Neural 3D Holography: Learning Accurate Wave Propagation Models for 3D Holographic Virtual and Augmented Reality Displays

SUYEON CHOI*, MANU GOPAKUMAR*, and YIFAN PENG, Stanford University
JONGHYUN KIM, NVIDIA and Stanford University
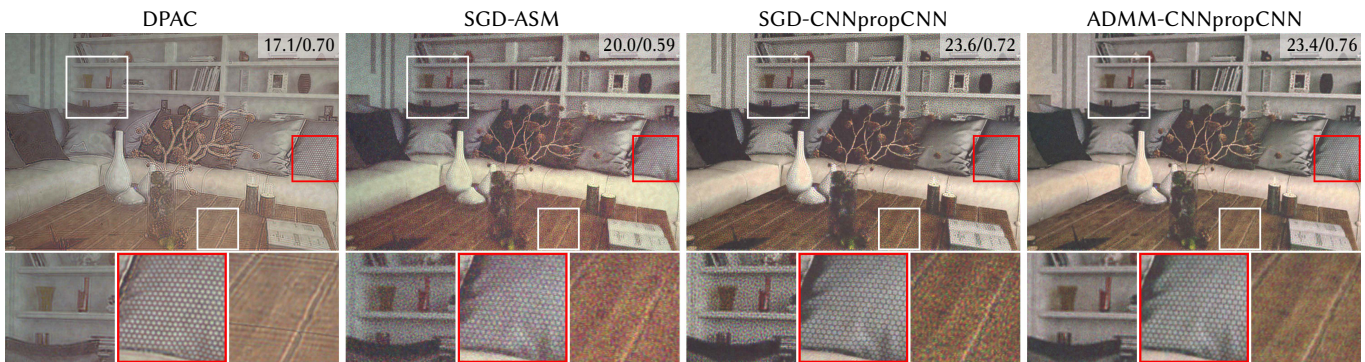GORDON WETZSTEIN, Stanford University

Fig. 1. Experimental 3D computer-generated holography (CGH) results captured with a display prototype. In this experiment, the camera is focused at an intermediate distance, i.e., the cushion. We compare several different 3D CGH algorithms under the same experimental conditions: our implementation of the double phase-amplitude coding (DPAC) approach [Maimone et al. 2017; Shi et al. 2021], a stochastic gradient descent solver used with a wave propagation model based on the angular spectrum method (SGD-ASM), SGD used with the proposed wave propagation model (SGD-CNNpropCNN), and a proximal gradient solver used with our model that additionally promotes piecewise smoothness of the complex wave field for the in-focus parts of our RGBD target images (ADMM-CNNpropCNN). Our wave propagation model enables accurate 3D holographic display with significantly improved image quality for in-focus 3D scene parts and our phase regularization strategy additionally improves strong out-of-focus speckle artifacts observed with other approaches. Quantitative evaluations for each result are included as PSNR/SSIM.

Holographic near-eye displays promise unprecedented capabilities for virtual and augmented reality (VR/AR) systems. The image quality achieved by current holographic displays, however, is limited by the wave propagation models used to simulate the physical optics. We propose a neural network–parameterized plane-to-multiplane wave propagation model that closes the gap between physics and simulation. Our model is automatically trained using camera feedback and it outperforms related techniques in 2D plane-to-plane settings by a large margin. Moreover, it is the first network-parameterized model to naturally extend to 3D settings, enabling high-quality 3D computer-generated holography using a novel phase regularization strategy of the complex-valued wave field. The efficacy of our approach is demonstrated through extensive experimental evaluation with both VR and optical see-through AR display prototypes.

*denotes equal contribution.

Authors' addresses: Suyeon Choi; Manu Gopakumar; Yifan Peng, Stanford University; Jonghyun Kim, NVIDIA and Stanford University; Gordon Wetzstein, Stanford University.

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

1 INTRODUCTION

Augmented and virtual reality (AR/VR) systems promise unprecedented user experiences, but the light engines of current AR/VR platforms are limited in their peak brightness, power efficiency, device form factor, support of perceptually important focus cues, and ability to correct visual aberrations of the user or optical aberrations of the downstream optics. Holographic near-eye displays promise solutions for many of these problems. Their unique capability of synthesizing a 3D intensity distribution with a single spatial light modulator (SLM) and coherent illumination, created by bright and power-efficient lasers, makes these displays ideal for applications in wearable computing systems.

Although the fundamentals of holography have been developed more than 70 years ago, until recently, high-quality holograms have

only been achieved using optical recording techniques [Benton and Bove 2008]. The primary challenge for generating high-quality digital holograms using SLMs in a computationally efficient manner are the algorithms used for computer-generated holography (CGH). Traditional CGH algorithms [Chang et al. 2020; Park 2017] rely on simulated wave propagation models that do not adequately represent the physical optics of a near-eye display, thus severely limiting the achievable quality. Recently, a class of machine learning–enabled holographic wave propagation models has been proposed that partly overcome these challenges. For example, Peng et al. [2020] and Chakravarthula et al. [2020] proposed automatic ways to calibrate neural network–parameterized 2D wave propagation models using cameras, thereby significantly improving upon previously reported holographic image quality. In these works, the networks parameterize the forward propagation from SLM to target image; they learn optical aberrations and other discrepancies between the physical optics and a propagation model to make the latter more accurate but not necessarily faster than classic models. Horisaki et al. [2018], Peng et al. [2020], Eybposh et al. [2020], Lee et al. [2020], and Shi et al. [2021] also introduced various neural network architectures for fast holographic image synthesis. These "inverse" networks are trained to learn a mapping from image plane(s) to SLM plane, so that a target image can be quickly converted to a phase-only SLM pattern without the need for iterative optimization. The image quality achieved by these methods, however, is fundamentally limited by the forward wave propagation models they are trained with. The only approach that combines a network-parameterized forward model with an inverse network is that of Peng et al. [2020], but their model is limited to 2D plane-to-plane propagation.

Our work aims to unlock the full potential of emerging holographic near-eye displays in synthesizing high-quality 3D holograms. We argue that the key technology necessary to achieve this goal is an accurate and differentiable plane-to-multiplane forward wave propagation model that adequately simulates the physical optics of a display. Here, we address this issue by combining the classic angular spectrum method (ASM) with convolutional neural networks (CNNs) to form a unique wave propagation model representing a hybrid between classic physics models and modern networks. The learnable parameters of our model are automatically calibrated with a camera and we demonstrate this model to be significantly more accurate in representing the physical optics of a display than previously proposed 2D wave propagation models. The model is also more general in being the first network-parameterized model to represent 3D plane-to-multiplane propagation rather than 2D plane-to-plane propagation. Once trained for a specific holographic display, our model can be used with a number of different solvers that take an RGBD image as input and optimize the phase pattern to be displayed on an SLM.

Although our camera-calibrated model significantly advances state-of-the-art CGH algorithms, we demonstrate it using iterative solvers that do not operate in real time. Real-time inverse networks can be trained using models like ours (see e.g., [Peng et al. 2020]), but we leave this as a future software engineering challenge. With our work, we help make holographic near-eye displays a practical technology for emerging AR/VR applications by optimizing the image quality for near-continuous 3D display settings.

Specifically, we make the following contributions:

- We propose a differentiable camera-calibrated model for the wave propagation in holographic near-eye displays. This model more accurately represents physical optics than previous approaches in 2D settings and it is the first network-parameterized model to operate in a 3D plane-to-multiplane setting.
- We develop a strategy for 3D multiplane CGH optimization that significantly reduces speckle in out-of-focus regions; this strategy places a piecewise smoothness constraint on the phase of in-focus regions and it is optimized using a proximal gradient solver.
- We evaluate our methods with both virtual and optical see-through augmented reality display prototypes, demonstrating the highest-quality 2D and 3D holographic display results to date.

Note that certain types of view-dependent effects, such as specular highlights or (dis)occlusion for large amounts of parallax, may not be supported by multiplane representations such as ours. Yet, the small eyebox afforded by the limited space–bandwidth product of current SLMs makes the lack of these effects negligible in the context of near-eye display applications.

2 RELATED WORK

Various aspects of holographic displays have been actively investigated over the last few decades. We summarize this body of work in the following, but refer the interested reader to the recent surveys by Park [2017] and Chang et al. [2020] for additional discussions and references.

