Time-multiplexed Neural Holography: A Flexible Framework for Holographic Near-eye Displays with Fast Heavily-quantized Spatial Light Modulators

SUYEON CHOI*, Stanford University, USA
MANU GOPAKUMAR*, Stanford University, USA
YIFAN PENG, Stanford University, USA
JONGHYUN KIM, NVIDIA and Stanford University, USA
MATTHEW O'TOOLE, Carnegie Mellon University, USA
GORDON WETZSTEIN, Stanford University, USA

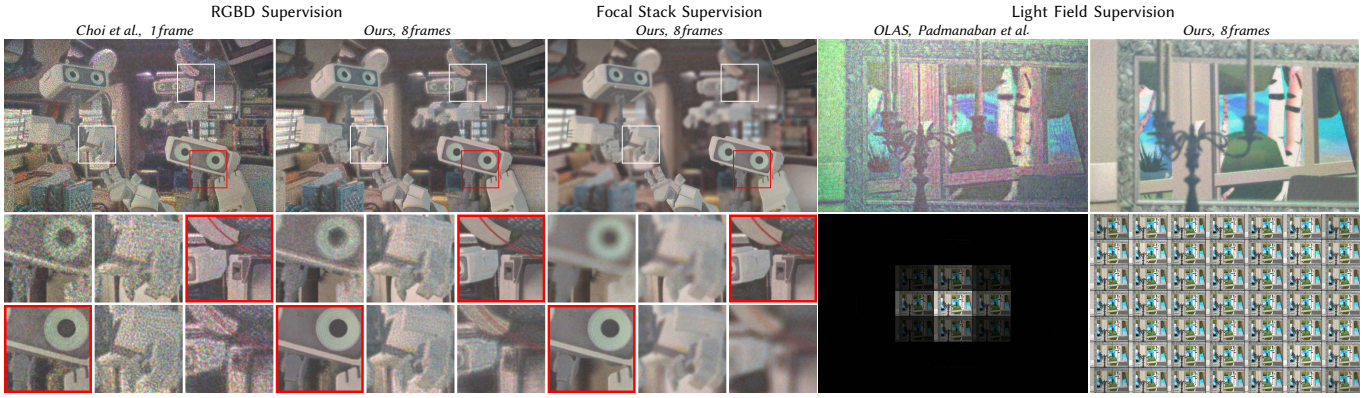

Fig. 1. Computer-generated holography (CGH) results captured with a display prototype that uses a fast, low-precision (i.e., 4 bit) phase spatial light modulator (SLM). When supervised with 2.5D RGBD images, our approach (2nd column) provides a better image quality than the state-of-the-art neural 3D holography algorithm [Choi et al. 2021a] (1st column) using this low-precision SLM. Our CGH framework is flexible in not only enabling 2.5D but also 3D focal stack and 4D light field supervision. The former approach (3rd column) results in the best in-focus (red boxes) and out-of-focus (white boxes) image quality among 2.5D and 3D CGH algorithms. Our 4D light field–supervised approach (5th column) outperforms the recently proposed OLAS method [Padmanaban et al. 2019] (4th column) by a large margin and utilizes the space–bandwidth product more effectively, as shown by the simulated light fields in the lower right images.

Holographic near-eye displays offer unprecedented capabilities for virtual and augmented reality systems, including perceptually important focus cues. Although artificial intelligence–driven algorithms for computer-generated holography (CGH) have recently made much progress in improving the image quality and synthesis efficiency of holograms, these algorithms are not directly applicable to emerging phase-only spatial light modulators (SLM) that are extremely fast but offer phase control with very limited precision. The speed of these SLMs offers time multiplexing capabilities, essentially enabling partially-coherent holographic display modes. Here we report advances in camera-calibrated wave propagation models for these types of holographic near-eye displays and we develop a CGH framework that robustly optimizes the heavily quantized phase patterns of fast SLMs. Our framework is flexible in supporting runtime supervision with different types of content, including 2D and 2.5D RGBD images, 3D focal stacks, and 4D light fields. Using our framework, we demonstrate state-of-the-art results for all of these scenarios in simulation and experiment.

CCS Concepts: • **Hardware** → **Emerging technologies**; • **Computing methodologies** → **Computer graphics**.

Additional Key Words and Phrases: computational displays, holography, virtual reality

1 INTRODUCTION

Holographic near-eye displays for virtual and augmented reality (VR/AR) applications offer many benefits to wearable computing systems over conventional microdisplays. These include high peak brightness, power efficiency, support of perceptually important focus cues and vision-correcting capabilities [Kim et al. 2021], as well as thin device form factors [Kim et al. 2022; Maimone and Wang 2020]. Yet, the image quality achieved by computer-generated holography (CGH) lags far behind that of conventional displays, requiring further advancements in the algorithms driving holographic displays.

Recently, artificial intelligence (AI) methods have enabled significant improvements in image quality [Chakravarthula et al. 2020;

*denotes equal contribution.

Authors' addresses: Suyeon Choi, suyeon@stanford.edu, Stanford University, USA; Manu Gopakumar, manugopa@stanford.edu, Stanford University, USA; Yifan Peng, evanpeng@stanford.edu, Stanford University, USA; Jonghyun Kim, jonghyunk@nvidia.com, NVIDIA and Stanford University, USA; Matthew O'Toole, mpotoole@cmu.edu, Carnegie Mellon University, USA; Gordon Wetzstein, gordon.wetzstein@stanford.edu, Stanford University, USA.

Choi et al. 2021a; Peng et al. 2020] and speed [Horisaki et al. 2018; Peng et al. 2020; Shi et al. 2021] of holographic displays. These algorithms, however, are primarily applicable to slow liquid crystal–based (LC) spatial light modulators (SLMs) that offer control of the phase of a coherent light source at high precision. Emerging micro-electromechanical (MEMS) phase SLMs [Bartlett et al. 2019] offer potential benefits over LC-based systems in being more light efficient, significantly faster, better suited to operate across a wide range of wavelengths, and more stable for varying temperatures. Indeed, MEMS-based amplitude SLMs are one of the most popular technology choices for many display applications, including projectors, so MEMS-based phase SLMs may also become increasingly important for holography applications. Unfortunately, the algorithms developed for high-precision LC-based phase SLMs suffer from a degradation in image quality and fail to fully utilize time-multiplexing when used with the high framerate, heavily quantized phase control that MEMS-based SLMs offer. For example, DLP's phase SLM by Texas Instruments only offers up to 4 bits of precision or, similarly, 16 unevenly distributed discrete levels of phase control at frame rates of 1440 Hz [Bartlett et al. 2019; Ketchum and Blanche 2021].

The focus of our work is to extend AI-driven CGH algorithms to operate with emerging fast but heavily quantized phase SLMs. This is a non-trivial task, because quantization is non-differentiable, so the standard machine learning toolset does not directly apply in these settings. Moreover, most of the degrees of freedom of a holographic display stem from their ability to create constructive and destructive interference, which can only be achieved instantaneously in time but not between time-multiplexed frames. It is thus not clear whether the partially-coherent holographic display mode enabled by the fast SLM speed is actually beneficial when combined with a limited precision of phase control or how it affects image quality. We propose an algorithmic CGH framework that robustly optimizes holograms in these mathematically challenging scenarios and explore the aforementioned tradeoff, demonstrating significant benefits in image quality and space–bandwidth utilization [Yoo et al. 2021] of higher-speed phase SLMs. Moreover, we develop a learned propagation model that is more flexible than previously proposed alternatives in allowing us to calibrate it using 3D multiplane supervision but leverage a variety of target content, including 2D images, 2.5D RGBD images, 3D focal stacks, and 4D light fields, for supervision during runtime.

Specifically, our contributions include the following:

- a new variant of a camera-calibrated wave propagation model for holographic displays, which is flexible in enabling runtime supervision by 2D, 2.5D, 3D, or 4D content;
- a framework for robust CGH optimization with fast but heavily quantized phase-only SLMs;
- experimental demonstration of improved image quality and better utilization of the SLM's space–bandwidth product enabled by our framework.

Source code for this paper is available at computationalimaging.org.

2 RELATED WORK

Many aspects of holographic displays, including optics, SLMs, and algorithms, have advanced considerably over the last few years. Detailed discussions of many of these advancements can be found in the survey papers by Yaras [2010], Park [2017], and Chang et al. [2020]. A recent roadmap article by Javidi et al. [2021] also outlines current and future research efforts of digital holography in non-display areas, including 3D imaging and microscopy.

Our work primarily focuses on advancing the algorithms driving holographic near-eye displays. In a nutshell, the CGH problem comprises several parts. First, the target content is specified in some format that needs to be converted to a complex-valued wavefield, such as point clouds [Fienup 1982; Gerchberg 1972; Maimone et al. 2017; Shi et al. 2017, 2021], polygons [Chen and Wilkinson 2009; Matsushima and Nakahara 2009], light rays [Wakunami et al. 2013; Zhang et al. 2011], image layers [Chen et al. 2021; Chen and Chu 2015; Zhang et al. 2017], or light fields [Benton 1983; Kang et al. 2008; Lucente and Galyean 1995; Padmanaban et al. 2019; Ziegler et al. 2007]. Second, this wavefield needs to be encoded by a phase-only SLM, which can be achieved by fast, direct phase coding approaches [Hsueh and Sawchuk 1978; Lee 1970; Maimone et al. 2017] or slow, iterative solvers, such as classic Gerchberg–Saxton-type algorithms [Fienup 1982; Gerchberg 1972] or variants of stochastic gradient descent [Chakravarthula et al. 2019; Peng et al. 2020].

Yet, the simulated wave propagation models used by most of these CGH algorithms do not always model the physical optics faithfully, thereby degrading image quality. Moreover, the computational complexity of these algorithms often prevents them from being practical in the power-constrained settings of a wearable computing system. Emerging artificial intelligence–driven CGH approaches have focused on addressing these limitations. For example, surrogate gradient methods that use a camera in the loop (CITL) for hologram optimization can significantly improve image quality [Choi et al. 2021b; Peng et al. 2021, 2020]. Alternatively, differentiable wave propagation models can be learned to calibrate for the gap between simulated models and physical optics [Chakravarthula et al. 2020; Choi et al. 2021a; Kavakli et al. 2022; Peng et al. 2020]. Moreover, neural networks can be trained to enable real-time CGH algorithms [Horisaki et al. 2021, 2018; Peng et al. 2020; Shi et al. 2021].

Note that our work is concurrently and independently developed from the very recent work by Lee et al. [2022]. Although both works share some similarity in applying constrained gradient descent methods to optimize binary or heavily-quantized phase holograms, our framework outperforms the counterpart with the use of a learned propagation model for better image quality, the ability to effectively handle SLMs with varied bit depths and non-linear quantizations, and compatibility with a wide range of supervision sources.

3 A FLEXIBLE FRAMEWORK FOR CGH

In Fresnel holography, a collimated coherent light beam illuminates an SLM with a source field u_{src}, and the light reflected in response reproduces a target intensity distribution. To generate this hologram, a phase-only SLM imparts a spatially-varying delay ϕ on the phase of the field. After propagating a distance z from the SLM, the resulting

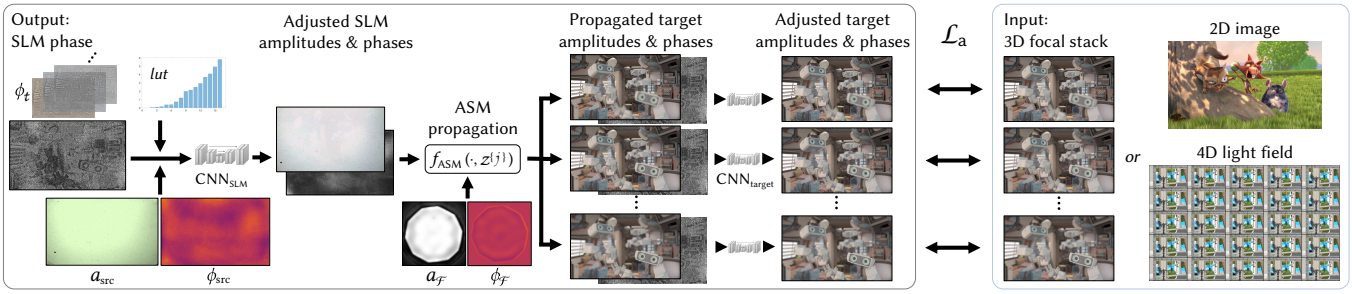

Fig. 2. Illustration of our calibrated wave propagation model and 2D/3D/4D supervision strategy for the multiplexed, quantized hologram generation. The complex-valued field at the SLM is adjusted by several learnable terms (amplitude and phase at the SLM plane as well as look-up table for phase mapping) and then processed by a CNN. The resulting complex-valued wave field is propagated to all target planes using the ASM wave propagation operator with two extra learnable terms (amplitude and phase at the Fourier domain). The wave fields at each target plane are processed again by smaller CNNs. The proposed framework applies to multiple input forms, including 2D, 2.5D, 3D, and 4D.

complex-valued field u_z is given by the following image formation model:

$$u_z\left(x,y,\lambda\right) = f\left(u_{\text{SLM}}\left(x,y,\lambda\right),z\right),$$
$$u_{\text{SLM}}\left(x,y,\lambda\right) = e^{iq(\phi(x,y,\lambda))}u_{\text{src}}\left(x,y,\lambda\right), \tag{1}$$

where λ is the wavelength of light, x, y are the transverse coordinates, and u_{SLM} is the modulated field at the SLM. The wave propagation operator f models free-space propagation between two parallel planes separated by a distance z. For notational convenience, we will omit the dependence on x, y, λ and the source field u_{src}. The intensity pattern generated by this display at distance z in front of the SLM when showing phase ϕ is therefore $\left|f\left(e^{iq(\phi)},z\right)\right|^2$.

When using low-bit SLMs for time-multiplexed holography, the effect of quantization is not negligible. To model a quantized phase-only SLM with $M \times N$ pixels, where every pixel offers phase control with limited precision, we define a quantization operator q:

$$q : \mathbb{R}^{M\times N} \to \boldsymbol{Q}^{M\times N}, \quad \phi \mapsto q(\phi) = \Pi_{\boldsymbol{Q}}\left(\phi\right), \tag{2}$$

where Π is the projection operator that maps the continuous phase value to the closest discrete phase in the feasible set $\boldsymbol{Q}$ supported by the SLM.

Our framework approaches computer-generated holography with a differentiable camera-calibrated image formation model (Sec. 3.1), an optimization procedure designed for quantized SLMs (Sec. 3.2), and a family of loss functions supervised on either 2D, 2.5D, 3D, or 4D content to produce time-multiplexed holograms (Sec. 3.3). Figure 2 illustrates our model and optimization pipeline.

3.1 Camera-calibrated Wave Propagation Model

Recent work on holographic displays has demonstrated that the naive application of simulated wave propagation models, like the angular spectrum method (ASM) [Goodman 2014], to holographic displays fails to account for the non-idealities of the physical optical system, such as phase distortions of the SLM, optical aberrations, and the limited diffraction efficiency of the SLM [Chakravarthula et al. 2020; Choi et al. 2021a; Peng et al. 2020]. This discrepancy between simulated and physical image formation adversely affects image quality, but can be overcome by learning to calibrate for the physical optics using a differentiable, neural network–parameterized propagation model.

Here, we propose a variant of the learned model recently proposed by Choi et al. [2021a]:

$$f_{\text{model}}\left(u_{\text{SLM}},z\right) = \text{CNN}_{\text{target}}\left(\mathcal{P}_{\text{ASM}}\left(\text{CNN}_{\text{SLM}}\left(a_{\text{src}}e^{i\phi_{\text{src}}}u_{\text{SLM}}\right),z\right)\right),$$
$$\mathcal{P}_{\text{ASM}}\left(u,z\right) = \iint \mathcal{F}\left(u\right)\cdot\mathcal{H}\left(f_x,f_y,\lambda,z\right)e^{i2\pi\left(f_x x+f_y y\right)}df_x df_y,$$
$$\mathcal{H}\left(f_x,f_y,\lambda,z\right) = a_{\mathcal{F}}e^{i\left(\frac{2\pi}{\lambda}z\sqrt{1-\left(\lambda f_x\right)^2-\left(\lambda f_y\right)^2}+\phi_{\mathcal{F}}\right)}, \tag{3}$$

where CNN_{SLM} and $\text{CNN}_{\text{target}}$ are convolutional neural networks that operate on the complex field at the SLM and target planes. The target plane is a distance z from the SLM. In addition, a_{src} and ϕ_{src} are learned to account for content-independent spatial variations in amplitude and phase of the incident source field at the SLM plane while $a_{\mathcal{F}}$ and $\phi_{\mathcal{F}}$ are added to the ASM propagation to learn spatial variations in amplitude and phase in the Fourier plane similarly to the learned complex convolutional kernel presented by Kavakli et al. [2022].

