heek, A. Levskaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner, and M. van Zee. Flax: A neural network library and ecosystem for JAX, 2020.
- [8] H. Hua and B. Javidi. A 3d integral imaging optical see-through head-mounted display. *Optics express*, 22(11):13484–13491, 2014.
- [9] A. Maimone, A. Georgiou, and J. S. Kollin. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)*, 36(4):1–16, 2017.
- [10] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein. Neural holography. In *ACM SIGGRAPH 2020 Emerging Technologies*, pages 1–2. 2020.