Holographic Display Optics. Much progress has recently been made by the computational optics community in advancing hardware aspects related to holographic near-eye displays. For example, advances have been reported in optimizing diffractive optical elements [Li et al. 2016; Maimone and Wang 2020; Yeom et al. 2015], laser scanning or steering mechanisms [Jang et al. 2018, 2017], and operation with incoherent emitters [Moon et al. 2014; Peng et al. 2021] or amplitude-only SLMs [Gao et al. 2016]. Kuo et al. [2020] showed how to expand the étendue of a near-eye display, which is a hardware limitation, using diffusive combiners.

One of the primary benefits of holographic near-eye displays over conventional microdisplays is their support of focus cues. Although this capability is also supported by near-eye light field displays [Hua and Javidi 2014; Huang et al. 2015; Lanman and Luebke 2013], light field displays typically sacrifice spatial resolution for this purpose. Our work is primarily focused on advancing wave propagation models to unlock the full potential of high-quality near-eye 3D holographic displays for virtual and augmented reality applications.

Computer-generated Holography. Many algorithms have been proposed to convert a target 2D or 3D intensity image into a phase-only pattern to be displayed on an SLM. These can be roughly classified as using point [Fienup 1982; Gerchberg 1972; Maimone et al. 2017; Shi et al. 2017, 2021], polygon [Chen and Wilkinson 2009; Matsushima and Nakahara 2009], light ray [Wakunami et al. 2013; Zhang et al. 2011], or layer [Chen et al. 2021; Chen and Chu

2015; Zhang et al. 2017] primitives for wave propagation. Alternatively, holographic stereograms convert light fields into holograms and, similarly to some of the aforementioned approaches, encode depth- and view-dependent effects [Benton 1983; Kang et al. 2008; Lucente and Galyean 1995; Padmanaban et al. 2019; Yaras et al. 2010; Zhang and Levoy 2009; Ziegler et al. 2007]. Among these CGH algorithms, direct methods are usually fast but rely on some kind of phase coding [Hsueh and Sawchuk 1978; Lee 1970; Maimone et al. 2017] to represent a complex-valued field using phase-only SLMs. Single-SLM phase coding approaches interlace two phase-only patterns representing the complex field. This is light inefficient because it creates multiple copies of the holographic image, which need to be optically filtered using additional opto-mechanical elements. Dual-SLM phase coding is challenging to implement, because of the increased bulk, cost, and required calibration. Direct methods thus often provide lower image quality or reduced brightness compared to iterative approaches [Chakravarthula et al. 2019; Dorsch et al. 1994; Fienup 1982; Gerchberg 1972; Peng et al. 2020, 2017]. Our approach builds on a recent class of CGH algorithms that leverage modern machine learning methods to overcome long-standing challenges of other CGH approaches, such as 3D holographic image quality, including out-of-focus behavior.

Machine-learning-based Holographic Displays. Horisaki et al. [2018] were the first to propose a neural network to synthesize phase patterns in holographic display applications, although the quality of their results was limited. Peng et al. [2020] recently proposed a network architecture that enabled real-time 2D holographic display with an image quality comparable to that of previous iterative methods. Shi et al. [2021] proposed a related network architecture that is more efficient and also works for 3D holographic images. Both Peng et al. [2020] and Chakravarthula et al. [2020] realized that the simulated wave propagation models used by prior work do not adequately capture the physics of a holographic display; both works proposed camera-based calibration techniques that were used to optimize the parameters of different neural network–parameterized 2D plane-to-plane wave propagation models. A related approach was also adopted by Choi et al. [2021] to optimize the achieved 2D and 3D holographic image quality using two SLMs. Finally, Eybposh et al. [2020] proposed a network architecture for fast CGH with 3D multiplane intensity input for applications in holographic multiphoton microscopy. Their wave propagation model, however, is based on classical optics propagation operators, which the proposed model outperforms by a large margin for holographic display applications (see SGD-ASM results in Figs. 4,5 and Tab. 1).

In this work, we propose a new wave propagation model for light transport in a holographic near-eye display. This is closely related to the 2D plane-to-plane wave propagation models proposed by Peng et al. [2020] and Chakravarthula et al. [2020] but our model is both more accurate and more general by also modeling 3D plane-to-multiplane light transport. Similar to Peng et al. [2021; 2020], Chakravarthula et al. [2020], and Choi et al. [2021], we use a camera in the loop for calibration purposes. We do not claim this to be a contribution of our work, although our technique is the first to demonstrate multiplane camera-based holographic display calibration with more than two planes. All of the prior methods

have been demonstrated using VR-type displays – here, we show that our camera-calibrated 3D wave propagation model is able to achieve high-quality 3D results in both VR and optical see-through AR scenarios.

Other applications. A variety of applications across science and engineering rely on holographic illumination, such as optical tweezers [Curtis et al. 2002], 3D printing [Shusteff et al. 2017], neuroimaging [Hernandez et al. 2016; Yang et al. 2015], and optogenetics [Papagiakoumou et al. 2010]. The methods developed here may also be applicable in those domains.

3 A 3D NEURAL NETWORK–BASED WAVE PROPAGATION MODEL

In this section, we briefly review conventional simulated wave propagation models before introducing a network-parameterized, camera-calibrated model.

3.1 Traditional Holographic Wave Propagation

We work with a Fresnel hologram configuration, where a collimated coherent laser beam is incident on a phase-only SLM that delays the phase of this source field u_{src} in a per-pixel manner. The task for any CGH algorithm is then to determine the best SLM phase pattern $\phi \in \mathbb{R}^{M \times N}$, i.e., the hologram, for a target 2D or 3D intensity distribution specified at some distance z in front of the SLM. A popular model that simulates the propagation of a complex wave $u = ae^{i\phi}$ from one plane to another, for example SLM to target plane, is the angular spectrum method (ASM) [Goodman 2005]:

$$f_{\mathrm{ASM}}(u, z) = \iint \mathcal{F}\left(a(x, y, \lambda) e^{i\phi(x, y, \lambda)} u_{\mathrm{src}}(x, y)\right) \tag{1}$$

$$\cdot \mathcal{H}(f_x, f_y, \lambda, z) e^{i2\pi(f_x x + f_y y)} df_x df_y,$$

$$\mathcal{H}(f_x, f_y, \lambda, z) = \begin{cases} e^{i\frac{2\pi}{\lambda}\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} z}, & \text{if } \sqrt{f_x^2 + f_y^2} < \frac{1}{\lambda}, \\ 0 & \text{otherwise} \end{cases}$$

where f_{ASM} is the propagation operator, λ is the wavelength, f_x, f_y are spatial frequencies, $\mathcal{H}$ is the transfer function, $\mathcal{F}(\cdot)$ denotes the Fourier transform, and $a(x, y) = const.$ when the field u describes a phase-only SLM. Note that f_{ASM} is a simulated model of the unknown wave propagation operator f describing the physical optics. This model operates on complex-valued fields that contain both amplitude and phase information. The intensity this model predicts at distance z when displaying phase ϕ on the SLM is $|f_{\mathrm{ASM}}(e^{i\phi}, z)|^2$.