Similar to Choi et al., we capture a training and a test set comprised of a large number of SLM phase patterns and corresponding amplitude images recorded at a set of distances $\{j\}, j = 1\ldots J$ with our prototype holographic display. Using a standard stochastic gradient descent–type solver, we then fit the parameters of the CNNs, CNN_{SLM} and $\text{CNN}_{\text{target}}$, as well as $a_{\text{src}}, a_{\mathcal{F}}, \phi_{\text{src}}, \phi_{\mathcal{F}}$ to learn the calibrated wave propagation model. The model used in this framework builds upon the model from Choi et al. by using the terms a_{src}, ϕ_{src}, $\phi_{\mathcal{F}}$, and $a_{\mathcal{F}}$ to learn many of the content-independent non-idealities of the holographic system. The source terms can efficiently model the effects of non-ideal illumination at the SLM plane, and the Fourier plane terms can compactly account for the effects of non-ideal optical filtering. Together these terms enable the use of smaller convolutional neural networks to learn the content-dependent non-idealities, such as the spatially varying pixel response at the SLM. Table 1 quantitatively assesses the effect of these physically-inspired parameters by evaluating the performance of different calibrated wave propagation models on a captured dataset. All models are trained over 6 intensity planes, corresponding to 0.0 D, 0.5 D, 1.0 D,

Table 1. Comparison of different calibrated wave propagation models. All models are trained on 6 of the 7 planes. PSNR is evaluated for training and test sets as well as for the 7th held-out plane. The number of parameters of each model is also reported. Training details are listed in Supplement S2.4.

Models	Params.	Train	Test	Held-out
NH [Peng et al. 2020]	4.1M	26.7	27.1	26.3
NH3D [Choi et al. 2021a]	68.5M	34.4	32.4	31.9
Our model, CNNs only	6.2M	31.6	29.7	30.0
+ a_{src}	7.2M	35.3	35.4	32.3
+ a_{src} + ϕ_{src}	8.2M	36.2	36.3	33.0
+ a_{src} + ϕ_{src} + $\phi_{\mathcal{F}}$	12.3M	36.5	36.4	32.8
+ a_{src} + ϕ_{src} + $\phi_{\mathcal{F}}$ + lut	12.3M	36.4	36.4	32.8
+ a_{src} + ϕ_{src} + $\phi_{\mathcal{F}}$ + $a_{\mathcal{F}}$ + lut	16.4M	**36.7**	**36.7**	32.6

1.5 D, 2.5 D, and 3.0 D in the physical space. A 7th plane at 2.0 D is set as the held-out plane for evaluation. In this table, we also ablate the performance of an additional *lut* parameter to optionally learn the feasible set $\boldsymbol{Q}$ of quantized values supported by the SLM. We observe that our model (bottom row) significantly reduces the number of parameters when compared to the original NH3D model, while still producing the highest PSNR metrics on the test set and the held-out plane. Notably, the lagging performance of the NH model, which is purely composed of physically-inspired terms, illustrates the substantial benefit of incorporating the flexibility of CNNs in a calibrated propagation model. Further details on our model architecture and training are included in Supplement S2.4

3.2 Optimizing Phase Patterns for Quantized SLMs

Emerging MEMS-based phase SLMs are fast but offer only a limited precision for controlling phase. DLP's phase SLM by Texas Instruments (TI) [Bartlett et al. 2019], for example, runs at a maximum framerate of 1440 Hz grayscale but only offers 4 bits, or 16 discrete phase levels, at each of the frames. We therefore need to derive methods that allow us to optimize phase patterns for heavily quantized phase SLMs. The primary problem is that the quantization function q is not differentiable. To this end, we discuss and evaluate several strategies for dealing with q assuming some simple 2D loss function $\mathcal{L}\left(s \cdot \left|f_{\text{model}}\left(e^{iq(\phi)}, 0\right)\right|, a_{\text{target}}\right)$, where a_{target} is the desired 2D amplitude, and s is a scale parameter that is optimized along with ϕ.

The naive solution to dealing with q is to simply ignore it. Specifically, the phase pattern ϕ can be optimized given a 2D target amplitude image a_{target} and quantized to the available precision after the optimization. This is the approach typically adopted by state-of-the-art CGH algorithms that work well for liquid crystal–type phase SLMs, because these SLMs offer 8 bit or higher precision phase modulation. TI's MEMS device enables time multiplexing but only offers 4 bits, which makes this approach impractical (see Fig. 3). Instead, the reference code supplied with the SLM implements a variant of projected gradient descent [Boyd et al. 2004], which projects the iteratively updated solution onto the feasible set of quantized values $\boldsymbol{Q}$. This approach is equivalent to a gradient descent–type update scheme that applies q after each iteration k as:

$$\widehat{\phi}^{(k)} \leftarrow \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial \phi}\right)^T \mathcal{L}\left(s \cdot \left|f_{\text{model}}\left(e^{i\phi^{(k-1)}}\right)\right|, a_{\text{target}}\right),$$
$$\phi^{(k)} \leftarrow \Pi_{\boldsymbol{Q}}\left(\widehat{\phi}^{(k)}\right) = q\left(\widehat{\phi}^{(k)}\right). \tag{4}$$

As an alternative solution to solving these types of problems, surrogate gradient methods are often used [Bengio et al. 2013; Zenke and Ganguli 2018]. Here, the forward pass is computed using the correct quantization function q but during the error backpropagation pass, the gradients of a differentiable proxy function $\widehat{q}$ are used. This enables improved optimization of phase patterns through a quantization layer with the minimal overhead of computing the proxy gradients:

$$\phi^{(k)} \leftarrow \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial q} \cdot \frac{\partial \widehat{q}}{\partial \phi}\right)^T \mathcal{L}\left(s \cdot \left|f_{\text{model}}\left(e^{iq\left(\phi^{(k-1)}\right)}\right)\right|, a_{\text{target}}\right). \tag{5}$$

Perhaps the most common choice for $\widehat{q}$ is a sigmoid function, whose slope can be gradually annealed during training [Bengio et al. 2013; Chung et al. 2016; Zenke and Ganguli 2018].

We propose the use of a continuous relaxation of categorical variables using Gumbel-Softmax [Jang et al. 2016; Maddison et al. 2016] for optimizing heavily quantized phase values in CGH applications. This approach has several desirable properties. First, the Gumbel noise and categorical relaxation prevent the optimization from getting stuck in local minima, which is perhaps the primary benefit over other surrogate gradient methods. Second, annealing of the temperature parameter τ of the softmax as well as the shape of the score function are directly supported. Formally, this approach is written as:

$$\widehat{q}(\phi) = \sum_{l=1}^{L} \boldsymbol{Q}_l \cdot \mathcal{G}_l\left(\mathbf{score}\left(\phi, \boldsymbol{Q}\right)\right), \tag{6}$$

$$\mathcal{G}_l(z) = \frac{\exp\left(\left(z_l + g_l\right)/\tau\right)}{\sum_{l=1}^{L} \exp\left(\left(z_l + g_l\right)/\tau\right)}, \tag{7}$$

$$\mathbf{score}_l\left(\phi, \boldsymbol{Q}\right) = \sigma\left(w \cdot \delta\left(\phi, \boldsymbol{Q}_l\right)\right)\left(1 - \sigma\left(w \cdot \delta\left(\phi, \boldsymbol{Q}_l\right)\right)\right), \tag{8}$$

where $g_l \sim \text{Gumbel}(0, 1)$ is the Gumbel noise for all of the $l = 1, \ldots, L$ categories, i.e., quantized phase levels, σ is a sigmoid function, δ is the signed angular difference, and w is a scale factor (see Jang et al. [2016] and the supplement for additional details).

3.3 Runtime Supervision of Time-multiplexed Holograms

Fast MEMS-based phase SLMs can produce higher-quality holograms through time multiplexing, i.e., intensity averaging of multiple frames. Given our camera-calibrated wave propagation model (Sec. 3.1), we optimize for time-multiplexed holograms using different target content at runtime.

2D Holography. In this case, we wish to synthesize a 2D intensity image at a distance z in front of the phase SLM. The distance can be fixed or dynamically varied in software to enable a varifocal

holographic display mode. For this purpose, we specify the loss:

$$\mathcal{L}_{\text{2D}} = \mathcal{L}\left(s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|f_{\text{model}}\left(e^{iq(\phi^{(t)})}, z\right)\right|^2}, a_{\text{target}}\right), \tag{9}$$

between the target amplitude image a_{target} and the simulated holographic image and solve for ϕ. We can easily formulate a time-multiplexed variant of the CGH problem using this loss function by summing over $t = 1 \dots T$ squared amplitudes, i.e., intensities, where T refers to the total number of time-multiplexed frames that can be displayed throughout the exposure time of the human eye. The simplest example of the loss function $\mathcal{L}$ is an ℓ_2 loss although other loss functions, such as perceptually motivated image quality metrics, could be applied as well.

2.5D Holography using RGBD Input. Using the multiplane loss function presented by Choi et al. [2021a], holograms can be synthesized to generate a 2D set of intensities at depths specified by a depth map. We refer the interested reader to Supplement S2.5 for the loss function and an additional discussion on utilizing time multiplexing to produce natural blur with 2.5D supervision.

3D Multiplane Holography. True 3D holography can be achieved by optimizing a single SLM phase pattern ϕ or a series of time-multiplexed patterns $\phi^{(t)}$ for the target amplitude of a focal stack $\text{fs}_{\text{target}}$. The corresponding loss function in our framework looks very similar to that of the 2D hologram above, although it is evaluated over the set of focal slices $\{j\}$:

$$\mathcal{L}_{\text{3D}} = \mathcal{L}\left(s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|f_{\text{model}}\left(e^{iq(\phi^{(t)})}, z\{j\}\right)\right|^2}, \text{fs}_{\text{target}}\right). \tag{10}$$

Effectively optimizing this focal stack loss using the full blur available within the diffraction angle of the SLM requires time multiplexing as illustrated in Supplement S2.6.

4D Light Field Holography. Finally, we can also supervise our CGH framework using the amplitudes of a 4D target light field $\text{lf}_{\text{target}}$. For this purpose, a differentiable hologram-to-light field transform is required, which can be calculated using the Short-time Fourier transform (STFT) [Padmanaban et al. 2019; Zhang and Levoy 2009]:

$$\mathcal{L}_{\text{4D}} = \mathcal{L}\left(s\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\text{STFT}\left(f_{\text{model}}\left(e^{iq(\phi^{(t)})}, z\right)\right)\right|^2}, \text{lf}_{\text{target}}\right). \tag{11}$$

By utilizing time multiplexing, our optimized holograms can uniquely reproduce a set of light field views that fully covers the SLM's space–bandwidth product as detailed in Supplement S2.7.

4 EXPERIMENTS

To evaluate our novel algorithms, we use a benchtop 3D holographic display prototype. This prototype includes a FISBA RGBeam fiber-coupled module with red, green, and blue optically aligned laser diodes for illumination and a TI DLP6750Q1EVM phase SLM for high-speed quantized phase modulation. We capture the images produced by this prototype with a FLIR Grasshopper3 12.3 MP color USB3 sensor through a Canon EF 35mm lens with focus controlled by an Arduino microcontroller. Further details of the prototype are included in Supplement S1.

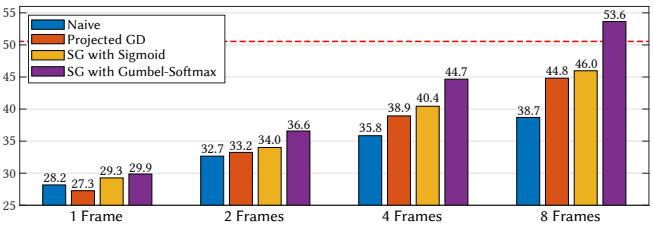

Fig. 3. Evaluation of CGH algorithms for fast, heavily quantized phase SLMs. We show simulations of 4 bit phase quantization with varying numbers of time-multiplexed frames, showing the average PSNR over 14 example images. The projected gradient descent (GD) improves upon the naive method, which ignores quantization. Surrogate gradient (SG) methods replace the gradients of the non-differentiable quantization operator in the backpropagation pass using either a sigmoid or a Gumbel-Softmax (GS) function. The latter is found to outperform other approaches by a large margin, especially with faster SLMs. Remarkably, our framework using only 4 bit precision with 8 time-multiplexed frames even outperforms a conventional 8 bit phase SLM without time multiplexing (red dashed line).

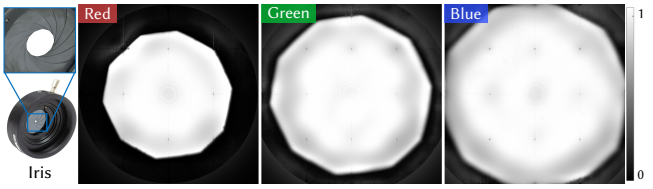

Fig. 4. Learned optical filters for three channels, corresponding to the amplitude distribution on the Fourier plane $a_{\mathcal{F}}$ that is indicated in Sec. 3.1 and Table 1. On the left we show the photograph of the physical iris used in the system acting as the optical filter. Our model accurately learns the shape of the physical iris and, as expected, its diameter in the learned model varies accordingly to wavelength.

Comparing CGH Algorithms. We compare several CGH approaches for the task of optimizing phase patterns for a fast phase SLM with 4 bits, or 16 phase levels, in Fig. 3. The naive approach, which quantizes the phase after optimization performs poorly, as measured by the peak signal-to-noise ratio (PSNR). The projected gradient descent approach performs better and shows improvements with an increasing SLM speed. The surrogate gradient (SG) method used with the gradients of sigmoid and those of the Gumbel-Softmax are significantly better than other methods, with Gumbel-Softmax outperforming all other methods by a large margin, especially for higher-speed SLMs. This experiment represents the TI SLM with 4 bits and up to 480 Hz color, i.e., 8 multiplexed frames each running at 60 Hz so a total of 480 Hz. We evaluate other bit depths in the supplement and show similar trends. Finally, Gumbel-Softmax can be used as part of an SG method (Eq. 5) using only its gradients $\frac{\partial \widehat{q}}{\partial \phi}$ or it can be used to replace q by $\widehat{q}$ also in the forward image formation. We found the former performs better in most settings, and therefore only report these results in the paper; see the supplement for evaluations of the latter approach.

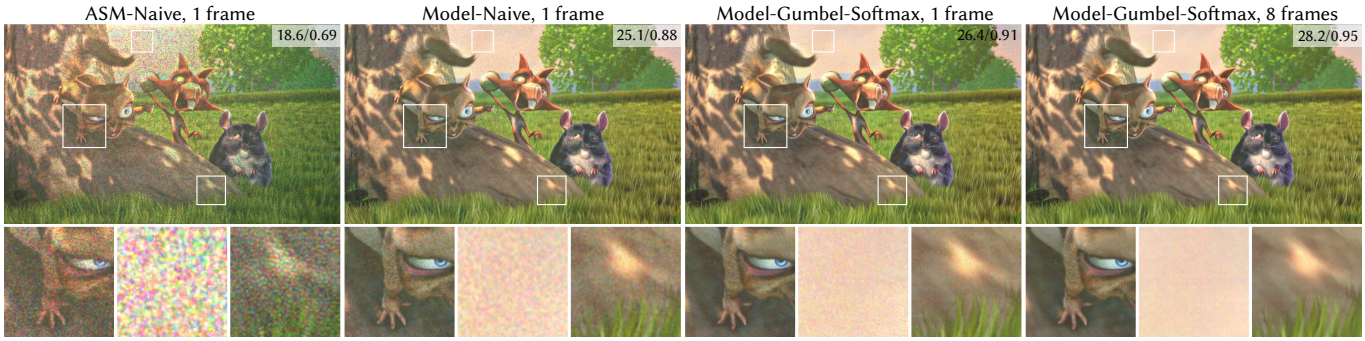

Fig. 5. Comparison of 2D CGH algorithms using experimentally captured data. Here, we compare SGD algorithms using the ASM w/ Naive (1st column), Model w/ Naive (2nd column), and Model w/ GS without time multiplexing (3rd column) and with 8 multiplexed frames (4th column). Our calibrated wave propagation model and Gumbel-Softmax quantization layer result in sharper images with higher contrast and less speckle than others under the same experimental conditions. Quantitative evaluations are included as PSNR/SSIM.

