

# Learned Hardware-in-the-loop Phase Retrieval for Holographic Near-Eye Displays

PRANEETH CHAKRAVARTHULA, UNC Chapel Hill  
ETHAN TSENG, TARUN SRIVASTAVA, Princeton University  
HENRY FUCHS, UNC Chapel Hill  
FELIX HEIDE, Princeton University

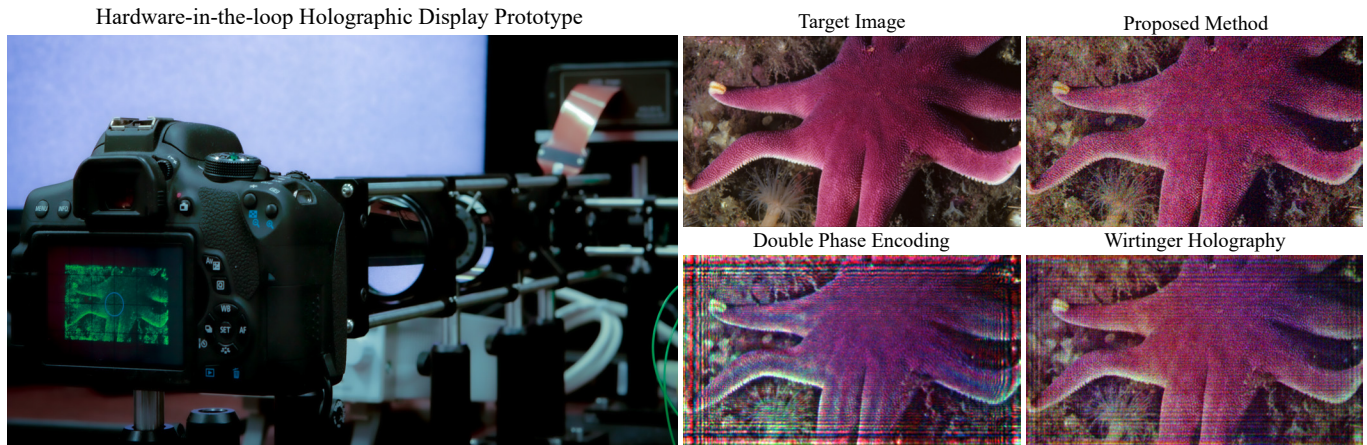


Fig. 1. *Hardware-in-the-loop Phase Retrieval*. Holographic displays often show poor image quality due to severe real world deviations in the light propagation compared to the simulated “ideal” light propagation model. We built a holographic display-camera setup (left) to generate data that is used to train a neural network for approximating the *unknown light propagation* in a real display and the resulting aberrations. We then use this trained network to compute phase holograms that compensate for real world aberrations in a hardware-in-the-loop fashion. This allows us to supervise the display states with unknown light propagation just by observing image captures of the display prototype (left). Compared to holographic images captured using state-of-the-art methods, Double Phase Encoding [Maimone et al. 2017] (bottom center) and Wirtinger Holography [Chakravarthula et al. 2019] (bottom right), the proposed method (top right) produces images on real hardware that are aberration-free and close to the target image (top center).

Holography is arguably the most promising technology to provide wide field-of-view compact eyeglasses-style near-eye displays for augmented and virtual reality. However, the image quality of existing holographic displays is far from that of current generation conventional displays, effectively making today’s holographic display systems impractical. This gap stems predominantly from the severe deviations in the idealized approximations of the “unknown” light transport model in a real holographic display, used for computing holograms.

In this work, we depart from such approximate “ideal” coherent light transport models for computing holograms. Instead, we learn the deviations of the real display from the ideal light transport from the images measured using a display-camera hardware system. After this unknown light propagation is learned, we use it to compensate for severe aberrations in real holographic imagery. The proposed hardware-in-the-loop approach is robust to spatial, temporal and hardware deviations, and improves the image quality of existing methods qualitatively and quantitatively in SNR and perceptual

quality. We validate our approach on a holographic display prototype and show that the method can fully compensate unknown aberrations and erroneous and non-linear SLM phase delays, without explicitly modeling them. As a result, the proposed method significantly outperforms existing state-of-the-art methods in simulation and experimentation – just by observing captured holographic images.

CCS Concepts: • **Computing methodologies** → **Computational photography**.