The ASM wave propagation model can be used to solve the inverse problem of computing a phase pattern ϕ for a single or a set of multiple target image amplitudes $a_{\mathrm{target}}^{\{j\}}$ located at the set of distances $z^{\{j\}}, j = 1 \dots J$ from the SLM by solving the following objective:

$$\underset{\phi}{\text{minimize }} \mathcal{L}_a\left(s \cdot |f_{\mathrm{ASM}}\left(e^{i\phi}, z^{\{j\}}\right)|, a_{\mathrm{target}}^{\{j\}}\right). \tag{2}$$

Here s is a fixed or learnable scale factor that accounts for possible differences in the range of values between the output of f_{ASM} and $a_{\mathrm{target}}^{\{j\}}$. The loss function $\mathcal{L}_a$ constrains the amplitudes at the target planes and this problem can be conveniently solved using variants
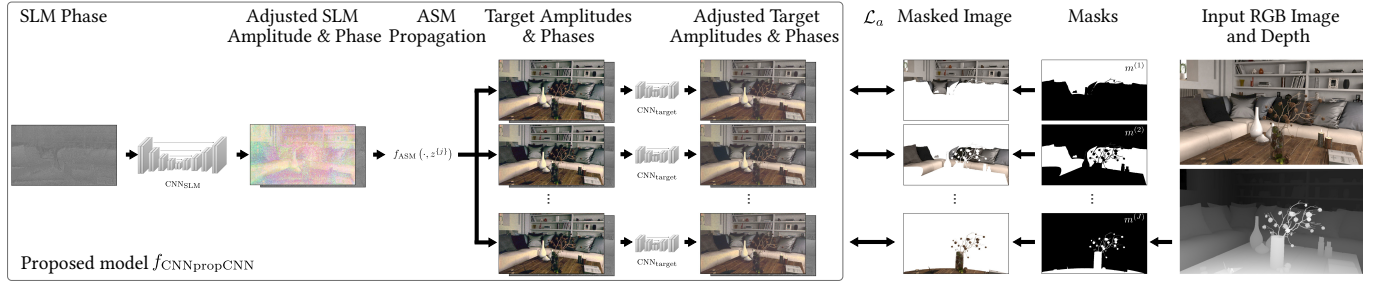
Fig. 2. Illustration of our 3D wave propagation model and RGBD supervision strategy. The phase pattern displayed by the SLM is processed by a CNN. The resulting complex-valued wave field is propagated to all target planes using a conventional ASM wave propagation operator. The wave fields at each target plane are processed again by smaller CNNs. The loss function constrains the masked amplitudes at the target planes to match the masked target RGB image, where the binary masks at each target plane are computed from the target depth map.

of stochastic gradient descent (SGD), as recently proposed by Peng et al. [2020]. In the remainder of this paper, we will refer to this approach to computing 2D or multiplane 3D holograms as the SGD-ASM method.

3.2 Camera-calibrated Wave Propagation Model

Analytic models, such as the ASM, are great for simulations. However, they are often poor representations of the true wave propagation operator of a physical optical system. Small imperfections, such as optical aberrations, phase nonlinearities of the SLM, stray light, or the finite diffraction efficiency of the SLM, make it difficult to use the ASM out of the box and calibrating all of these possible sources of imperfection is a tedious or impossible task.

To overcome these challenges, both Peng et al. [2020] and Chakravarthula et al. [2020] recently proposed neural network–parameterized models for 2D plane-to-plane wave propagation. Both approaches use a camera to automatically calibrate the respective model by showing a set of training phase patterns on the SLM, capturing the resulting intensity on the target plane, and then fitting the model parameters to these phase–intensity pairs. Peng et al.'s model (NH) is focused on being interpretable whereas Chakravarthula et al. applied a CNN on the intensity $|f_{\text{ASM}}\left(e^{i\phi}, z\right)|^2$ performing image-to-image translation from ideal predicted image to captured image (HIL). Both of these models improve 2D holographic image quality and Peng et al. also demonstrated varifocal and multiplane holographic display modes that either selected or interpolated between several 2D models trained for different distances. Moreover, Peng et al. also proposed a camera-in-the-loop (CITL) approach that uses the physical optical forward model, which is by definition the ideal forward model, albeit with approximated gradients for the backward pass. Therefore, all of these approaches are limited in either the accuracy of their respective forward model, or its gradients, or both.

None of these existing approaches naturally extends to 3D and neither of the two models actually works very well, as we will show in Section 4. We speculate that this is due to the following reasons: while the NH model is interpretable, it does not have sufficient degrees of freedom to learn aspects that are not modeled well by their interpretable model, such as undiffracted light. While the HIL

model is flexible enough to learn pretty much any deviation between the intensity of a holographic image and a target image, it has difficulty learning a physically meaningful mapping because it does not utilize the phase information on the target plane. Although the CITL approach uses the ideal forward "model", it is limited in requiring errors to be backpropagated into the unknown SLM phase map using the gradients of the ASM, which turn out to be a poor approximation of the physical optics. We provide evidence of these hypotheses in Section 4.

We propose a new wave propagation model that combines the strengths of these previous approaches while naturally and efficiently extending them to a 3D multiplane setting. To overcome the limitations of Peng's model, we sacrifice interpretability and let the network learn the difference between physical and simulated light transport. To combat the limitations of HIL, we design our network architecture to operate on the complex-valued wave field directly rather than on the intensity at a specific target plane. Thus, our approach allows the network to learn a physically meaningful residual between the wave field simulated by the ASM and that of the physical optical system, which we demonstrate by the improved quality and generalization behavior of our model in Section 4.

Specifically, our proposed model combines the ASM with CNNs as

$$f_{\text{CNNpropCNN}}\left(u, z^{\{j\}}\right) = \text{CNN}_{\text{target}}\left(f_{\text{ASM}}\left(\text{CNN}_{\text{SLM}}\left(e^{i\phi}\right), z^{\{j\}}\right)\right), \quad (3)$$

where $\text{CNN}_{\text{SLM}} : \mathbb{R}^{2 \times M \times N} \to \mathbb{R}^{2 \times M \times N}$ takes as input two channels with the real and imaginary values of the field on the SLM and outputs the real and imaginary components of the adjusted field at the SLM plane, correcting SLM nonlinearities, spatially varying source intensity, optical aberrations, and other factors. The adjusted field is then propagated to one or several target planes using the ASM. We apply another CNN at each of the target planes, $\text{CNN}_{\text{target}} : \mathbb{R}^{2 \times M \times N} \to \mathbb{R}^{2 \times M \times N}$, on the real and imaginary channels of the complex-valued ASM-propagated fields at the target planes *before* converting the resulting fields to intensity. Using the same shared CNN at each target plane worked well for our optical setup in practice. Indeed, this strategy was necessary for our model to achieve a good performance for the held-out plane and thus generalize across depths (see Sec. 4.4). The model could also use different CNNs for each target plane if required, but that would require more memory,

it did not improve the results in our experiments, and it led to lower performance on the held-out plane. Figure 2 shows an illustration of our final model.

A unique characteristic of our model, compared to NH and HIL, is that we apply CNNs on both SLM and target planes rather than only on the target planes. Indeed, as shown in Table 1, CNN_{SLM} is one of the most critical components of such a model as it significantly improves the accuracy of the wave propagation model. This is not only true for the 2D plane-to-plane variant but also for 3D plane-to-multiplane propagation. We speculate that CNN_{SLM} is more efficient in learning nonlinear behavior of the SLM.

3.3 Network Architecture, Dataset Acquisition, and Model Training

Both CNN_{SLM} and $\text{CNN}_{\text{target}}$ are implemented using the well-known UNet architecture [Ronneberger et al. 2015] with two input and output channels for the real and imaginary components of the fields. These CNNs do not directly implement complex number arithmetic, as this is currently not supported by PyTorch, but the CNN still learns to operate on the complex-valued field. CNN_{SLM} uses skip connections and 8 consecutive downsampling operations using strided convolutions as well as 8 consecutive upsampling operations using transposed convolutions. This CNN uses 32 feature channels after the input layer with feature channels doubling for each downsampling layer to a maximum of 512 channels. The smaller $\text{CNN}_{\text{target}}$ network has 5 downsampling and upsampling layers with 8 feature channels after the input, doubling with each downsampling layer to a maximum of 128 channels. Both networks use instance normalization [Ulyanov et al. 2016], Leaky ReLU (slope -0.2) for the down blocks, and ReLU nonlinearities for the up blocks.

To train this CNN-parameterized model, we capture many pairs of SLM phase maps and resulting single or multiplane intensities using a camera. Instead of using random phases, we generate a pool of phase patterns using traditional CGH algorithms with high-resolution target images from the DIV2K dataset [Agustsson and Timofte 2017]. Specifically, for each of our 8 target planes, we generate 100 phase patterns using the DPAC algorithm and 1,000 using SGD from random images of the dataset. For the SGD-optimized phases, we also randomize the number of iterations and initialize the phase patterns with random phase values. Our training data thus consists of 8,800 phase patterns and corresponding captured intensity images in total.

The model parameters, i.e., the weights and bias terms of CNN_{SLM}, $\text{CNN}_{\text{target}}$ are then optimized using the ADAM solver in PyTorch. One model is optimized separately for each of the three color channels. We use a learning rate of $5e^{-4}$, batch size of 1, and the ℓ_1 loss. Training takes about two days and is stopped if the validation loss does not decrease after 10 epochs.

Additional implementation details and pseudo-code for all algorithms can be found in the supplemental material. Source code is available on the project website[1].