Learning Physical Filters. We visualize in Figure 4 the performance of our learned model in accurately approximating the optical filter, which is an iris in the physical display system. As expected, values outside the filters are all zeros. The shape of blade edges is robustly learned with our model and scales with wavelength as expected. The variance of diameter size also aligns with the variance of wavelength. Refer to Figure S7 in the supplement for visualization of the full model.

Assessing 2D Holography. We present in Figure 5 experimental results of 2D holographic display assessing different CGH algorithms and different multiplexing schemes. In this experiment, we compare SGD algorithms using the ASM with Naive quantization, our model with Naive quantization, and our model with Gumbel-Softmax (GS). We observe two insights. First, the use of our calibrated wave propagation model corrects for most artifacts present in the physical display. Second, applying the GS operation leads to better performance in such heavily-quantized optimization problems. Refer also to Figures S8–9, as well as Tables S1 and S2 in the supplementary document for both quantitative and qualitative assessments of other examples.

Assessing 3D Holography. We present in Figure 6 experimental results of 3D holographic display assessing different CGH algorithms. In this experiment, we compare SGD algorithms with the prior state-of-the-art NH3D model and Naive quantization using RGBD input [Choi et al. 2021a] with 1 frame and 8 multiplexed frames, respectively, our model with Gumbel-Softmax (GS), and our model with GS using focal stack supervision. PSNR metrics are provided in the caption. Using only a single frame results in speckly in-focus content (shown with red squares in Figure 6). Even with multiple frames, RGBD supervision produces speckle in the unconstrained out-of-focus regions. However, with our focal stack supervision and time multiplexing, we observe natural out-of-focus blur, while still preserving sharpness for the in-focus content. For example, the branch at the intermediate depth is sharp, and the sky in the background is smooth. In the supplement, we show extensive evaluations and ablations of 3D multiplane CGH methods for more 3D scenes (Figures S3–4 and S10–16).

Assessing 4D Light Field Holography. We present in Figure 7 experimental results of 4D light field–supervised holographic display, assessing different CGH algorithms. In this experiment, we compare the OLAS [Padmanaban et al. 2019] algorithm, our approach using light field–supervision with the ASM and naive quantization (ASM-Naive), and our approach with the camera-calibrated wave propagation model and Gumbel-Softmax (Model-GS) to account for the low bit depth of the SLM. The OLAS algorithm requires light field and depth maps for each light field view as input and it does not support time multiplexing. Both variants of our method do not require depth maps and jointly optimize 8 time-multiplexed frames using SGD. For each example scene, we show close-ups of content at two distances (far, near). We observe that our framework exhibits the best image quality for both in-focus (red squares) and out-of-focus regions (white squares). Refer also to Figures S5 and S17 in the supplementary document for additional simulation and experimental results.

5 DISCUSSION

In summary, we present a new framework for computer-generated holography. This framework includes a camera-calibrated wave propagation model that combines parts of the recently proposed model in a novel way to achieve a better performance with fewer model parameters. We explore surrogate gradient methods for optimizing the heavily quantized SLM patterns of emerging MEMS-based phase SLMs and show the Gumbel-Softmax algorithm to outperform other approaches. Our framework is flexible in supporting 2D, 2.5D, 3D, and 4D supervision at runtime and we show state-of-the-art results in all of these scenarios with our near-eye holographic display prototypes.

Limitations and Future Work. Image quality could be further improved by increasing the precision and framerate of the employed phase SLMs and, importantly, by improving their diffraction efficiency. In Figure S6 of our supplement, we explore the simulated image quality with varying levels of time multiplexing and bit depth, but analytically deriving this landscape remains an interesting direction for future work to explore. Our algorithms do not run in real

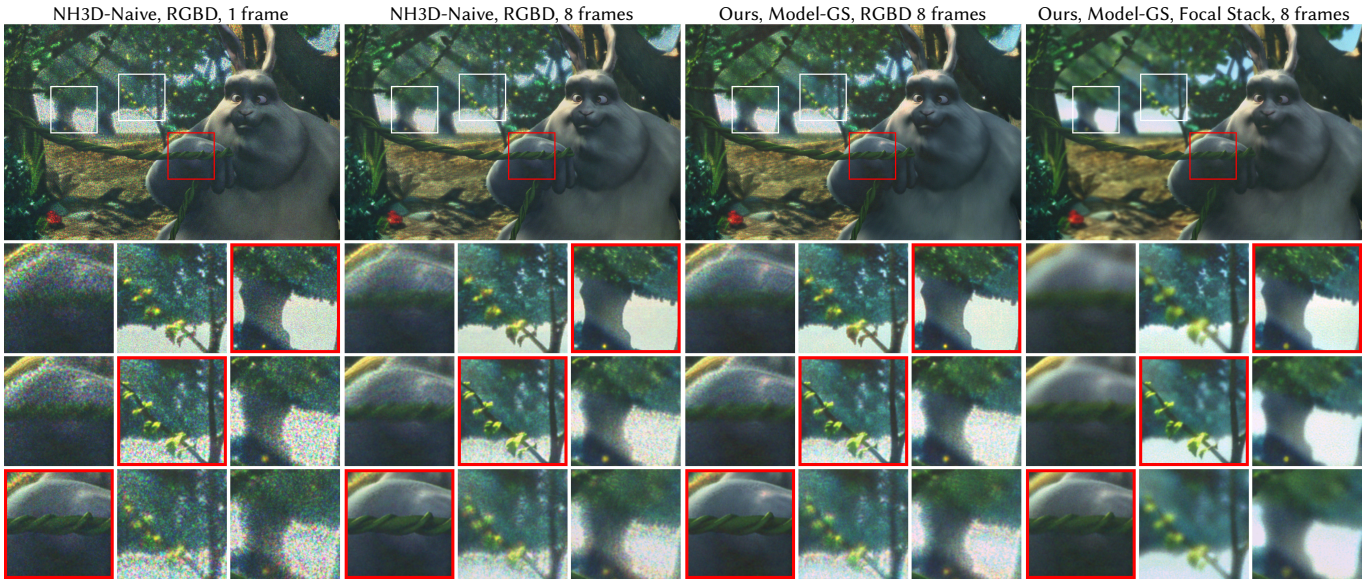

Fig. 6. Comparison of 3D CGH algorithms using experimentally captured data. Here, we compare SGD algorithms with the prior state-of-the-art NH3D model and Naive quantization using RGBD input [Choi et al. 2021a] with 1 frame and 8 multiplexed frames, respectively, our model with Gumbel-Softmax (GS), and our model with GS using focal stack supervision. The corresponding PSNR metrics are 24.3 dB, 25.8 dB, and 26.7 dB with respect to the RGBD all-in-focus targets (left 3 columns), and 26.9 dB with respect to the focal stack (right column). For close-ups, red squares indicate where the camera is focused at three distances (from top to bottom: far, intermediate, and near).

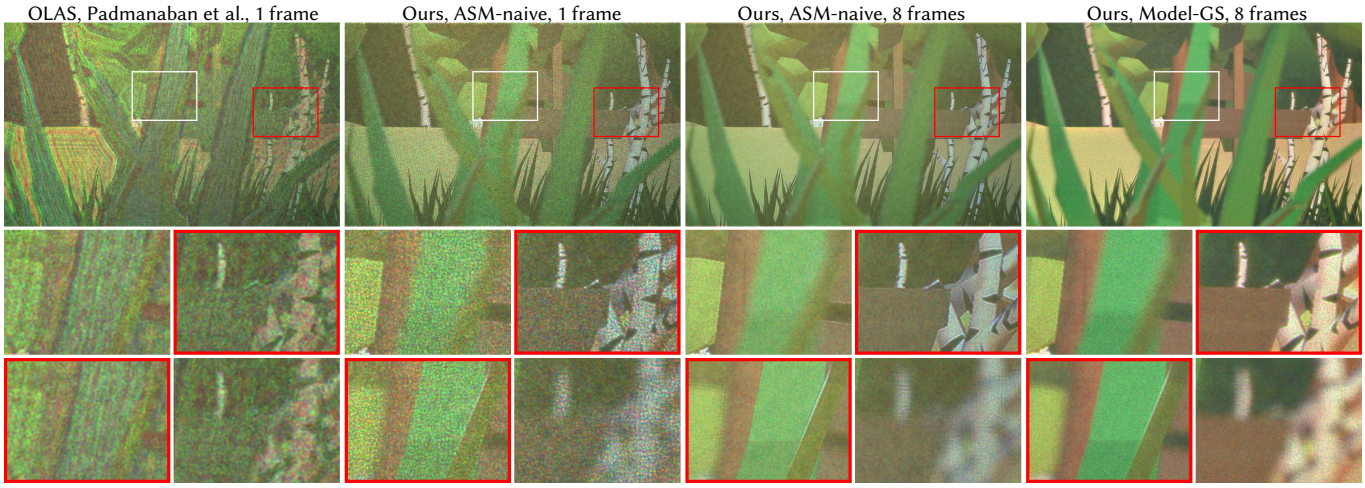

Fig. 7. Comparison of 4D light field–supervised CGH algorithms using experimentally captured data. Here, we compare the OLAS algorithm [Padmanaban et al. 2019] (1st column) without time multiplexing, and three variants of our approach: ASM-Naive without time multiplexing (2nd column) and with 8 multiplexed frames (3rd column) and Model-GS with 8 multiplexed frames (4th column). For close-ups, red squares indicate where the camera is focused at two distances (top: far, bottom: near). Since OLAS deterministically computes a single phase pattern for a target light field, there would be no variation between time-multiplexed frames.

time, but require on the order of tens of seconds to a few minutes to compute a hologram. Neural networks could be employed to speed up the computation, as recently demonstrated by Horisaki et al. [2018], Peng et al. [2020], and Shi et al. [2021]. Due to their limited space–bandwidth product, holographic near-eye displays only provide a limited eye box, which could be addressed by dynamically steering it using eye tracking [Jang et al. 2017]. The depth of field of 3D-supervised holograms in AR scenarios should match that of the user's eye, which requires tracking their pupil diameter. Finally, we demonstrated our results on benchtop prototype displays, which will have to be miniaturized into the impressive device form factors presented by Maimone et al. [2017] and Wang and Maimone [2020].

Conclusion. The algorithmic advances presented in this work help make holographic near-eye displays a practical technology for next-generation VR/AR systems.

ACKNOWLEDGMENTS

We thank Cindy Nguyen for helpful discussions. This project was in part supported by a Kwanjeong Scholarship, a Stanford SGF, Intel, NSF (award 1839974), a PECASE by the ARO (W911NF-19-1-0120), and Sony.

REFERENCES

Terry A. Bartlett, William C. McDonald, and James N. Hall. 2019. Adapting Texas Instruments DLP technology to demonstrate a phase spatial light modulator. In *SPIE OPTO, Proceedings Volume 10932, Emerging Digital Micromirror Device Based Systems and Applications XI*. 109320S.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. 2013. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432* (2013).

Stephen A. Benton. 1983. Survey Of Holographic Stereograms. In *Proc. SPIE*, Vol. 0367.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. 2019. Wirtinger Holography for Near-eye Displays. *ACM Trans. Graph.* 38, 6 (2019).

Praneeth Chakravarthula, Ethan Tseng, Tarun Srivastava, Henry Fuchs, and Felix Heide. 2020. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Trans. on Graph. (TOG)* 39, 6 (2020), 1–18.

Chenliang Chang, Kiseung Bang, Gordon Wetzstein, Byoungho Lee, and Liang Gao. 2020. Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective. *Optica* 7, 11 (2020), 1563–1578.

Chun Chen, Byounghyo Lee, Nan-Nan Li, Minseok Chae, Di Wang, Qiong-Hua Wang, and Byoungho Lee. 2021. Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function. *Opt. Express* 29, 10 (2021), 15089–15103.

Jhen-Si Chen and Daping Chu. 2015. Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications. *Opt. Express* 23, 14 (2015), 18143–18155.

Rick H-Y Chen and Timothy D Wilkinson. 2009. Computer generated hologram with geometric occlusion using GPU-accelerated depth buffer rasterization for three-dimensional display. *Applied optics* 48, 21 (2009), 4246–4255.

Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, and Gordon Wetzstein. 2021a. Neural 3D Holography: Learning Accurate Wave Propagation Models for 3D Holographic Virtual and Augmented Reality Displays. *ACM Trans. Graph. (SIGGRAPH Asia)* (2021).

Suyeon Choi, Jonghyun Kim, Yifan Peng, and Gordon Wetzstein. 2021b. Optimizing image quality for holographic near-eye displays with michelson holography. *Optica* 8, 2 (2021), 143–146.

Junyoung Chung, Sungjin Ahn, and Yoshua Bengio. 2016. Hierarchical multiscale recurrent neural networks. *arXiv preprint arXiv:1609.01704* (2016).

James R Fienup. 1982. Phase retrieval algorithms: a comparison. *Applied optics* 21, 15 (1982), 2758–2769.

Ralph W Gerchberg. 1972. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35 (1972), 237–246.

Joseph W. Goodman. 2014. Holography Viewed from the Perspective of the Light Field Camera. In *Fringe 2013*, Wolfgang Osten (Ed.). Springer Berlin Heidelberg, 3–15.

Ryoichi Horisaki, Yohei Nishizaki, Katsuhisa Kitaguchi, Mamoru Saito, and Jun Tanida. 2021. Three-dimensional deeply generated holography. *Appl. Opt.* 60, 4 (2021), A323–A328.

Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. 2018. Deep-learning-generated holography. *Applied optics* 57, 14 (2018), 3859–3863.

Chung-Kai Hsueh and Alexander A. Sawchuk. 1978. Computer-generated double-phase holograms. *Applied optics* 17, 24 (1978), 3874–3883.

Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoungho Lee. 2017. Retinal 3D: augmented reality near-eye display via pupil-tracked light field projection on retina. *ACM Trans. Graph. (SIGGRAPH Asia)* 36, 6 (2017).

Eric Jang, Shixiang Gu, and Ben Poole. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144* (2016).

Bahram Javidi, Artur Carnicer, Arun Anand, George Barbastathis, Wen Chen, Pietro Ferraro, J. W. Goodman, Ryoichi Horisaki, Kedar Khare, Malgorzata Kujawinska, Rainer A. Leitgeb, Pierre Marquet, Takanori Nomura, Aydogan Ozcan, YongKeun Park, Giancarlo Pedrini, Pascal Picart, Joseph Rosen, Genaro Saavedra, Natan T. Shaked, Adrian Stern, Enrique Tajahuerce, Lei Tian, Gordon Wetzstein, and Masahiro Yamaguchi. 2021. Roadmap on digital holography. *Opt. Express* 29, 22 (2021).

Hoonjong Kang, Takeshi Yamaguchi, and Hiroshi Yoshikawa. 2008. Accurate phase-added stereogram to improve the coherent stereogram. *Appl. Opt.* 47, 19 (2008).

Koray Kavakli, Hakan Urey, and Kaan Akşit. 2022. Learned holographic light transport. *Appl. Opt.* 61, 5 (2022), B50–B55.

Remington S Ketchum and Pierre-Alexandre Blanche. 2021. Diffraction efficiency characteristics for MEMS-based phase-only spatial light modulator with nonlinear phase distribution. In *Photonics*, Vol. 8. Multidisciplinary Digital Publishing Institute, 62.

Dongyeon Kim, Seung-Woo Nam, Kiseung Bang, Byounghyo Lee, Seungjae Lee, Youngmo Jeong, Jong-Mo Seo, and Byoungho Lee. 2021. Vision-correcting holographic display: evaluation of aberration correcting hologram. *Biomed. Opt. Express* 12, 8 (2021), 5179–5195.

Jonghyun Kim, Manu Gopakumar, Suyeon Choi, Yifan Peng, Ward Lopes, and Gordon Wetzstein. 2022. Holographic glasses for virtual reality. In *Proceedings of the ACM SIGGRAPH*.

Byounghyo Lee, Dongyeon Kim, Seungjae Lee, Chun Chen, and Byoungho Lee. 2022. High-contrast, speckle-free, true 3D holography via binary CGH optimization. *arXiv preprint arXiv:2201.02619* (2022).