3.4 Inference with a Trained Model

Once trained, we can use the wave propagation model to compute phase patterns of 2D or multiplane 3D target images by solving Equation 2 using $f_{\text{CNNpropCNN}}$ instead of f_{ASM} with SGD. This is an iterative approach, which takes a few tens of seconds or minutes to complete and is thus not real time. However, our model could also be used as a loss function to train another neural network for real-time inference, as recently demonstrated by Peng et al. [2020]; we did not attempt this in our work and leave it as future work.

The input of the 2D variant of our approach is simply a 2D image at some specific distance z from the SLM. When used with a 3D multiplane holographic display mode, the naive approach would be to constrain all planes simultaneously using a rendered focal stack of the target scene. This approach, however, is not ideal for several reasons. First, it requires the focal stack of the target scene to be rendered, which is computationally costly. Second, supervising with a focal stack over-constrains the system, because Equation 2 would have many more target observations than unknown phase values. This approach would therefore likely exceed the degrees of freedom of the SLM. Third, it requires the defocus blur of the system to be explicitly modeled in order to render the focal stack. This is not trivial, because if one wanted to make this defocus blur perceptually correct, one would have to track the pupil diameter of the user, which requires additional system complexity. If one wanted to model the defocus blur naturally supported by the holographic display, one would have to take the space–bandwidth product of the SLM into account. Moreover, the physical defocus behavior of a coherent wave field is unintuitive and different from the incoherent light we typically see in our natural environments.

To mitigate these challenges, we propose a computationally efficient approach that requires only an RGBD image of the target scene, rather than a multiplane volume or a focal stack. Depth maps are readily available for all computer-generated content and they can be approximated for photographic content using computer vision techniques known as monocular depth estimation. Working with RGBD images for multiplane holography is not new and requires the values of the depth map to be quantized to the nearest holographic display plane. Thus, each pixel location across all target planes j is only constrained for one of the target depth planes – the one closest to the corresponding depth value at that location. This is formalized by quantizing a target depth map D by converting it to the set of binary masks $m^{(j)} \in \{0, 1\}^{J \times M \times N}$, such that

$$m^{(j)}(x, y) = \begin{cases} 1, & \text{if } |z^{(j)} - D(x, y)| < |z^{(k)} - D(x, y)|, \forall k \neq j \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Intuitively, mask pixel $m^{(j)}(x, y)$ is set to 1 if the value of the depth map at this location, $D(x, y)$, is closer to the axial location $z^{(j)}$ of mask layer j than to any of the other mask layers.

3.5 Speckle-free 3D Holography using Phase Regularization

The inference procedure described above is computationally efficient and works well by constraining the in-focus parts of the scene. But it leaves the out-of-focus behavior of the wave field unconstrained, leading to unpredictable behavior that usually results in

---

[1]https://www.computationalimaging.org/publications/neuralholography3d/

strong speckle artifacts (see Section 4). This is intuitive, because the phase of in-focus parts on the target planes are usually random and thus decorrelate quickly, resulting in speckle. To counteract this decorrelation behavior, we propose an indirect way to constrain the out-of-focus amplitudes by regularizing the in-focus phase values of the optimized propagating wave field. For this purpose, we promote piecewise smoothness of the in-focus phase values. This allows for sparse discontinuities in the phase patterns around texture or depth edges, but it suppresses the random in-focus phase noise that makes the field decorrelate as it propagates out of focus. Specifically, the complete loss function incorporating both multiplane amplitude and phase constraints is

$$\mathcal{L} = \mathcal{L}_a + \mathcal{L}_\phi = \sum_{j=1}^{J} \left\| \left(s \cdot |f_{\text{CNNpropCNN}}\left(e^{i\phi}, z^{(j)}\right)| - a_{\text{target}}^{(j)} \right) \circ m^{(j)} \right\|_2^2$$

$$+ \gamma \sum_{j=1}^{J} \left\| \Delta\Phi \left(f_{\text{ASM}} \left(\text{CNN}_{\text{SLM}}\left(e^{i\phi}\right), z^{(j)} \right) \right) \circ m^{(j)} \right\|_1, \quad (5)$$

where $\circ$ denotes element-wise multiplication, Δ is the Laplace operator, $\Phi(\cdot)$ is an operator that extracts the phase of a complex-valued field, and γ is a user-defined regularization weight that balances the two loss terms. Note that the second term of the loss regularizes the phase directly after the f_{ASM} operator, because the phase output of $f_{\text{CNNpropCNN}}$ is not constrained by the model training and may therefore not be physically meaningful.

If only the first term of this loss function is optimized, as is the case for 2D holographic display scenarios, this objective can be efficiently solved using variants of SGD. Solutions to the complete loss function found by these solvers, however, are often unsatisfactory, as these solvers are not effective in inducing the sparsity that the ℓ_1 norm is meant to promote. As an alternative, proximal gradient solvers are typically better at these tasks [Bach et al. 2012], which we verify in the supplemental material. Thus, we propose an ADMM-based solver to minimize the loss function of Equation 5. This solver is derived in detail in the supplement.

4 EXPERIMENTAL ASSESSMENT

4.1 Prototype Holographic VR Display

Our VR prototype uses a FISBA RGBeam fiber-coupled module with three optically aligned laser diodes with a maximum output power of 50 mW and wavelengths of 636.4, 517.7, and 440.8 nm, respectively. We use a Holoeye Leto phase-only liquid crystal on silicon SLM with a resolution of 1920 × 1080 pixels, a pixel pitch of 6.4 μm, and a precision of 8 bits. Color images are captured as separate exposures for each wavelength and combined in post-processing. All images are captured with a FLIR Grasshopper3 2.3 MP color USB3 vision sensor through a Canon EF 50mm lens for the VR setup. The SLM, Canon lens, and sensor are synchronized in hardware with an Arduino Uno to enable programmable focus settings (see Bando et al. [2013] for a tutorial) over a total depth range of 2 D, from optical infinity to approx. 0.5 m. This setup is shown in Figure 3. Our optical see-through AR (OST-AR) prototype uses a Quantum gem-532 laser module with a wavelength of 532.0 nm, a Holoeye Pluto phase-only SLM with a pixel pitch of 8.0 μm, and a micro-prism-based lightguide (LLVision LEION) as the optical combiner. All images are captured
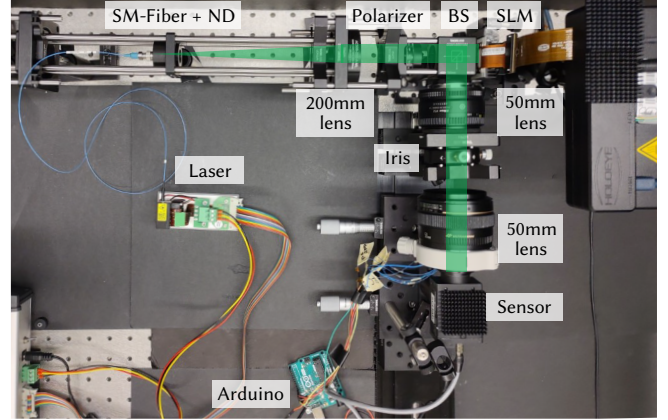
Fig. 3. Prototype VR display. An RGB laser module is coupled to a single mode (SM) fiber to illuminate our phase-only SLM through a neutral density (ND) filter and a polarizer. A beam splitter (BS) redirects the phase-modulated light, through a lens and optical filter, towards the camera. The camera lens provides programmable focus settings, which are controlled by an Arduino.

with a FLIR Grasshopper3 2.3 MP color USB3 vision sensor through a Canon EF 35mm lens for the AR setup. Additional details of these prototypes are included in the supplement.

For all 2D holographic image results shown in the paper and supplement, we positioned the target plane on which the image is optically recorded at a distance of 4.4 mm from the SLM, which corresponds to a distance of 1.26 D, or 0.80 m, from the camera. For all 3D holographic image results, we positioned 8 target planes, on which the multiplane images are shown and optically recorded, equally spaced in dioptric space throughout a range of 0–2 D from the camera. The inter-plane distance of 0.31 D perceived by a user thus corresponds to the depth of field of the human eye [Campbell 1957; Marcos et al. 1999] and can therefore be considered approximately continuous in depth. We measured these distances to correspond to 0.0, 1.1, 2.1, 3.2, 4.4, 5.7, 7.0, and 8.2 mm away from the SLM physically.

4.2 Assessing 2D Holographic Display Modes

Figures 4 and S10–S12 show several test images that were experimentally captured. We compare results achieved by the following methods: an SGD-based phase retrieval algorithm using the ASM wave propagation model (SGD-ASM), the model-based approach proposed by Chakravarthula et al. [2020] (SGD-HIL), the model-based approach proposed by Peng et al. [2020] (SGD-NH), the camera-in-the-loop approach proposed by Peng et al. [2020] (CITL-ASM), and SGD used with the proposed model (SGD-CNNpropCNN). We see that our model provides the best contrast, sharpness, lack of speckle artifacts, and overall image quality. Quantitatively, our approach greatly improves upon other methods – about 2 dB of peak signal-to-noise-ratio (PSNR) over CITL-ASM and 3–4 dB over other models. Supplemental Tables S1, S2 confirm these improvements for a large number of test images.
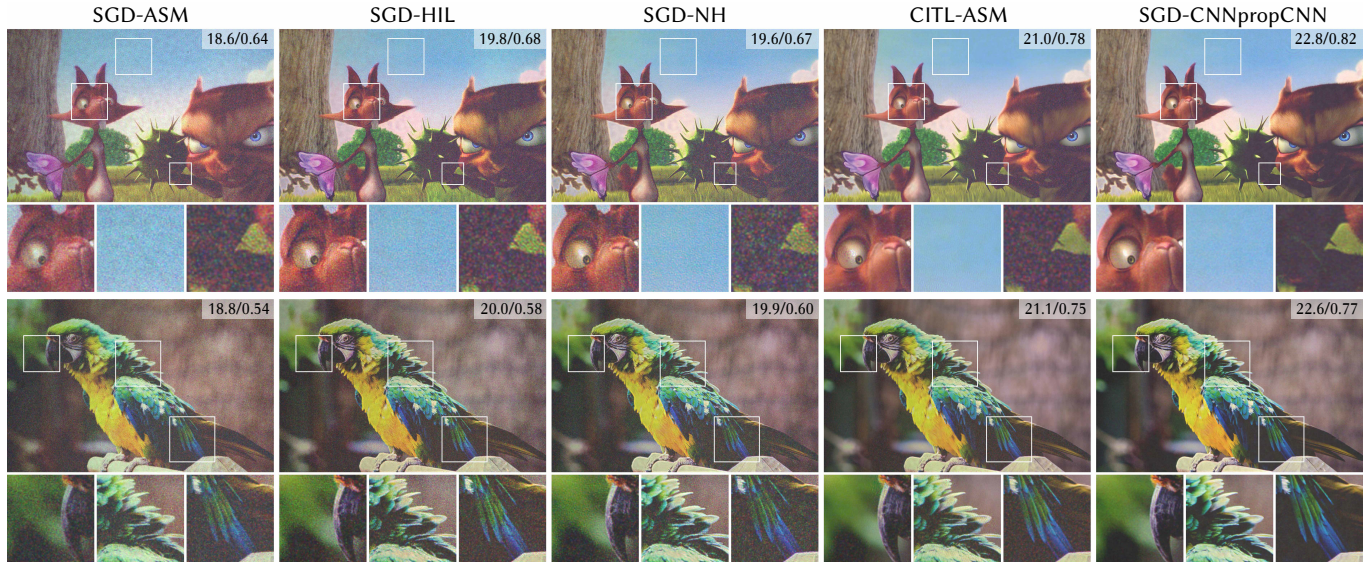
Fig. 4. Comparison of 2D wave propagation models using experimentally captured data. We compare an SGD solver used with the conventional ASM, hardware-in-the-loop model (HIL) and neural holography (NH) models, the camera-in-the-loop CGH optimization approach (CITL-ASM), and our proposed model (SGD-CNNpropCNN). Our model results in sharper images with higher contrast and less speckle than other models under the same experimental conditions. Quantitative evaluations for each result are included as PSNR/SSIM.

4.3 Assessing 3D Holographic Display Modes

Figures 5 and S13–S21 show experimentally captured results of several multiplane 3D scenes, each focused at a near, an intermediate, and a far distance. We compare experimental results with our implementation of the double phase–amplitude coding method [Maimone et al. 2017] (DPAC, see supplement for implementation details), a multiplane SGD optimization that uses ASM wave propagation (SGD-ASM), and the proposed multiplane 3D model used with an SGD solver (SGD-CNNpropCNN) without phase constraints (see Sec. 3.4) and the same model used with an ADMM solver (ADMM-CNNpropCNN) that enforces piecewise smooth phase constraints of the in-focus multiplane images (see Sec. 3.5). In these results, we see that our implementation of DPAC shows overall reasonably good quality for in-focus (red boxes) and out-of-focus (white boxes) parts of the scenes, although the contrast is somewhat low[2]. The SGD-ASM solver significantly improves the contrast over DPAC, but it is much more noisy in both in-focus and out-of-focus image regions (center left column). The proposed model adequately models the wave propagation from the SLM to all target planes and a multiplane SGD solver that constraints the in-focus parts of the target image achieves a very good image quality with significantly reduced speckle and better image quality in these in-focus parts (center right column). However, because the out-of-focus behavior is unconstrained, as the wave field propagates away from the constrained in-focus parts, its unconstrained out-of-focus behavior results in significant out-of-focus speckle. Using the proposed

piecewise smooth in-focus phase constraints mitigates this out-of-focus speckle and results in the best in-focus and out-of-focus image quality (right column). Note that multiplane SGD methods do not suffer from light leakage (i.e., boundary artifacts from content at multiple depths) since the phase is optimized to simultaneously produce in-focus content at all depths. In the supplement, we show extensive evaluations, comparisons, and additional ablations of 3D multiplane CGH methods for these and additional 3D scenes.

4.4 Ablation Study

The previous subsections evaluate the performance of our model for CGH reconstruction tasks. With Table 1, we quantitatively evaluate and ablate the performance of our model in accurately predicting the physical optical wave propagation behavior in various scenarios. The upper block shows three different 2D variants of our model trained for a single target plane. We compare these variants to ASM, HIL, and NH. On the training set, our model outperforms both HIL and NH by more than 8 dB PSNR, which is a significant improvement. To demonstrate that this is not due to overfitting, we evaluate the 2D model performance using a test set of 1,100 images (see supplement). The improvement of our model over these previous models is still about 7 dB for the test set of unseen images, confirming that our model is indeed more accurate and generalizes from training to test set. When evaluating these methods on another target plane that the model was not trained on, which is not possible for HIL, we see that the performance is best when our model only uses the CNN on the SLM plane, CNN_{SLM}. Either variant that uses a CNN on the target plane, $\text{CNN}_{\text{target}}$, overfits to that plane and does not generalize well to other target planes. Note that HIL has about the same number