Wai Hon Lee. 1970. Sampled Fourier transform hologram generated by computer. *Applied Optics* 9, 3 (1970), 639–643.

Mark Lucente and Tinsley A Galyean. 1995. Rendering interactive holographic images. In *ACM SIGGRAPH*. 387–394.

Chris J Maddison, Andriy Mnih, and Yee Whye Teh. 2016. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712* (2016).

Andrew Maimone, Andreas Georgiou, and Joel S Kollin. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Trans. Graph. (SIGGRAPH)* 36, 4 (2017), 85.

Andrew Maimone and Junren Wang. 2020. Holographic Optics for Thin and Lightweight Virtual Reality. *ACM Trans. Graph. (SIGGRAPH)* 39, 4 (2020).

Kyoji Matsushima and Sumio Nakahara. 2009. Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method. *Applied optics* 48, 34 (2009), H54–H63.

Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. 2019. Holographic Near-eye Displays Based on Overlap-add Stereograms. *ACM Trans. Graph.* 38, 6 (2019).

Jae-Hyeung Park. 2017. Recent progress in computer-generated holography for three-dimensional scenes. *Journal of Information Display* 18, 1 (2017), 1–12.

Yifan Peng, Suyeon Choi, , Jonghyun Kim, and Gordon Wetzstein. 2021. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Science Advances* (2021).

Yifan Peng, Suyeon Choi, Nitish Padmanaban, and Gordon Wetzstein. 2020. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* 39, 6 (2020), 1–14.

Liang Shi, Fu-Chung Huang, Ward Lopes, Wojciech Matusik, and David Luebke. 2017. Near-eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics. *ACM Trans. Graph.* 36, 6 (2017).

Liang Shi, Beichen Li, Changil Kim, Petr Kellnhofer, and Wojciech Matusik. 2021. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* 591, 7849 (2021), 234–239.

Koki Wakunami, Hiroaki Yamashita, and Masahiro Yamaguchi. 2013. Occlusion culling for computer generated hologram based on ray-wavefront conversion. *Optics express* 21, 19 (2013), 21811–21822.

Fahri Yaras, Hoonjong Kang, and Levent Onural. 2010. State of the Art in Holographic Displays: A Survey. *Journal of Display Technology* 6, 10 (2010), 443–454.

Dongheon Yoo, Youngjin Jo, Seung-Woo Nam, Chun Chen, and Byoungho Lee. 2021. Optimization of computer-generated holograms featuring phase randomness control. *Opt. Lett.* 46, 19 (2021), 4769–4772.

Friedemann Zenke and Surya Ganguli. 2018. Superspike: Supervised learning in multilayer spiking neural networks. *Neural computation* 30, 6 (2018), 1514–1541.

Hao Zhang, Liangcai Cao, and Guofan Jin. 2017. Computer-generated hologram with occlusion effect using layer-based processing. *Applied optics* 56, 13 (2017).

Hao Zhang, Neil Collings, Jing Chen, Bill A Crossland, Daping Chu, and Jinghui Xie. 2011. Full parallax three-dimensional display with occlusion effect using computer generated hologram. *Optical Engineering* 50, 7 (2011), 074003.

Zhengyun Zhang and M. Levoy. 2009. Wigner distributions and how they relate to the light field. In *Proc. ICCP*. IEEE, 1–10.

Remo Ziegler, Simon Bucheli, Lukas Ahrenberg, Marcus Magnor, and Markus Gross. 2007. A Bidirectional Light Field-Hologram Transform. In *Computer Graphics Forum (Eurographics)*, Vol. 26. 435–446.

Time-multiplexed Neural Holography: A Flexible Framework for Holographic Near-eye Displays with Fast Heavily-quantized Spatial Light Modulators—Supplemental Material

SUYEON CHOI*, Stanford University, USA
MANU GOPAKUMAR*, Stanford University, USA
YIFAN PENG, Stanford University, USA
JONGHYUN KIM, NVIDIA and Stanford University, USA
MATTHEW O'TOOLE, Carnegie Mellon University, USA
GORDON WETZSTEIN, Stanford University, USA

This supplementary document includes implementation details of our holographic display prototype, complementary derivations related to wave propagation and optimization models, and additional experimental results. Refer also to the supplementary video for better visualization.

Here we list the abbreviations and notations used across this document. These are consistent with those in the main paper.

SLM:	a spatial light modulator
CGH:	computer-generated holography
STFT:	the Short-time Fourier transform (STFT)
ASM:	the angular spectrum method [Goodman 2005]
GS:	the Gumber-Softmax operation [Jang et al. 2016; Maddison et al. 2016]
CITL:	the camera-in-the-loop optimization technique [Peng et al. 2020]
SGD:	stochastic gradient descent phase retrieval [Peng et al. 2020]
NH:	a 2D wave propagation model that is trained using an SGD-based camera-in-the-loop training strategy aka neural holography [Peng et al. 2020]; once trained, this model is used to generate new holograms using an SGD solver
NH3D:	wave propagation model using CNNs operating on the complex-valued field at the SLM plane before ASM propagation and also directly after propagation to the target planes [Choi et al. 2021]

S1 ADDITIONAL DETAILS ON HARDWARE

In this section, we describe the hardware implementation of our benchtop 3D holographic display prototype. Figure S1 shows the system schematic and photograph of our implementation, including a display and a capture unit, that are connected under a closed-loop framework. Specifically, the SLM is TI DLP6750Q1EVM with a resolution of 1,280 × 800, a pixel pitch of 10.8 μm, and a bit depth of 4 bits per pixel. The laser is a FISBA RGBeam fiber-coupled module with three optically aligned laser diodes with a maximum output power of 50 mW. The measured wavelengths are 636.4, 517.7, and 440.8 nm. In our implementation, color images are captured as separate exposures for each wavelength and then cast in post-processing.

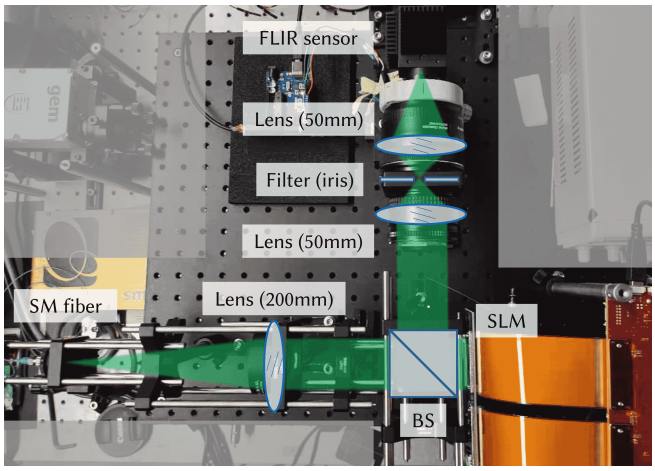

Fig. S1. Schematic and prototype photograph of our holographic display.

Other components including the collimating lenses, the relay imaging lenses, the filtering iris, and the beam splitter (Thorlabs BS016) are shown in Figure S1. All images are captured with a FLIR Grasshopper3 12.3 MP color USB3 sensor through a Canon EF 50mm lens. The Canon lens and sensor are synchronized in hardware via Arduino (Uno SMD) controller to enable programmable varifocal display and acquisition. The capture unit is assembled on a motorized translation stage to enable the acquisition capability from different horizontal viewpoints. In such a way, we are able to acquire holographic images to both form the training dataset and showcase diverse 3D cues.

In the calibration step, we use a similar procedure to that described in the relevant work [Peng et al. 2020] and apply a planar

*denotes equal contribution.

Authors' addresses: Suyeon Choi, suyeon@stanford.edu, Stanford University, USA; Manu Gopakumar, manugopa@stanford.edu, Stanford University, USA; Yifan Peng, evanpeng@stanford.edu, Stanford University, USA; Jonghyun Kim, jonghyunk@nvidia.com, NVIDIA and Stanford University, USA; Matthew O'Toole, mpotoole@cmu.edu, Carnegie Mellon University, USA; Gordon Wetzstein, gordon.wetzstein@stanford.edu, Stanford University, USA.

homography from the field of computer vision to accurately register the captured images to the ground-truth images. Our implementation uses a target binary pattern consisting of 18 × 11 white dots with the interval between the centers of neighboring two dots set 70 pixels. Accordingly, the region of interest has a resolution of 1,190 × 700 pixels.

S2 ADDITIONAL DETAILS ON SOFTWARE

S2.1 From a variant of projected gradient descent to surrogate gradient with unit Jacobian

Here, we derive the relationship between the variant of projected gradient descent in the manuscript and the gradient descent with the surrogate gradient of unit matrix Jacobian.

1) Based on the projected gradient descent rule described in the main paper (Eq. 4), we consider the projection step of the last iteration together with the phase update step of the current iteration as one iteration:

$$\phi^{(k-1)} \leftarrow \Pi_Q\left(\widehat{\phi}^{(k-1)}\right) = q\left(\widehat{\phi}^{(k-1)}\right).$$
$$\widehat{\phi}^{(k)} \leftarrow \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial \phi}\right)^T \mathcal{L}\left(s\left|f_{\text{CNN}}\left(e^{i\phi^{(k-1)}}\right)\right|, a_{\text{target}}\right).$$

2) Then, we substitute the first row into the second row:

$$\widehat{\phi}^{(k)} \leftarrow q\left(\widehat{\phi}^{(k-1)}\right) - \alpha\left(\frac{\partial \mathcal{L}}{\partial q}\right)^T \mathcal{L}\left(s\left|f_{\text{CNN}}\left(e^{iq\left(\widehat{\phi}^{(k-1)}\right)}\right)\right|, a_{\text{target}}\right).$$

Note that all of these variants (including the one in the manuscript) generally fail to work well for holographic phase retrieval, because the projection taken every step vanishes the phase update with the gradient term.

3) Thus, we relax this hard constraint by leaving out the projection in the first term; this variant of projected gradient descent applying the projection only in the second term does not suffer from being stuck. This is also a special case of surrogate gradient where we use the surrogate gradient of unit Jacobian $\frac{\partial \widehat{q}}{\partial \phi} = I$. This means the gradient with respect to the quantized phase q is simply passed to the gradient with respect to the continuous phase ϕ. We use the following algorithm as the representative of the projected gradient descent:

$$\widehat{\phi}^{(k)} \leftarrow \widehat{\phi}^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial q}\right)^T \mathcal{L}\left(s\left|f_{\text{CNN}}\left(e^{iq\left(\widehat{\phi}^{(k-1)}\right)}\right)\right|, a_{\text{target}}\right) \quad (1)$$
$$= \widehat{\phi}^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial q} \cdot \frac{\partial \widehat{q}}{\partial \phi}\right)^T \mathcal{L}\left(s\left|f_{\text{CNN}}\left(e^{iq\left(\widehat{\phi}^{(k-1)}\right)}\right)\right|, a_{\text{target}}\right). \quad (2)$$

Note that Eq. 2 is identical to Eq. 5 in the main paper. The recent paper by Lee et al. [2022] uses hard-sigmoid as a surrogate gradient which has unit gradient within the valid range, and we note that this falls into the category of variants of gradients we describe in this section and our Gumbel-Softmax based approach outperforms it with a large margin as shown in Fig. 3.

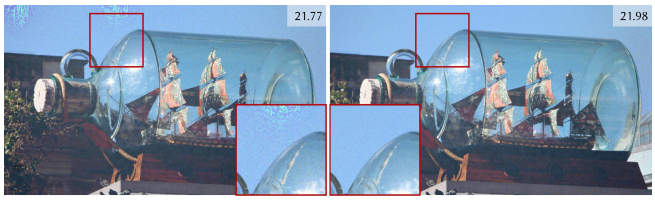

Fig. S2. Experimentally captured results with the camera-in-the-loop calibration using the naive (left) and the surrogate gradient (right) quantization. PSNR metrics are indicated.

S2.2 Camera-in-the-loop with highly quantized SLMs

Here we describe a more accurate camera-in-the-loop procedure for highly quantized SLMs. The camera-in-the-loop procedure proposed by Peng et al. [2020] approximates the gradient of the physical forward model f to that of simulated model $\widehat{f}$:

$$\phi^{(k)} \leftarrow \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial \phi}\right)^T$$
$$\simeq \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial f} \cdot \frac{\partial \widehat{f}}{\partial \phi}\right)^T. \quad (3)$$

However, note that we always have to quantize the phase before displaying it on the SLM. Thus, technically, f should reads as $f(q(\phi))$. Again, while we do not have access to the gradient of the quantization function q, we can approximate it with the surrogate gradient $\frac{\partial \widehat{q}}{\partial \phi}$:

$$\phi^{(k)} \leftarrow \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial f(q)} \cdot \frac{\partial \widehat{f}(q)}{\partial \phi}\right)^T$$
$$= \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial f} \cdot \frac{\partial \widehat{f}}{\partial q} \cdot \frac{\partial q}{\partial \phi}\right)^T$$
$$\simeq \phi^{(k-1)} - \alpha\left(\frac{\partial \mathcal{L}}{\partial f} \cdot \frac{\partial \widehat{f}}{\partial q} \cdot \frac{\partial \widehat{q}}{\partial \phi}\right)^T. \quad (4)$$

In Fig. S2, we compare two update rules Eq. (3) and Eq. (4). We see that with the approximation with the surrogate gradient, image quality is noticeably improved.

S2.3 Setting parameters for quantized phase optimization

In this subsection we describe parameters used in quantized phase optimization. We run 2,000 iterations with early stopping with a learning rate of 0.01 for 1 frame and that of 0.02 for 8 frames. We note that the surrogate gradients, if used with the Sigmoid or functions with clipping, require a higher learning rate to avoid getting stuck in local minima which leads to poor performance. During optimization, gradually annealing the slope helps the optimization by better approximating the step function gradually while allowing exploration of a large parameter space with a lower slope at the beginning. The Sigmoid function can be annealed with a parameter s multiplied with the input x, so we refer to the Sigmoid function

as $\sigma(s \cdot x)$. The Gumbel-Softmax can be annealed with three parameters including the temperature parameter of Softmax τ, the width parameter w which corresponds to the interval of discrete levels, and a scale multiplied to the score function in Eq. 8 of the manuscript. In the experiments, we tuned w considering the number of phase levels (interval between neighbour phase displacements) and the scale multiplied to the score function was increased from 300 to 1,000 during optimization. We used an annealing schedule of $\tau = \tau_0 \cdot e^{-c \cdot (t/t_{\max})}$ at iteration t with $c \sim \ln 2$ and $\tau_0 \sim 4$.

S2.4 Model architecture and training details

The two convolutional neural networks in our model are based on the U-net architecture as in Choi et al. [2021]. We made a slight modification on the CNN archtectures such that each CNN has 5 layers and 4 input channels of amplitude, phase, real, and imaginary values of the input field. The output of CNN_{SLM} is two channels that are used as real and imaginary values of the adjusted SLM field. The output of $\text{CNN}_{\text{target}}$ is 1 channel that is used as a corrected amplitude. As stated in Table 1 of the main paper, the model is trained over 6 intensity planes, corresponding to 0.0 D, 0.5 D, 1.0 D, 1.5 D, 2.5 D, and 3.0 D in the physical space. The propagation distances from the SLM are 7.9, 8.1, 8.25, 8.4, 8.6, 8.8, 9.1 cm and the held-out plane is set to 8.6 cm. We use a batch size of 2, and a learning rate of $4e^{-4}$. We note that the variety of phasemaps are important for model training. For example, we note that a dataset mainly generated using the SGD algorithm usually consists of holographic images that have very narrow angular spectrum. Thus, we generate the dataset with the STFT-based regularizer we present in Eq. S7. In addition, we generate phasemaps with a set of random parameters, including learning rates, initial phase distribution, and propagation distances. We generate 3,000 phases for each channel and capture the intensity at 7 target planes. Other than the held-out plane, the dataset is divided into training, validation, and test sets with a ratio of 8:1:1. The training takes around 24 hours to converge. To parameterize and train a look up table for phase mapping, the phase maps are first one-hot encoded, multiplied with the parameterized lookup table, and then summed up per pixel before passing through the full forward model pipeline.