---

[2]Our implementation of DPAC followed the description of two recent papers [Maimone et al. 2017; Shi et al. 2021] but, despite our best efforts, we were not able to achieve the same quality of results these authors demonstrated. This may be due to slight differences in the hardware setup, which (if mitigated) would likely improve the results of all methods.
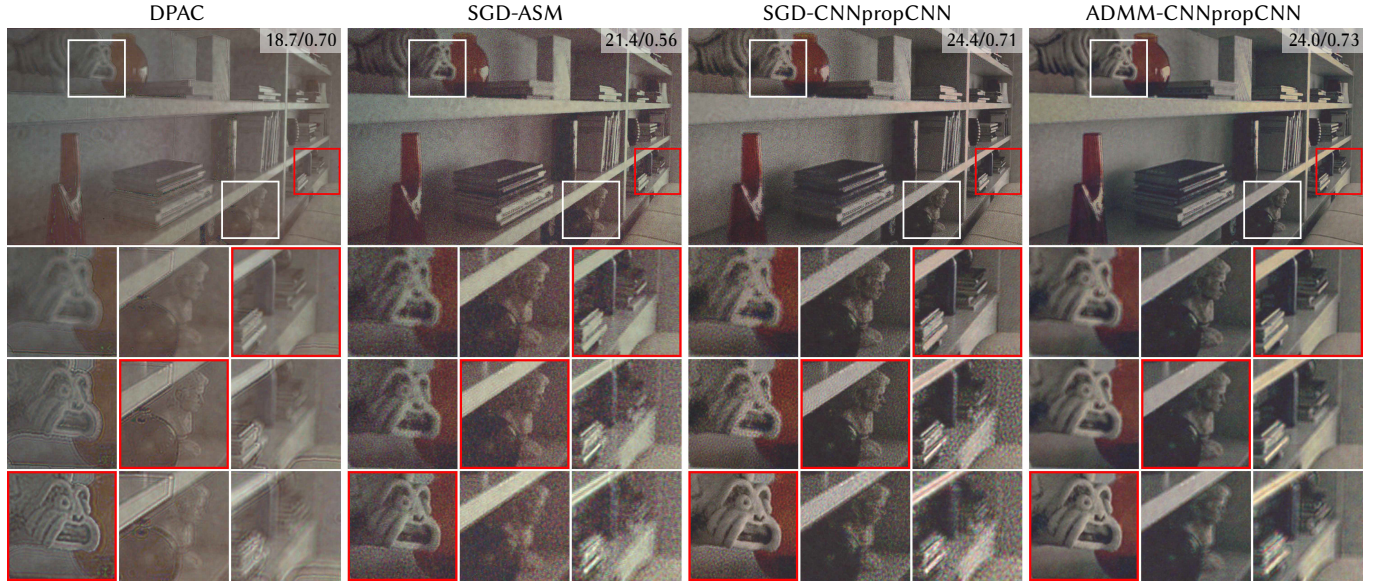
Fig. 5. Comparison of 3D CGH methods using experimentally captured data. We compare the DPAC algorithm (left), a multiplane SGD solver using the ASM model (center left), and two variants of our multiplane 3D model. The first variant uses an SGD solver but only constrains in-focus scene parts, resulting in good image quality in those regions but significant out-of-focus speckle artifacts (center right). The same model used with an ADMM solver that promotes piecewise smooth phases for the in-focus parts of the scene exhibits very good image quality for both in-focus and out-of-focus parts (right). Quantitative evaluations for each result are included as PSNR/SSIM.

of parameters as our CNNpropCNN model (~65M), but NH uses significantly fewer (~8M).

This observation motivates us to train our model using multiplane supervision, which is shown in the center block of Table 1. Here, we train multiplane 3D variants of our model trained on 7 of the 8 planes. We intentionally left out the third plane from the SLM at 0.6 D or 2.1 mm from the training procedure. Again, our CNNpropCNN model shows the best quality for both training and test sets. Evaluating the performance of our multiplane model variants on the held-out plane, which was not part of the training, demonstrates that our 3D model generalizes well to planes in between the training planes. The same trends are also observed when using the same multiplane model and evaluating its performance for all 7 target planes for both training and test set (lower block of Tab. 1). Overall, these experiments demonstrate that the 2D variant of our model is more accurate than previously proposed wave propagation models, which are limited to 2D settings, and ours is also the only one that generalizes to 3D multiplane settings. Moreover, we evaluate several variants of our model and note that a CNN correcting the wave field on the SLM is necessary for good generalization behavior in between target planes of the training set. Any corrections that this CNN performs on the SLM plane, however, are shift invariant on the target plane due to the shift invariance of the ASM propagation operator that follows. Using the additional CNN on the target planes helps correct for shift varying artifacts, optimizing image quality when combined with the CNN on the SLM.

Table 1. Comparison of different models on the captured dataset. Top: all models are trained on a single intensity plane with the training set; the PSNR is evaluated on training and test sets as well as for a held-out plane. Center: model trained on 7 of the 8 intensity planes; PSNR is evaluated on a single plane for training and test sets as well as for a held-out plane. Bottom: model trained on 7 of the 8 intensity planes; PSNR is evaluated on all 7 planes for training and test sets as well as for the 8th held-out plane.

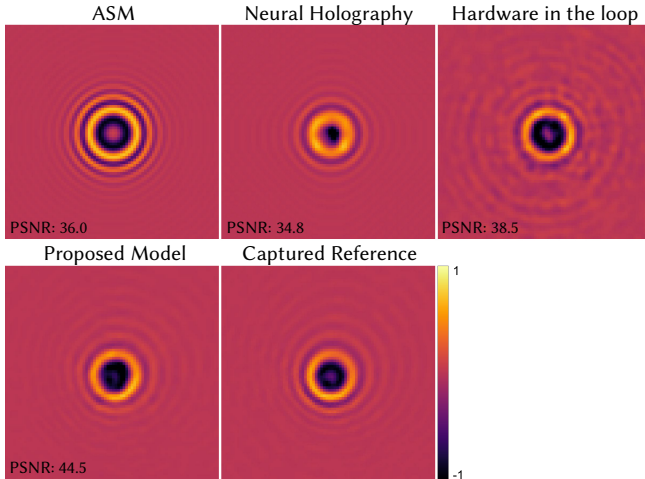|	Model	Training Set	Test Set	Held-out Plane
Tr. 1 Plane Ev. 1 Plane	ASM	21.0	21.0	20.8
	HIL	31.6	31.1	—
	NH	31.8	31.4	25.2
	propCNN	32.6	32.4	25.7
	CNNprop	38.5	37.5	**31.1**
	CNNpropCNN	**40.0**	**38.8**	26.0
Tr. 7 Planes Ev. 1 Plane	ASM	21.0	21.0	20.8
	propCNN	32.1	32.0	29.3
	CNNprop	35.1	35.2	33.0
	CNNpropCNN	**37.9**	**37.9**	**33.1**
Tr. 7 Planes Ev. 7 Planes	ASM	20.9	21.0	20.8
	propCNN	32.1	31.9	29.3
	CNNprop	35.2	34.9	33.0
	CNNpropCNN	**37.8**	**37.6**	**33.1**

Fig. 6. Comparing wave propagation model gradients. We show simulated gradients for the ASM, NH, HIL, and the proposed models as well as a gradient of the physical optical system captured using the finite differences method. Our model provides the best approximation to the physical gradient.
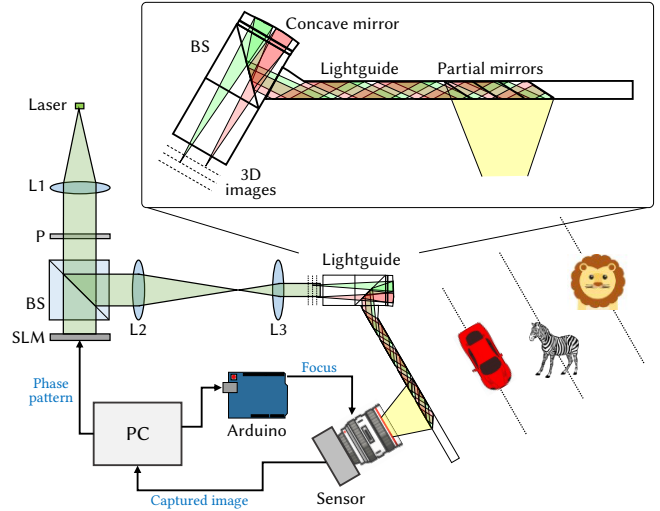

Fig. 7. Illustration of our OST-AR holographic display. The holographic light engine, including a laser, lenses (L1, L2, L3), a beam splitter (BS), polarizer (P), and SLM create a 3D multiplane image simultaneously at three target planes. These are coupled into a lightguide and are optically in focus with physical objects at a near, intermediate, and far distance. The sensor, SLM, and focus mechanism of the camera lens are synchronized in hardware.

4.5 Understanding Wave Propagation Models

A good wave propagation model for holographic display applications should satisfy two criteria. First, it should model the mapping from SLM phase to intensity values observed on the target planes well. We confirmed that this is the case for our model, but not for previous models, in the last subsection. Second, for a model to be effective at generating phase patterns of target images never seen during training, an iterative CGH algorithm like SGD has to be able to backpropagate through this model. Thus, the gradients of the model should match those of the unknown physical optics. In this section, we provide a best attempt to make the gradients of all wave propagation models more intuitive by visualizing and analyzing them.

Specifically, we note that the gradients of the physical optics can actually be captured using a camera with the finite difference method. This general idea is not new and has, for example, recently been explored to optically differentiate structured illumination systems [Chen et al. 2020]. For our application, we record camera images of two SLM phase patterns that only differ in one location and calculate their difference. This is shown in Figure 6 (lower right) where an image captured using a phase pattern of all zeros is subtracted from an image captured with the center block of 3×3 pixels set to π. Note that just changing one pixel value resulted in a signal-to-noise ratio that was too low to be visualized. Comparing this captured gradient to the corresponding gradient of the ASM model (Fig. 6, upper left) reveals that there is a significant gap between the simulation and the physics. We show these gradients also for the NH and HIL models. Interestingly, the gradient of the NH model is actually worse than ASM, even though it models the forward mapping from phase to target intensity better (Tab. 1). Our CNNpropCNN model approximates the captured gradient the best among all of these models.

We note that gradients of the physical wave propagation captured by the finite difference method require long exposure times. Therefore, capturing such gradients for each of the SLM pixels and potentially using them for supervising the model during training seems infeasible. Yet, these captured gradients provide an excellent tool for analyzing the gradients of any model and comparing them to the physical optics. For a fair comparison, we also computed the gradients of all models using the finite differences method and not using automatic differentiation.

5 TOWARDS 3D SEE-THROUGH AR HOLOGRAPHIC DISPLAYS

In this section, we present preliminary results achieved by an OST-AR prototype. The primary difference to the VR setup is that the holographic image is optically routed through a lightguide that allows for the holographic image to be superimposed on a physical scene. No lightguide we are aware of is specifically designed for a holographic display or the étendue our prototype provides, so we do not expect to observe as high of an image quality as for the VR setup. Moreover, lightguides and waveguides are typically designed for only a single target image distance, usually at optical infinity. One could try to physically actuate a conventional (incoherent) microdisplay along the optical axis to implement a varifocal display mode, but this would practically lead to optical aberrations and degraded image quality. In theory, a hologram could potentially correct for some of these aberrations and, when used with an appropriate optical combiner, correct for some of the optical aberrations of the downstream optics. We explore this idea in the following.
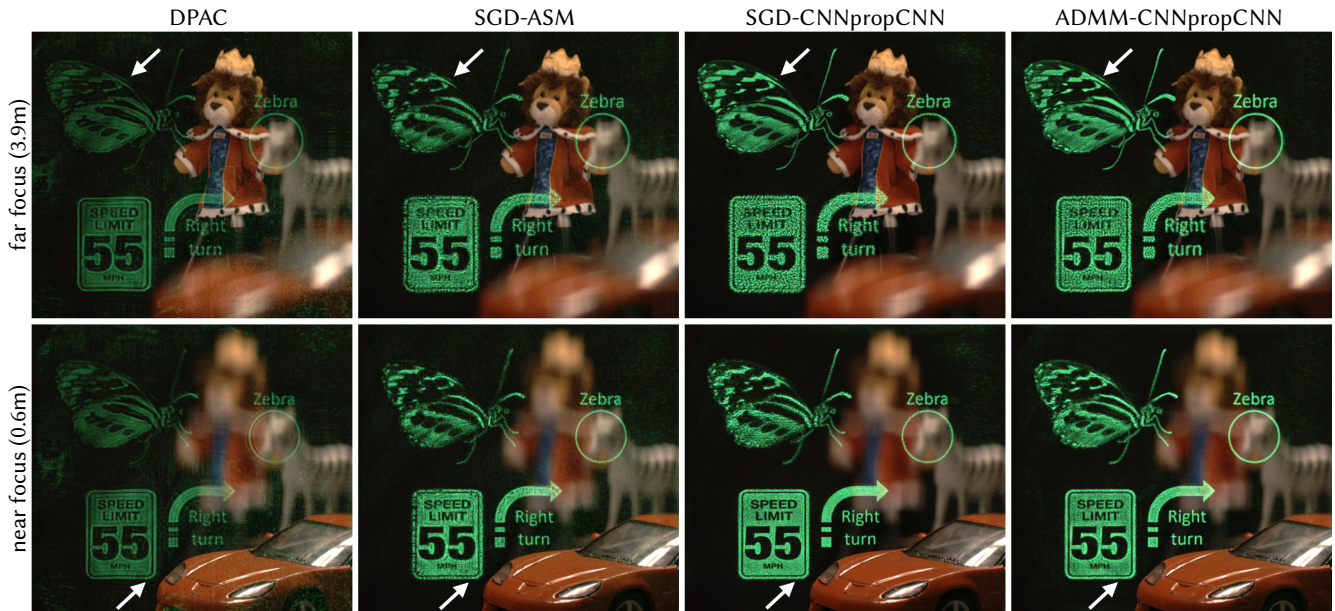
Fig. 8. Experimental results for an OST-AR system showing a comparison of 3D CGH algorithms with the lens focused at two different depths. We compare our implementation of the DPAC algorithm (left), SGD-ASM (center left), the SGD-CNNpropCNN method that uses the proposed multiplane wave propagation model (center right), and our ADMM-CNNpropCNN approach (right). White arrows indicate virtual objects that are focused at a particular depth. We see that algorithms using our wave propagation model perform better than those not using it. The SGD-CNNpropCNN method achieves the best in-focus results with ADMM-CNNpropCNN improving the out-of-focus content. Results for additional focus settings are shown in the supplement.

Specifically, we focus a monochromatic version of our holographic display into a lightguide with built-in horizontal exit pupil expansion via micro prisms (see Secs. 4.1 and S1 for details). We then change the distance of the holographic image from the SLM to create a best-focused image at three distances from the camera: 0.6 m, 1.2 m, and 3.9 m. We train our wave propagation model as described in Section 3.3 for these three target planes. Figure 7 shows an illustration of our OST-AR system and the configuration of the scene, including the car, zebra, and lion at near, intermediate, and far distances. Additional system details are included in the supplement.

Figure 8 shows experimentally captured results. We show the scene at two of the three focus settings and compare the results of several 3D CGH algorithms, including our implementation of DPAC, SGD-ASM, SGD-CNNpropCNN, and ADMM-CNNpropCNN. As with our VR results, SGD-CNNpropCNN, which uses the proposed camera-calibrated multiplane wave propagation model, demonstrates the best image quality for in-focus objects, such as the butterfly in the far-focus setting and the traffic sign and arrow in the near-focus setting. Also as before, ADMM-CNNpropCNN further improves the out-of-focus image quality although we observe a slightly increased amount of speckle artifacts.

6 DISCUSSION

In summary, we propose a new wave propagation model for holographic near-eye displays. Our model is parameterized by neural networks that are automatically trained using footage captured from a physical optical system using camera feedback. Our model significantly outperforms related techniques in 2D plane-to-plane

settings and it is the first such model to naturally extend to 3D plane-to-multiplane settings, enabling high-quality 3D computer-generated holography. We demonstrate that the 3D variant of our model can be directly supervised with RGBD target images, which makes our approach not only computationally efficient but also compatible with both computer-generated and readily-available cinematic content. To constrain the out-of-focus behavior of our 3D holograms, which is not directly constrained by the RGBD images, we propose a regularization strategy of the phase components of the wave field when they are in focus as well as optimizers to enforce this regularization efficiently. The efficacy of our approach is demonstrated to outperform existing methods through extensive experimental evaluation with both VR and optical see-through AR prototype displays.

Limitations and Future Work. Our approach is not without limitations. It focuses primarily on developing accurate and efficient neural network–parameterized wave propagation models for holographic near-eye displays, but not real-time holographic image synthesis. Although it was recently shown that the latter can be achieved [Eybposh et al. 2020; Horisaki et al. 2018; Lee et al. 2020; Peng et al. 2020; Shi et al. 2021], and that differentiable wave propagation models like ours can directly be used as part of the loss function for training CGH networks [Peng et al. 2020], we did not attempt this in our work.

Our prototypes use state-of-the-art phase-only SLMs, but these unfortunately only offer a very limiting étendue to the downstream

optics. In practice, this implies that the eyebox of our and other holographic near-eye displays is small. Higher-resolution SLMs, pupil steering [Jang et al. 2017], or étendue expansion [Kuo et al. 2020] could help mitigate this limitation. The limited étendue also implies that the depth of field observed by a user or camera is limited and may not match that of a physical scene. For OST-AR systems, such as the one shown in Section 5, this results in a noticeable mismatch between the refocusing effects of the physical scene and the superimposed holographic 3D scene. Future work could potentially address this by supervising the hologram generation using a 3D focal stack, rather than RGBD images, but this would come at the expense of increased computational complexity. Direct supervision by focal stacks may also reduce the remaining out-of-focus speckle and improve the overall image quality, particularly in the OST-AR scenario. However, focal stack supervision may also exceed the degrees of freedom of the SLM because it makes the inverse problem overconstrained.