S2.5 Natural defocus blur with 2.5D supervision on Quantized SLMs

The 2.5D supervision results in our paper are generated using the multiplane loss function presented by Choi et al. [2021]. For this approach, the depth map D from an RGBD input is first decomposed into a set of binary masks $m^{\{j\}}$ corresponding to a set of distances $z^{\{j\}}$ from the SLM using closest distance matching,

$$m^{(j)}(x,y) = \begin{cases} 1, & \text{if } |z^{(j)} - D(x,y)| < |z^{(k)} - D(x,y)|, \forall k \neq j, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

These binary masks are then used to constrain a multiplane loss that pushes the wavefront to reconstruct the desired RGB amplitude, a_{target}, at the corresponding in-focus distances from the SLM

$$\mathcal{L}_{\text{2.5D}} = \frac{1}{J}\sum_{j=1}^{J} \mathcal{L}\left(m^{(j)} \circ s \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left| f_{\text{model}}\left(e^{iq(\phi^{(t)})}, z^{(j)}\right)\right|^2}, \; m^{(j)} \circ a_{\text{target}} \right), \tag{6}$$

where $\circ$ is element-wise multiplication. One challenge with this approach is that it leaves the out-of-focus parts of the displayed intensities volume unconstrained, but this can be addressed using additional smooth phase regularization strategies as discussed further by Choi et al. [2021]. However, the ADMM technique for smooth phase proposed by this work can only produce very slight blur, and cannot effectively be adapted to quantization.

Alternatively, with time multiplexing, some prior works including [Yoo et al. 2021] have proposed phase randomness approaches that can produce a much shallower depth of field. These techniques aim to randomly send light in differerent directions from each scene point. Over many frames, this results in scene points that diffusely send light in all directions. Unfortunately, this technique struggles with producing good image quality in the presence of quantization because it independently optimizes frames. Quantization also adds artifacts to the out-of-focus blur with this technique. To overcome these quantization artifacts and reduce the number of frames needed for smooth out-of-focus blur, an additional STFT-based loss can be applied to the in-focus content.

$$\mathcal{L}_{\text{STFT}} = \frac{1}{J}\sum_{j=1}^{J} \overline{\sigma_\theta^2} \left(m^{(j)} \circ s \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left| \text{STFT}\left(f_{\text{model}}\left(e^{iq(\phi^{(t)})}, z\right)\right)\right|^2} \right), \tag{7}$$

where $\overline{\sigma_\theta^2}$ is the variance of the STFT over angles averaged over the spatial locations across the wavefront. This loss pushes the output of the holographic display to emit light evenly in all directions from the in-focus points. This mimics the diffuse behavior of most natural coherent scenes. As demonstrated in Fig. S3, this enables natural blur with 2.5D supervision on quantized SLMs.

S2.6 Time-multiplexing for 3D supervision

Our 3D focal stack supervision technique enables very high image quality with natural defocus effects. This technique relies on the time multiplexing in order to reproduce the defocus effects as illustrated in Fig. S4. Some prior works such as Shi et al. [2021] have used similar focal stack supervision with a single frame but that is only possible with much less blur. This blur is produced by a low frequency coherent wavefront and cannot match the natural blur produced by a scene sending light in all directions.

S2.7 Time-multiplexing for 4D supervision

Our approach can uniquely use a holographic display to reproduce a full set of light field views. Prior holographic stereogram works did not account for how interference attenuates and amplifies different rays after converting a light field into a hologram. With the recently proposed overlap-add stereogram (OLAS) method [Padmanaban et al. 2019], this interference results in most rays outside of the

Fig. S3. Simulated evaluation of different RGBD supervised techniques for generating smooth defocus blur on ideal quantized SLMs. From left to right: Model ADMM from Choi et al. [Choi et al. 2021], Randomness Prior from Yoo et al. [Yoo et al. 2021], our Randomness Prior Jointly-optimized implementation, and our STFT RGBD implementation. Focused regions are highlighted with red boxes, that from top to bottom, indicate far, center, and near distances. The ADMM phase smoothness technique from Choi et al. [2021] produces smooth but very small blur and has strong artifacts because it cannot be adapted to quantization effectively. The phase randomness technique with individually optimized frames produces more substantial blur but has poor in-focus image quality because it uses individually optimized frames. The phase randomness technique with jointly optimized frames improves on this in-focus image quality, and both of these phase randomness techniques suffer from out-of-focus artifacts on the quantized SLMs. Adding the STFT-based loss function greatly improves the out-of-focus blur and has high image quality.

central light field view being heavily attenuated by destructive interference. Additionally, smooth content in the central view gets amplified by constructive interference. Along with modeling this interference, time multiplexing is needed to accurately reproduce a set of light field views that fully cover the diffraction angle of the SLM. Without time multiplexing, a single coherent wavefront produced with phase modulation of the SLM's resolution will not have the degrees of freedom to produce arbitrary light field views that could naturally occur. The phenomena discussed here is illustrated in Fig. S5.

S3 ADDITIONAL EXPERIMENTAL RESULTS

In this section, we present extra simulated and experimental results of our 3D holographic display prototype.

Exploring the trade-off between number of bits and frames. We explore the trade-off space between the number of frames and the number of bits an SLM supports. We optimize phase patterns for 14 target images using the ASM model for the green channel using 5 different methods, including the Naive approach that quantizes only at the end, a variant of the projected gradient descent approach that we elaborate in Sec. S2.1, the Surrogate gradients approach with

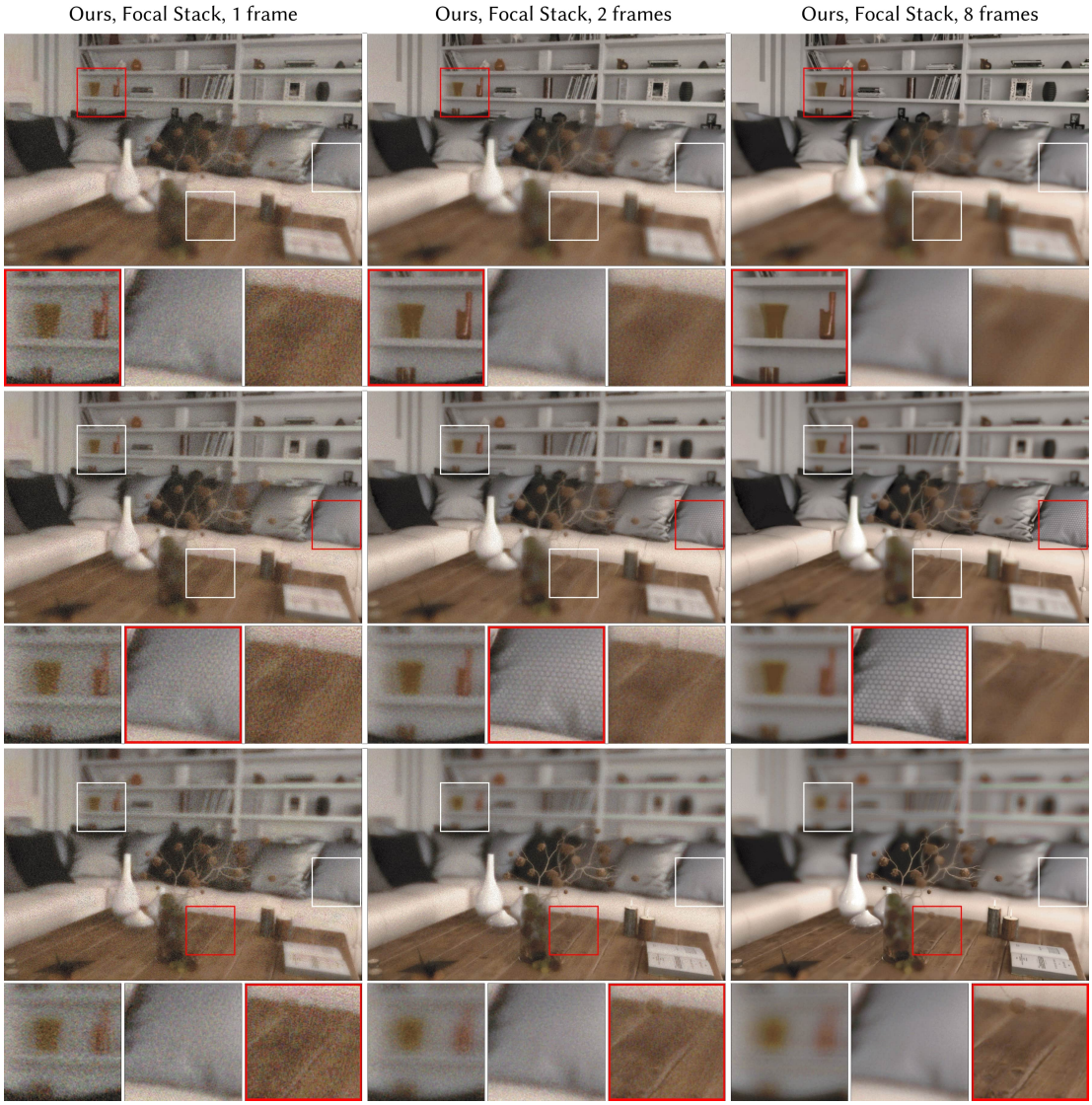

Fig. S4. Simulated evaluation of focal stack supervision with 1, 2, and 8 frames from left to right on ideal continuous SLMs. Focused regions are highlighted with red boxes, that from top to bottom, indicate far, center, and near distances. Even without quantization on an ideal SLM, the supervision with only 1 or 2 frames is overconstrained by the focal stack and cannot fully reproduce the desired natural defocus blur.

Sigmoid gradient, the Surrogate gradients with Gumbel-Softmax gradient, and the Gumbel-Softmax approach. We show averaged PSNR metrics in colormaps in Fig. S6.

Overall, the projected gradient descent approach improves upon the naive method, and notably, the surrogate gradient method with the Gumbel-Softmax gradient outperforms other methods by a large margin. We also denote a line for each method that roughly matches 50 dB performance that the 8 bit–1 frame type SLM achieves. Accordingly, we observe the trend that advanced algorithms shift the line towards the bottom left, which means it can effectively save the number of bits and frames without sacrificing performance. Note that the last approach does not quantize during the optimization but only replaces the forward model with our forward model described in Eq. 6 in the manuscript. This is analogous to the Naive approach for the Surrogate gradients approach with Gumbel-Softmax. We note that this continuous relaxation is beneficial especially in more constrained cases.

Full model visualization. Figure S7 visualizes our calibrated model, that includes learned intensity on SLM plane, learned phase on SLM plane, learned amplitude of the optical filter on Fourier plane, learned phase of the optical filter on Fourier plane, and learned lookup table for phase mapping. Note that this visualization includes many interesting aspects that present in the setup. First, the amplitude at SLM plane a_{SLM} reveals the envelope of the incident

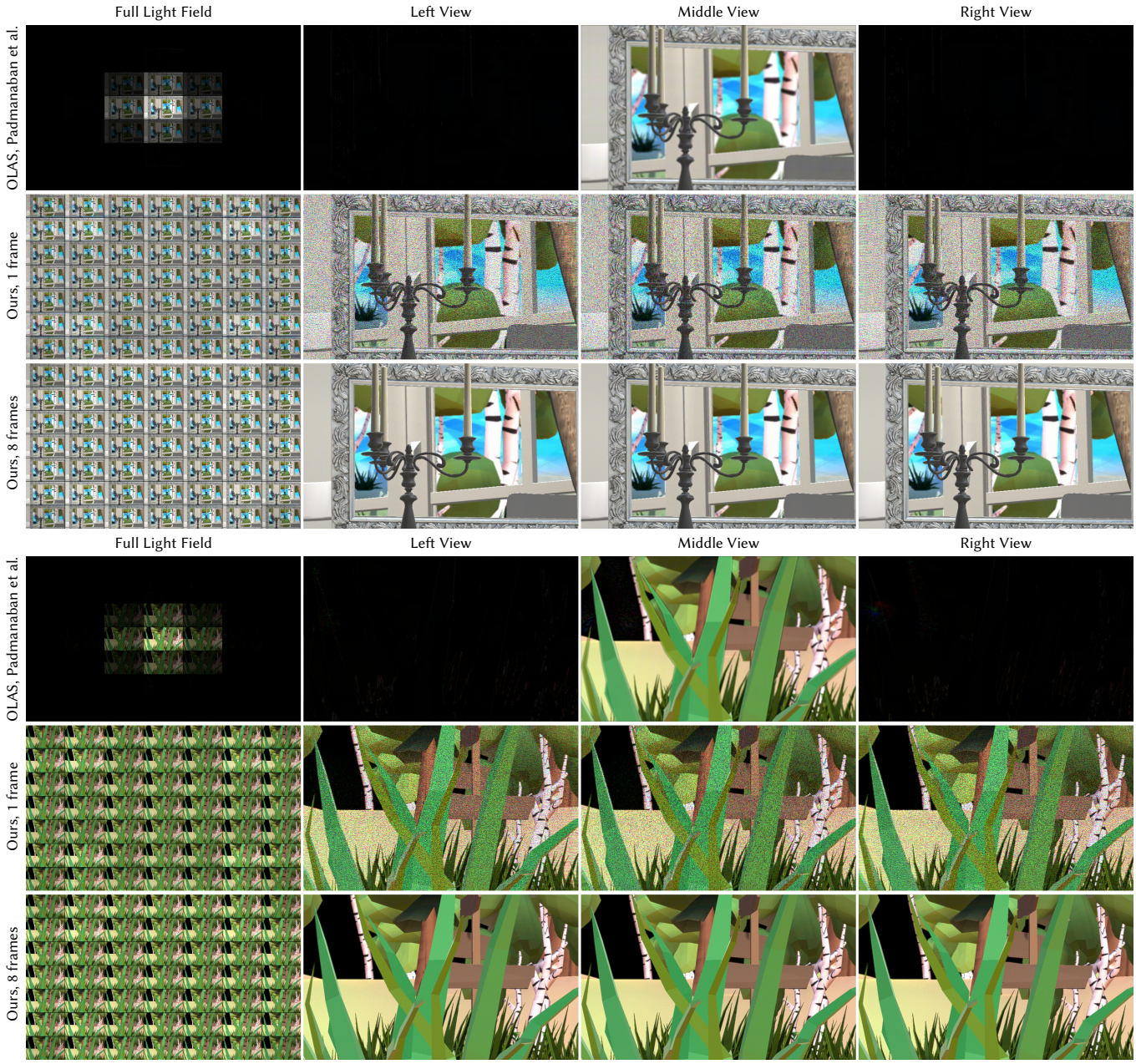

Fig. S5. Simulated evaluation of different light field to hologram techniques on ideal continuous SLMs. In the first column we have the full set of reproduced light field views. From the second column to the fourth column, we present the comparison of selected views. (Top) OLAS by Padmanaban et al. [2019] which does not account for the interference of rays has heavily amplified smooth content in the central view and heavily attenuated rays in other views. (Middle) Even without quantization on an ideal SLM, our proposed light field supervision technique with a single frame better covers the light field views but lacks the degrees of freedom to reproduce all the rays across all the light field views. (Bottom) Our proposed technique jointly over 8 frames has the degrees of freedom to fully reproduce the light field views.

beam as well as ripples and rings that occur in the physical display system. The phase at SLM plane ϕ_{SLM} shows the phase distortion. The terms at Fourier plane learn the shape of the physical filter we use in the setup and especially phase term $\phi_{\mathcal{F}}$ learns a radial phase ramp, and we note that potential propagation distance error can be learned through this parameter.

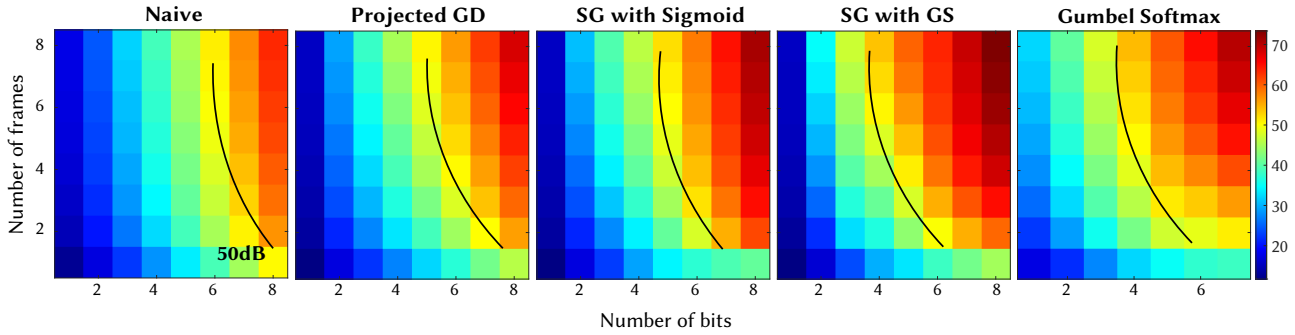

Fig. S6. Trade-off between the number of frames and bits for quantized SLMs, using different optimization algorithms. We optimize SLM phase patterns with a different number of bits and frames for 14 target images. We simulate the setup using the ASM model for the green channel using 5 different methods, including the Naive approach that quantizes only at the end, a variant of the projected gradient descent approach that we elaborate in Sec. S2.1, the Surrogate gradients approach with Sigmoid gradient, the Surrogate gradients with Gumbel-Softmax gradient, and the Gumbel-Softmax approach. We show averaged PSNR metrics as colormaps. In addition, we mark roughly where it reaches 50 dB PSNR as a black line, which is achieved by the 8 bit–1 frame type SLM. Note that the trend of black lines shifts.