The phase regularization strategy proposed in Section 3.5 proves very valuable in our VR prototype and noticeably improves out-of-focus behavior without sacrificing in-focus image quality. For the OST-AR prototype, it also worked well but it may be challenging to optically create flat phase fronts at all of our target planes due to the design of the lightguide. As demonstrated with our SGD-CNNpropCNN approach in that scenario, it is indeed possible to create high-quality intensities at all target planes simultaneously, which is a significant benefit over conventional (incoherent) microdisplays and thus proves the point of our preliminary experiments. However, simultaneously creating piecewise smooth phase fronts, as encouraged by our phase regularization, may be challenging in this scenario. It is therefore interesting to explore alternative optimization approaches for holographic 3D OST-AR systems in the future, such as the computationally more intensive focal stack supervision approach discussed above.

Optically recorded holograms are remarkable in being able to show view-dependent effects, such as specular highlights and semitransparent materials, which substantially add to the perceived realism. Our approach uses a 3D multiplane approach for the wave propagation model and also for optimizing the holograms. Although recent light field holographic displays have the potential to support these types of effects [Padmanaban et al. 2019], by definition, a multiplane approach does not support them. We argue that the limited eyebox size afforded by current-generation holographic near-eye displays may not require an accurate light field to be synthesized over the exit pupil to make these displays perceptually realistic. Yet, detailed user studies on how perceptually realistic holographic near-eye displays really are remains an exciting avenue of future work.

Finally, our prototype VR and AR displays are benchtop systems and not wearable. We did not attempt to miniaturize these systems, although recent work has demonstrated that this is possible in certain optical configurations [Maimone et al. 2017].

7 CONCLUSION

Holographic near-eye displays are a promising technology with the potential to address many long-standing challenges in augmented and virtual reality systems. With this work, we take steps towards enabling high-quality 3D computer-generated holography by combining modern artificial intelligence–driven methods with physics-based models in these emerging applications.

REFERENCES

Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *CVPR*.

Francis Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski. 2012. Optimization with Sparsity-Inducing Penalties. 4, 1 (2012), 1–106.

Yosuke Bando, Henry Holtzman, and Ramesh Raskar. 2013. Near-invariant blur for depth and 2D motion via time-varying light field analysis. *ACM Trans. Graph.* 32, 2 (2013), 1–15.

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

Stephen A. Benton and V. Michael Bove. 2008. *Holographic Imaging*. Wiley-Interscience.

Fergus W Campbell. 1957. The depth of field of the human eye. *Optica Acta: International Journal of Optics* 4, 4 (1957), 157–164.

Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. 2019. Wirtinger Holography for Near-eye Displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 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. Graph. (SIGGRAPH Asia)* 39, 6 (2020), 186.

Chenliang Chang, Kiseung Bang, Gordon Wetzstein, Byounghyo 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. *OSA 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. *OSA 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.

Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N Kutulakos. 2020. Auto-Tuning Structured Light by Optical Stochastic Gradient Descent. In *CVPR*. 5970–5980.

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

Jennifer E Curtis, Brian A Koss, and David G Grier. 2002. Dynamic holographic optical tweezers. *Optics communications* 207, 1-6 (2002), 169–175.

Rainer G Dorsch, Adolf W Lohmann, and Stefan Sinzinger. 1994. Fresnel ping-pong algorithm for two-plane computer-generated hologram display. *OSA Applied optics* 33, 5 (1994), 869–875.

M. Hossein Eybposh, Nicholas W. Caira, Mathew Atisa, Praneeth Chakravarthula, and Nicolas C. Pégard. 2020. DeepCGH: 3D computer-generated holography using deep learning. *Opt. Express* 28, 18 (2020), 26636–26650.

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

Qiankun Gao, Juan Liu, Jian Han, and Xin Li. 2016. Monocular 3D see-through head-mounted display via complex amplitude modulation. *OSA Opt. Express* 24, 15 (2016).

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. 2005. *Introduction to Fourier optics*. Roberts and Company.

Oscar Hernandez, Eirini Papagiakoumou, Dimitrii Tanese, Kevin Fidelin, Claire Wyart, and Valentina Emiliani. 2016. Three-dimensional spatiotemporal focusing of holographic patterns. *Nature communications* 7, 1 (2016), 1–11.

Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. 2018. Deep-learning-generated holography. *Appl. Opt.* 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.

Hong Hua and Bahram Javidi. 2014. A 3D integral imaging optical see-through head-mounted display. *Optics express* 22, 11 (2014), 13484–13491.

Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graph. (SIGGRAPH)* 34, 4 (2015), 60.

Changwon Jang, Kiseung Bang, Gang Li, and Byoungho Lee. 2018. Holographic Near-eye Display with Expanded Eye-box. *ACM Trans. Graph. (SIGGRAPH Asia)* 37, 6 (2018).

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).

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

Grace Kuo, Laura Waller, Ren Ng, and Andrew Maimone. 2020. High Resolution étendue expansion for holographic displays. *ACM Trans. Graph. (SIGGRAPH)* 39, 4 (2020).

Douglas Lanman and David Luebke. 2013. Near-eye light field displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 32, 6 (2013), 220.

Juhyun Lee, Jinsoo Jeong, Jaebum Cho, Dongheon Yoo, Byounghyo Lee, and Byoungho Lee. 2020. Deep neural network for multi-depth hologram generation and its training strategy. *Opt. Express* 28, 18 (2020), 27137–27154.

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

Gang Li, Dukho Lee, Youngmo Jeong, Jaebum Cho, and Byoungho Lee. 2016. Holographic display for see-through augmented reality using mirror-lens holographic optical element. *OSA Opt. Lett.* 41, 11 (2016), 2486–2489.

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

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).

Susana Marcos, Esther Moreno, and Rafael Navarro. 1999. The depth-of-field of the human eye from objective and subjective measurements. *Vision research* 39, 12 (1999), 2039–2049.

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.

Eunkyong Moon, Myeongjae Kim, Jinyoung Roh, Hwi Kim, and Joonku Hahn. 2014. Holographic head-mounted display with RGB light emitting diode light source. *OSA Opt. Express* 22, 6 (2014), 6526–6534.

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

Eirini Papagiakoumou, Francesca Anselmi, Aurélien Bègue, Vincent De Sars, Jesper Glückstad, Ehud Y Isacoff, and Valentina Emiliani. 2010. Scanless two-photon excitation of channelrhodopsin-2. *Nature methods* 7, 10 (2010), 848–854.

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 Transactions on Graphics (TOG)* 39, 6 (2020), 1–14.

Yifan Peng, Xiong Dun, Qilin Sun, and Wolfgang Heidrich. 2017. Mix-and-match holography. *ACM Trans. Graph.* 36, 6 (2017), 191.

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

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. (SIGGRAPH Asia)* 36, 6, Article 236 (2017), 236:1–236:17 pages.

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

Maxim Shusteff, Allison E. M. Browar, Brett E. Kelly, Johannes Henriksson, Todd H. Weisgraber, Robert M. Panas, Nicholas X. Fang, and Christopher M. Spadaccini. 2017. One-step volumetric additive manufacturing of complex polymer structures. *Science Advances* 3, 12 (2017). https://doi.org/10.1126/sciadv.aao5496

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. 2016. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022* (2016).

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.

Samuel J. Yang, William E. Allen, Isaac Kauvar, Aaron S. Andalman, Noah P. Young, Christina K. Kim, James H. Marshel, Gordon Wetzstein, and Karl Deisseroth. 2015. Extended field-of-view and increased-signal 3D holographic illumination with time-division multiplexing. *Opt. Express* 23, 25 (Dec 2015), 32573–32581.

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.

Han-Ju Yeom, Hee-Jae Kim, Seong-Bok Kim, HuiJun Zhang, BoNi Li, Yeong-Min Ji, Sang-Hoo Kim, and Jae-Hyeung Park. 2015. 3D holographic head mounted display using holographic optical elements with astigmatism aberration compensation. *OSA Opt. Express* 23, 25 (2015), 32025–32034.

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*. 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.