Additional 2D results. Figure S8 and Figure S9 present full resolution holographic images and their close-ups reconstructed with different CGH algorithms, including the ASM with naive quantization, the ASM with Gumbel-Softmax quantization, our Model with naive quantization, and our Model with Gumbel-Softmax quantization. For each algorithm, we assess with displaying one single frame on the SLM as well as multiplexing 8 frames (which is jointly optimized). Note that the benefits of Gumbel-Softmax are also less prominent in the ASM case when the image quality degradation is dominated by the model mismatch, but even then the quantitative evaluations indicate improved performance. Thus, the benefits of our quantization techniques are most significant when the model mismatch is mitigated using the learned model (See columns 5-8). Corresponding quantitative evaluation is presented in Table 1 and Table 2, indicating results with 8 multiplexed frames and 1 frame, respectively. PSNR and SSIM metrics are listed.

Additional 3D results. Figure S10 and Figure S11 further present comprehensive simulation results of focal stacks and their close-ups reconstructed with different CGH algorithms, including the AADPM from Shi et al. [Shi et al. 2021], the Model ADMM from Choi et al. [Choi et al. 2021], the SGD-RGBD from Choi et al. [Choi et al. 2021], the randomness control from Yoo et al. [Yoo et al. 2021], our STFT RGBD implementation, and our focal stack implementation. All of these holograms used are with quantization operations. For each algorithm, reconstructed images with the camera focus at three different distances (far, center, near) are shown. We observe that ours outperform the others in preserving sharp content for in-focus regions while providing more natural blur for out-of-focus regions, with the focal stack implementation on the right being the best. Accordingly, we experimentally captured results of the scene in Figure S10 optimized with native quantization and Gumbel-Softmax (GS) quantization, as shown in Figure S12 and Figure S13.

Figure S14 and Figure S15 show additional experimental results of 3D holographic display assessing different CGH algorithms (complimentary to Figure 6 in the main paper). In this experiment, we compare algorithms of the SGD-NH3D using RGBD input [Choi et al. 2021] with 1 frame and 8 multiplexed frames, respectively, SGD-ours using RGBD input without and with Gumbel-Softmax (GS), and SGD-ours using Focal Stack with GS. Quantitative assessments are provided as PSNR metrics in the caption, as well as summarized in Table 3. We also show the behaviour of the interpolation between supervised planes in Fig. S16.

Additional 4D results. Figure S17 presents experimental results of light field reconstructed with different CGH algorithms, including the ASM-Naive with 1 single frame, our ASM-GS with 1 single frame, the ASM-Naive with 8 multiplexed frames, and our ASM-GS with 8 multiplexed frames. For each example scene, we show close-ups of content at three distances (far, intermediate, near). Our framework leads to overall higher image fidelity for both the in-focus and out-of-focus regions.

Robustness to possible viewpoint Shifts. Figure S18 presents a set of captured results of a holographic scene that validates the robustness of our image synthesis to possible viewpoint shifts. The camera is manually translated in horizontal from left to right, for a few millimeters. We observe no noticeable degradation in image quality over the viewpoint shifts.

Table 1. PSNR and SSIM metrics of captured 2D results with 8 multiplexed frames. Among all the methods, the proposed model, in tandem with the Gumbel-Softmax (GS) quantization, achieves the highest PSNR and SSIM. Images assessed are shown in Figure S8 (index 1 to 5) and Figure S9 (index 6 to 10) .

	Methods (algorithm-propagation operator)			
	SGD-ASM	SGD-ASM-GS	SGD-ours	SGD-ours-GS
# 1	20.29 / 0.821	20.61 / 0.829	27.41 / 0.947	28.22 / 0.954
# 2	17.24 / 0.796	17.52 / 0.807	22.39 / 0.909	23.00 / 0.916
# 3	20.33 / 0.655	20.68 / 0.663	25.67 / 0.803	26.14 / 0.811
# 4	18.43 / 0.632	18.63 / 0.643	22.35 / 0.815	22.70 / 0.811
# 5	16.57 / 0.508	16.52 / 0.486	19.78 / 0.731	19.90 / 0.718
# 6	18.26 / 0.789	18.48 / 0.799	23.33 / 0.911	23.82 / 0.915
# 7	15.16 / 0.066	15.08 / 0.063	16.82 / 0.088	16.74/ 0.088
# 8	18.09 / 0.654	18.15 / 0.643	21.23 / 0.764	21.24 / 0.758
# 9	18.54 / 0.748	18.86 / 0.751	22.97 / 0.934	23.31 / 0.832
# 10	18.62 / 0.820	18.83 / 0.827	23.09 / 0.889	23.50 / 0.898
Avg.	18.16 / 0.649	18.33 / 0.652	22.51/ 0.769	22.85 / 0.770

Table 2. PSNR and SSIM metrics of captured 2D results with 1 single frame. Among all the methods, the proposed model, in tandem with the Gumbel-Softmax (GS) quantization, achieves the highest PSNR and SSIM. Images assessed are shown in Figure S8 (index 1 to 5) and Figure S9 (index 6 to 10) .

	Methods (algorithm-propagation operator)			
	SGD-ASM	SGD-ASM-GS	SGD-ours	SGD-ours-GS
# 1	18.63 / 0.686	18.85 / 0.703	25.14 / 0.882	26.42 / 0.910
# 2	16.31 / 0.722	16.30 / 0.721	20.89 / 0.864	22.30 / 0.888
# 3	18.80 / 0.511	18.93 / 0.512	24.09 / 0.720	24.72 / 0.744
# 4	17.31 / 0.480	17.38 / 0.481	21.26 / 0.709	21.75 / 0.715
# 5	15.46 / 0.391	15.54 / 0.394	19.03 / 0.673	19.40 / 0.676
# 6	17.48 / 0.719	17.53 / 0.724	22.33 / 0.875	22.96 / 0.889
# 7	14.87 / 0.057	14.53 / 0.052	16.68 / 0.082	16.53 / 0.080
# 8	16.88 / 0.555	17.01 / 0.563	20.32 / 0.709	20.57 / 0.713
# 9	17.55 / 0.632	17.57 / 0.642	21.75 / 0.771	22.27 / 0.783
# 10	17.73 / 0.758	17.71 / 0.762	21.90 / 0.858	22.57 / 0.871
Avg.	17.10 / 0.551	17.13 / 0.556	21.34 / 0.714	21.95 / 0.727

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 Trans. Graph. (SIGGRAPH Asia)* (2021).

Joseph W Goodman. 2005. *Introduction to Fourier optics.* Roberts and Company.

Eric Jang, Shixiang Gu, and Ben Poole. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144* (2016).

Byounghyo Lee, Dongyeon Kim, Seungjae Lee, Chun Chen, and Byoungho Lee. 2022. High-contrast, speckle-free, true 3D holography via binary CGH optimization. *arXiv preprint arXiv:2201.02619* (2022).

Chris J Maddison, Andriy Mnih, and Yee Whye Teh. 2016. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712* (2016).

Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. 2019. Holographic Near-eye Displays Based on Overlap-add Stereograms. *ACM Trans. Graph.* 38, 6 (2019).

Yifan Peng, Suyeon Choi, Nitish Padmanaban, and Gordon Wetzstein. 2020. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* 39, 6 (2020), 1–14.

Liang Shi, Beichen Li, Changil Kim, Petr Kellnhofer, and Wojciech Matusik. 2021. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* 591, 7849 (2021), 234–239.

Dongheon Yoo, Youngjin Jo, Seung-Woo Nam, Chun Chen, and Byoungho Lee. 2021. Optimization of computer-generated holograms featuring phase randomness control. *Opt. Lett.* 46, 19 (2021), 4769–4772.

Table 3. PSNR metrics of captured 3D results using different CGH algorithms, including the SGD-NH3D using RGBD input [Choi et al. 2021] with 1 frame and 8 multiplexed frames, respectively, SGD-our model using RGBD input, SGD-our model using RGBD input with Gumbel-Softmax (GS), and SGD-our model using Focal Stack (FS) supervision with GS. Images assessed are shown in Figure S14, Figure S15, and Figure 6 in the main paper. For each cell, the first PSNR is evaluated with respect to the RGBD all-in-focus targets, while the second one with respect to the focal stack. Note that the first four columns are supervised on RGBD input, where ours achieves the best all-in-focus PSNR, and the fifth column is supervised on a focal stack, and achieves the best performance on the PSNR metric on the target focal stack.

	Methods (SGD-propagation operator)				
	NH3D, 1 frame	NH3D, 8 frames	ours, 8 frames	ours, w/ GS, 8 frames	ours, w/ GS (FS), 8 frames
Robot	23.5 / 19.7	25.6 / 21.6	27.9 / 23.2	**28.7** / 23.6	27.7 / **26.1**
Sintel Bamboo	28.3 / 24.3	30.0 / 26.5	30.4 / 26.9	**31.0** / 27.0	30.3 / **30.1**
Hyperism Room	21.2 / 18.5	22.5 / 20.3	24.0 / 21.7	**24.7** / 22.3	24.2 / **23.7**
Big Buck Bunny	24.3 / 21.3	25.8 / 23.2	26.1 / 24.0	**26.7** / 24.5	25.9 / **26.9**

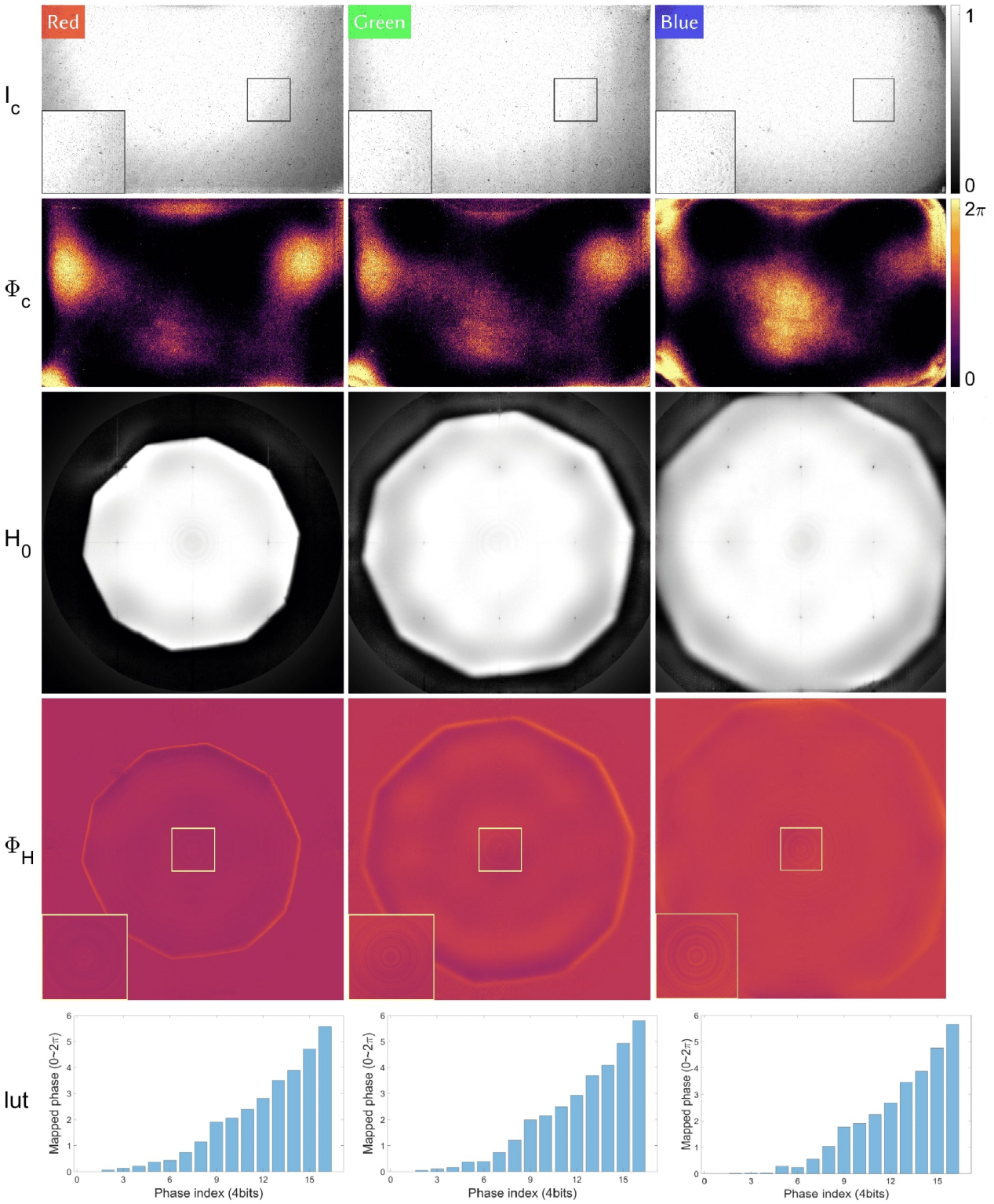

Fig. S7. Parameters visualization of the calibrated model of our holographic display prototype (refer to Section 3 in main text). From left to right: red, green, and blue channels. From top to bottom: learned intensity on SLM plane, learned phase on SLM plane, learned amplitude of the optical filter on Fourier plane, learned phase of the optical filter on Fourier plane, and learned look up table for phase mapping.

Fig. S8. Experimental holographic images reconstructed with different CGH algorithms when displaying one single frame and multiplexing 8 frames.

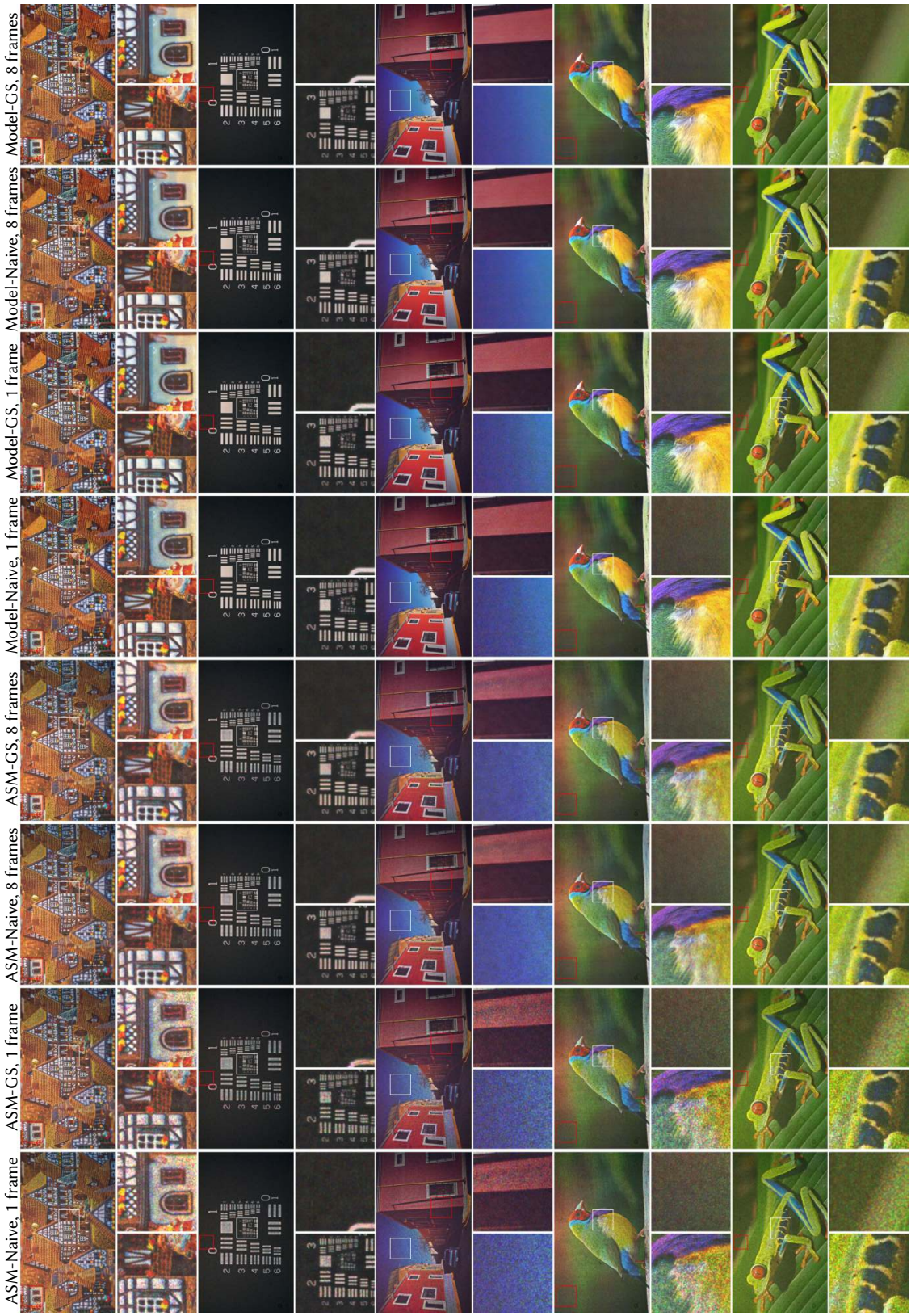

Fig. S9. Experimental holographic images reconstructed with different CGH algorithms when displaying one single frame and multiplexing 8 frames.

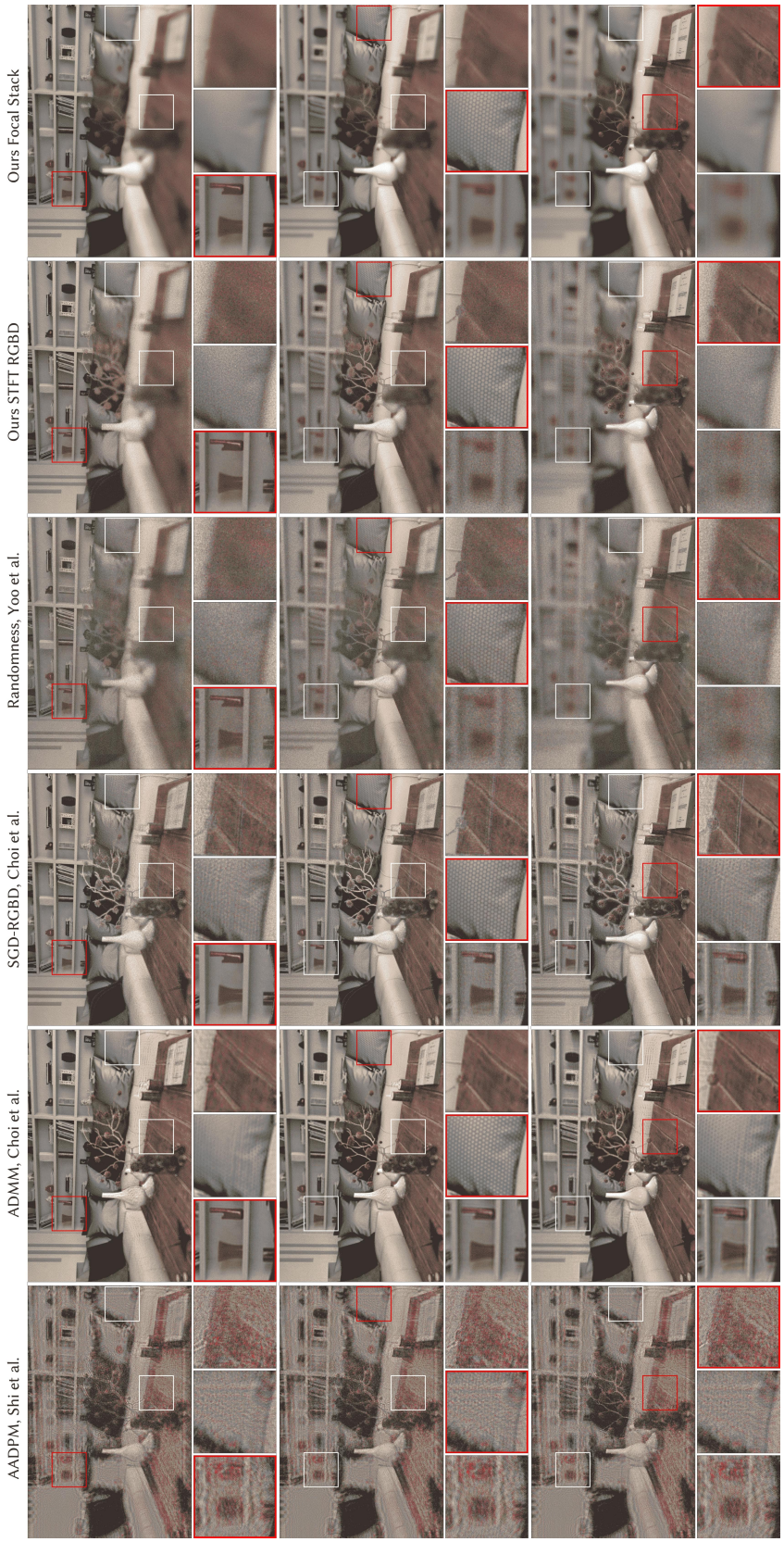

Fig. S10. Simulated focal stacks reconstructed with different CGH algorithms on ideal quantized SLMs. From left to right: AADPM from Shi et al. [Shi et al. 2021], Model ADMM from Choi et al. [Choi et al. 2021], SGD-RGBD from Choi et al. [Choi et al. 2021], Randomness Control from Yoo et al. [Yoo et al. 2021], our STFT RGBD implementation, and our focal stack implementation. Focused regions are highlighted with red squares, that from top to bottom, indicate far, center, and near distances.

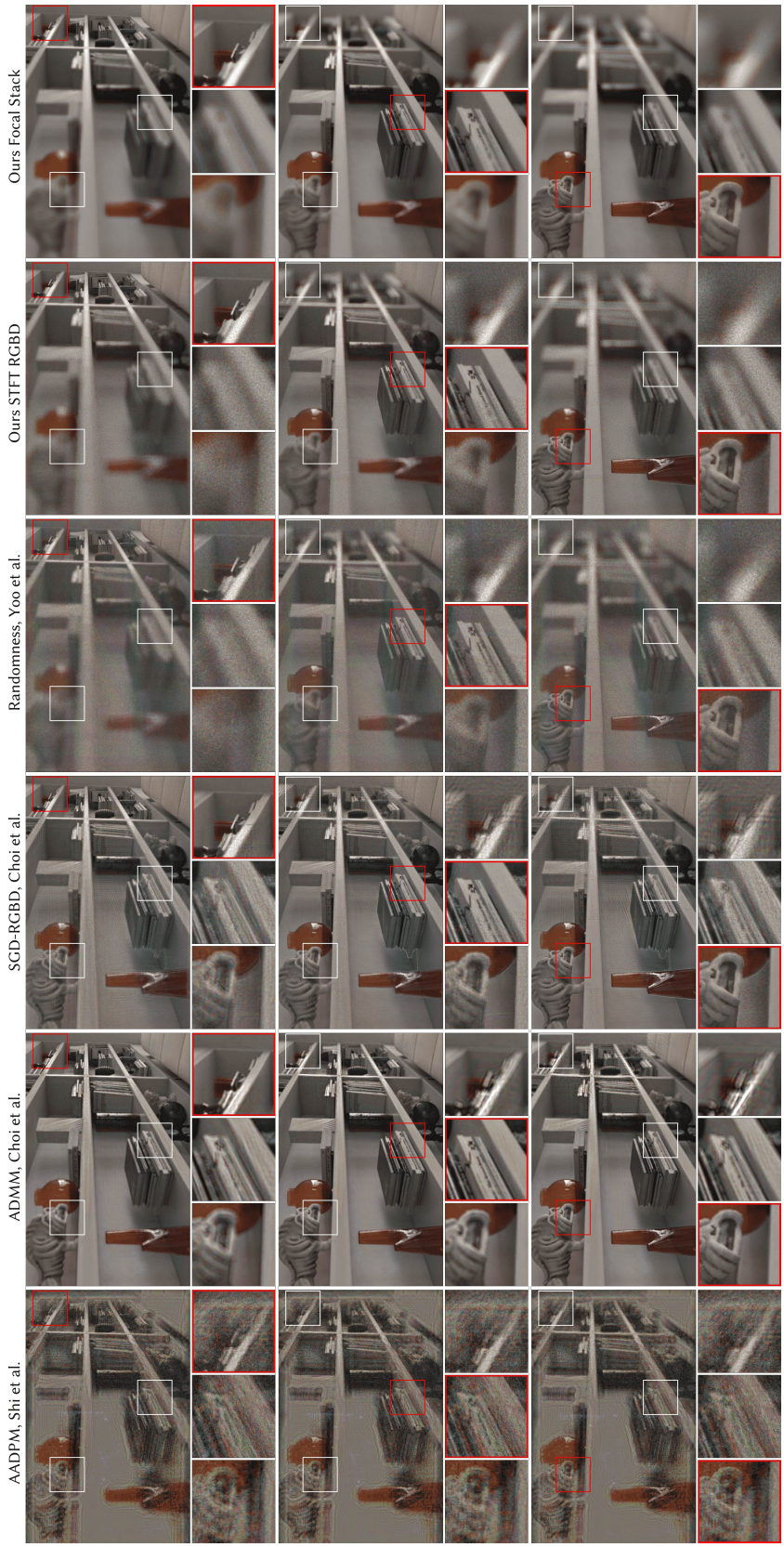

Fig. S11. Simulated focal stacks reconstructed with different CGH algorithms on ideal quantized SLMs. From left to right: AADPM from Shi et al. [Shi et al. 2021], Model ADMM from Choi et al. [Choi et al. 2021], SGD-RGBD from Choi et al. [Choi et al. 2021], Randomness Control from Yoo et al. [Yoo et al. 2021], our STFT RGBD implementation, and our focal stack implementation. Focused regions are highlighted with red squares, that from top to bottom, indicate far, center, and near distances.

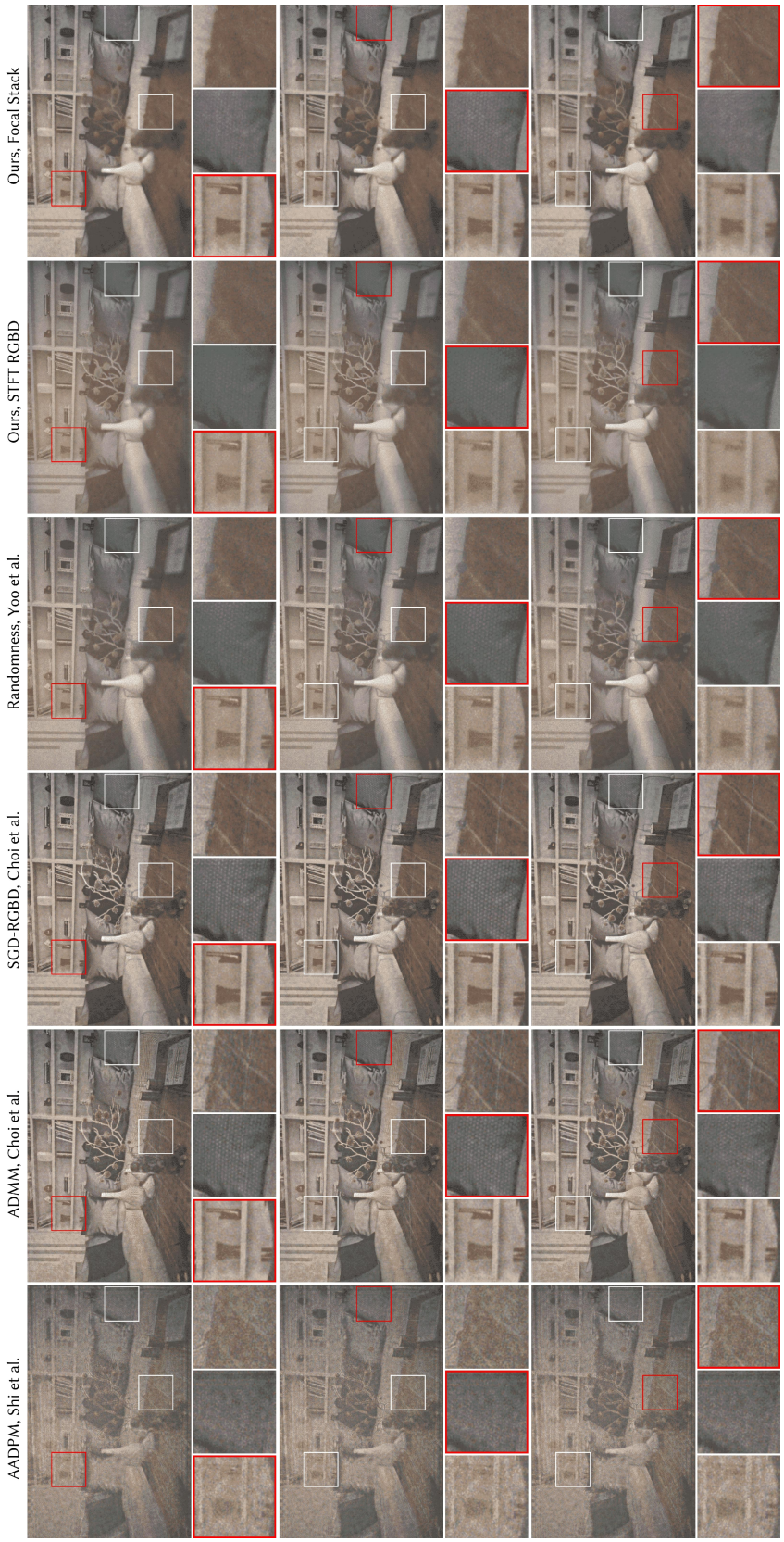

Fig. S12. Experimentally captured focal stacks reconstructed with different CGH algorithms on ideal quantized SLMs. From left to right: AADPM from Shi et al. [Shi et al. 2021], Model ADMM from Choi et al. [Choi et al. 2021], SGD-RGBD from Choi et al. [Choi et al. 2021], Randomness Control from Yoo et al. [Yoo et al. 2021], our STFT RGBD implementation, and our focal stack implementation. Focused regions are highlighted with red squares, that from top to bottom, indicate far, center, and near distances.

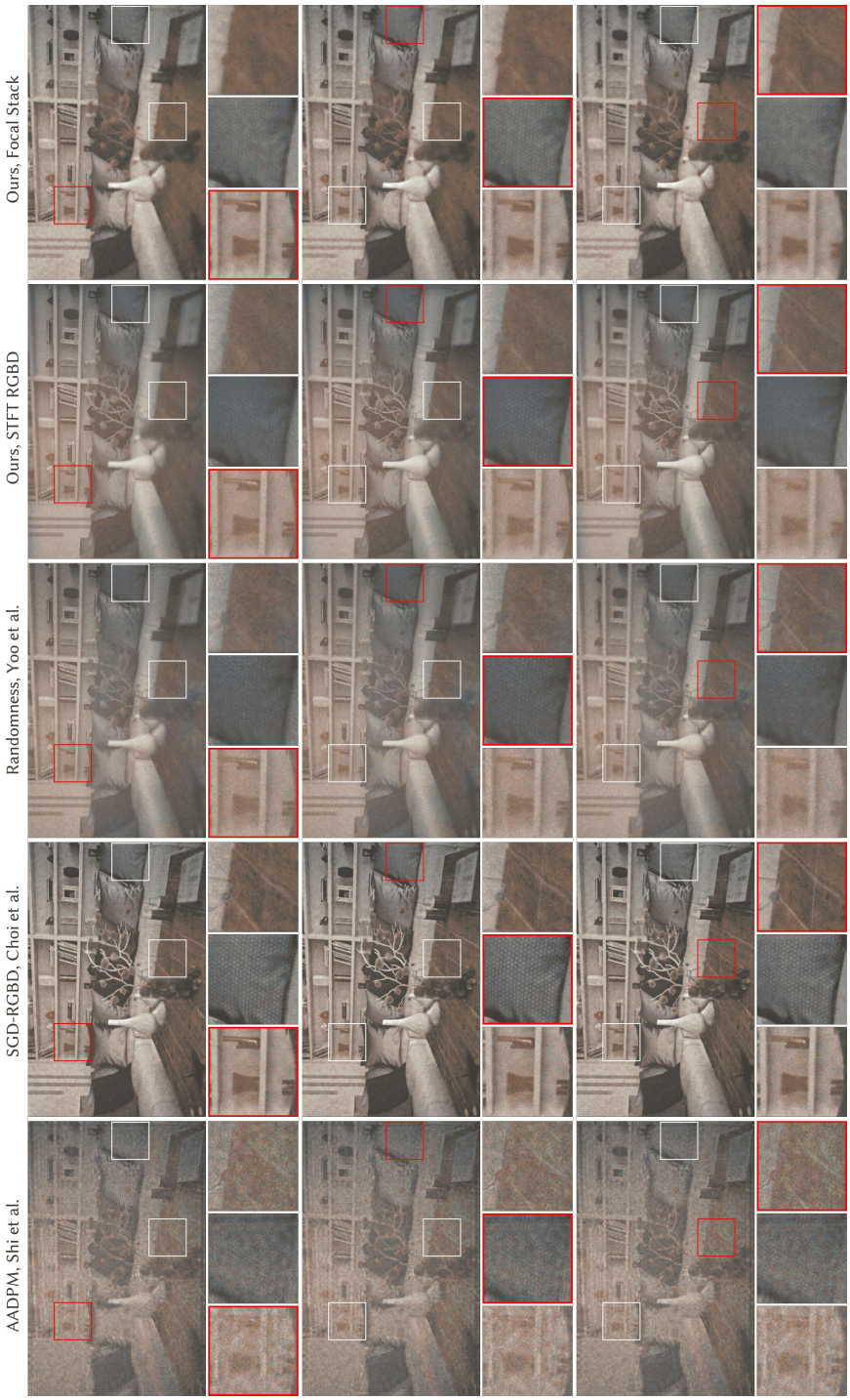

Fig. S13. Experimentally captured focal stacks reconstructed with different CGH algorithms on the Gumbel-Softmax (GS) quantized SLMs. From left to right: AADPM from Shi et al. [Shi et al. 2021], Model ADMM from Choi et al. [Choi et al. 2021], SGD-RGBD from Choi et al. [Choi et al. 2021], Randomness Control from Yoo et al. [Yoo et al. 2021], our STFT RGBD implementation, and our focal stack implementation. Note that ADMM with GS doesn't work in this case. Focused regions are highlighted with red squares, that from top to bottom, indicate far, center, and near distances.

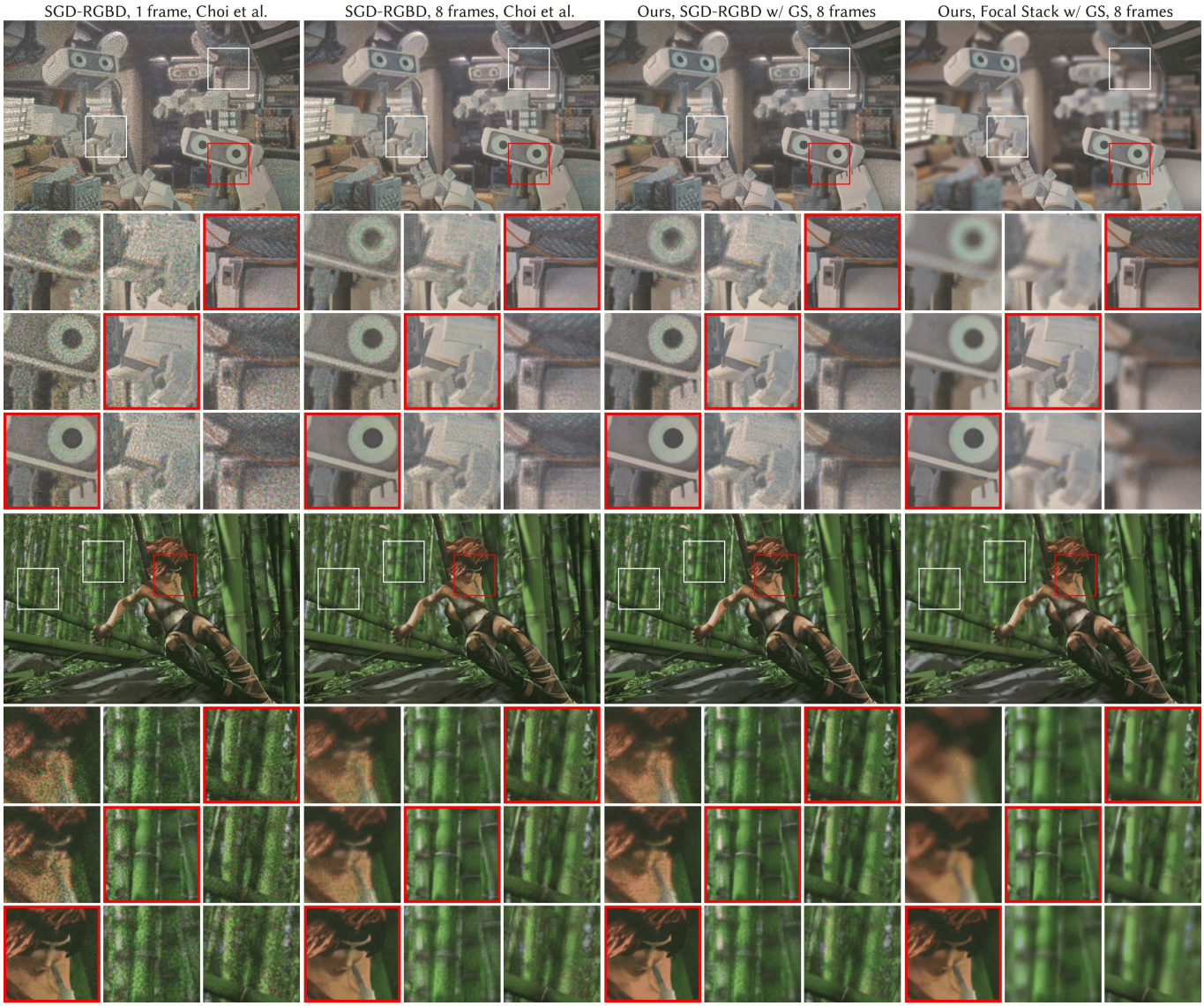

Fig. S14. Comparison of 3D CGH algorithms using experimentally captured data. In this experiment, we compare algorithms of SGD-NH3D using RGBD input [Choi et al. 2021] with 1 frame and 8 multiplexed frames, respectively, SGD-ours using RGBD input with Gumbel-Softmax (GS), and SGD-ours using Focal Stack with GS. For the top scene from left to right, the corresponding PSNR metrics are 23.5 dB, 25.6 dB, 28.7 dB, 27.7 dB with respect to the RGBD all-in-focus targets, and 19.7 dB, 21.6 dB, 23.6 dB, 26.1 dB with respect to the focal stack. Same metrics for the bottom scene are 28.3 dB, 30.0 dB, 31.0 dB, 30.3 dB and 24.3 dB, 26.5 dB, 27.0 dB, 30.1 dB. For close-ups, red squares indicate where the camera is focused at three distances (from top to bottom: far, intermediate, and near).

Fig. S15. Comparison of 3D CGH algorithms using experimentally captured data. In this experiment, we compare algorithms of SGD-NH3D using RGBD input [Choi et al. 2021] with 1 frame and 8 multiplexed frames, respectively, our SGD-RGBD with Gumbel-Softmax (GS), and our SGD-Focal Stack with GS. From left to right, the corresponding PSNR metrics are 21.2 dB, 22.5 dB, 24.7 dB, 24.2 dB with respect to the RGBD all-in-focus targets, and 18.5 dB, 20.3 dB, 22.3 dB, 23.7 dB with respect to the focal stack. For close-ups, red squares indicate where the camera is focused at three distances (from top to bottom: far, intermediate, and near).

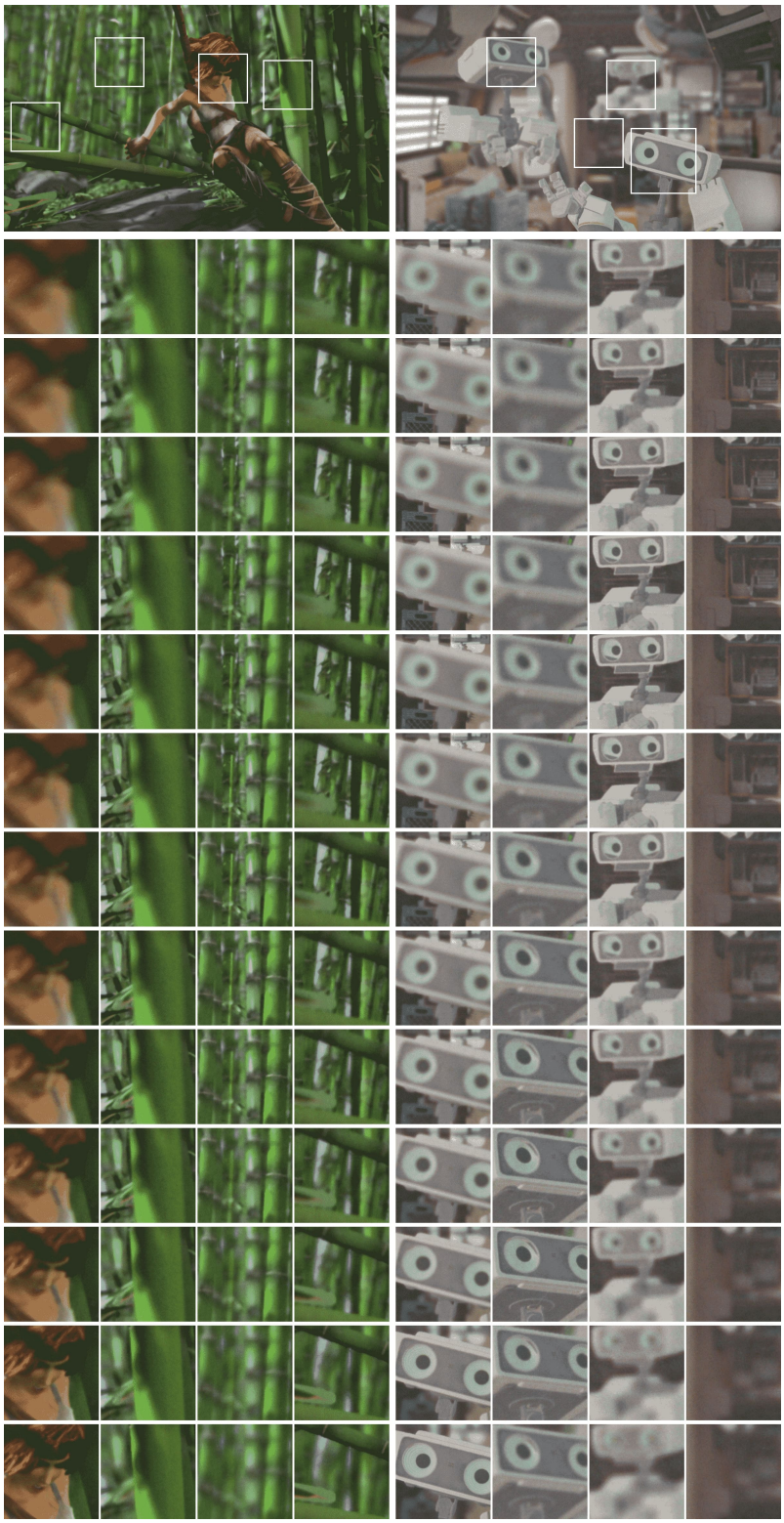

Fig. S16. Interpolation behaviour of 3D focal stack supervised holograms. We experimentally capture 13 planes and show them at each row of closeups. Note that only odd rows are supervised while unsupervised planes (even rows) interpolate it smoothly.

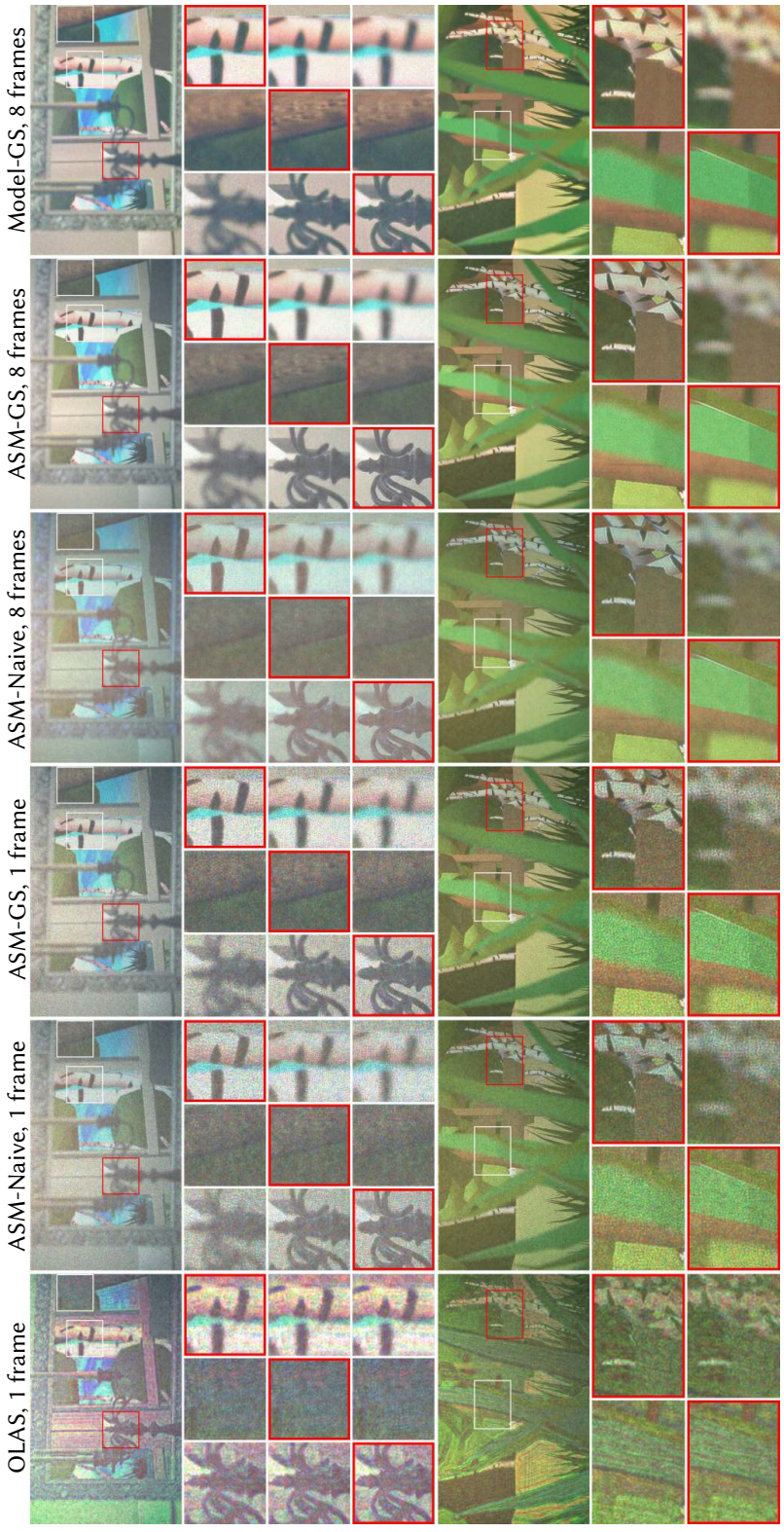

Fig. S17. Comparison of 4D CGH algorithms using experimentally captured data. In this experiment, we compare the OLAS algorithm [Padmanaban et al. 2019], and the SGD algorithms using our ASM-Naive with 1 single frame, our ASM-GS with 1 single frame, our ASM-Naive with 8 multiplexed frames, our ASM-GS with 8 multiplexed frames, and our Model-GS with 8 multiplexed frames. For close-ups, red squares indicate where the camera is focused at three distances (from top to bottom: far, intermediate, and near).

Fig. S18. Frames extracted from the camera with different spatial shifts.