## ACM Reference Format:

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.* 39, 6, Article 186 (December 2020), 18 pages. <https://doi.org/10.1145/3414685.3417846>

## 1 INTRODUCTION

Personal displays play an essential role in how we interact with smart devices and the immediate environment, including diverse applications in communication, entertainment, medical assistance, navigation, and in emerging environments such as self-driving vehicles. As an emerging next personal computing and social interaction platform, virtual reality (VR) and augmented reality (AR) promise such applications. Indeed, existing consumer products offer wearable displays with an acceptable form factor, resolution and field

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. 0730-0301/2020/12-ART186 \$15.00

<https://doi.org/10.1145/3414685.3417846>

of view, which make a wide variety of compelling applications in VR possible. However, AR displays still remain limited to large hardware, suffer from low image quality, vergence-accommodation conflict, and they do not allow for occlusion effects. Overcoming these limitations using conventional optics result in large setups. Resorting to approaches such as light field displays promise compact form factors, but is a poor alternative due to the spatio-angular resolution trade-off. For near-eye displays to be widely adopted as a personal and social computing platform, tomorrow's displays must provide wide field of view and high resolution in a small form factor, closely mimicking ordinary pairs of eyeglasses a user can wear for long hours.

Holographic displays, in principle, promise such an eyeglasses-like form factor by shifting the optical complexity to computation. Unlike conventional displays, which need a cascade of refractive optics to physically modulate the emitted wavefront of light, a holographic display produces this combined effect via a digital hologram using a spatial light modulator (SLM). In other words, today's holographic display methods *move the complexity of traditional optics to complexity of the modulation setup and careful regularization*. While theoretically an elegant and promising technology, holographic setups, in practice, suffer from *severe non-linearities* in the phase modulation, inhomogeneity of the SLM itself, and imperfections in the minimal optics and illumination setup.

Recent works have achieved impressive results by *careful hand-tuning of these non-linearities* [Maimone et al. 2017] in a light-weight wearable form-factor. However, the achieved image quality still does not approach conventional displays. An underlying fundamental limitation of the existing methods is the *approximation of the light propagation model*, which is generally inadequate in describing the experimental setup. Moreover, the forward models are often deliberately simplified with various assumptions and approximations to make the underlying phase retrieval problem tractable.

In this work, *we depart from such approximate forward models and propose a hardware-in-the-loop method to learn the non-linear image formation model of a real holographic projection system*. Relying on an initial phase estimate computed from an approximate forward model that does not consider the hardware non-linearities, we learn the mapping between the image generated by the holographic projector with *unknown light transport*, and the intended ideal reconstruction. These mappings serve as upper bounds on the unknown objective function (and our forward model) that we aim to minimize. We validate the proposed method both in simulation and on an experimental prototype and demonstrate that our improved approach eliminates severe artifacts present in existing approaches. We assess these capabilities qualitatively and quantitatively on representative test data, and we verify that the proposed method generalizes and is robust to hardware, spatial and temporal drift.

In particular, we make the following contributions:

- We introduce a method for the estimation of unknown light propagation models in holographic display forward models.
- The proposed approach relies on a *learned hardware-in-the-loop* optimization which iteratively refines an upper bound on the estimated objective function derived in this work.

- We validate the proposed method by solving phase retrieval for holographic projections with constant focus over the image. The resulting holographic reconstructions improve by more than 10 dB in simulation and 2.5 dB in PSNR on the hardware prototype.
- We assess the proposed framework experimentally with a prototype near-eye holographic display setup. The proposed method reduces severe artifacts of existing holographic display approaches, which we quantify with perceptual and SNR metrics.

*Overview of Limitations.* Although the proposed method achieves unparalleled image quality and predicts real-world aberrations at real-time frame rates, applying the method may require end-of-manufacturing-line per-device training data to account for optical fabrication tolerances. However, this would only be a one-time calibration step which we envision to be automated for production of future holographic eyeglasses-style displays and automotive heads-up displays. Similar to aberration calibration for smartphone camera modules manufacturing, this process would require a custom display and capture stage after the fabrication of each device with the proposed hardware-in-the-loop optimization method, producing a learned aberration approximation network that might be embedded in the firmware of the device. While the proposed method may have the potential to generalize per device type, evaluating this potential capability would require access to production samples with representative tolerances, which is out of scope.

## 2 RELATED WORK

In this section, we review relevant holographic display technologies and computer-generated holography (CGH) algorithms.

### 2.1 Holographic Near-Eye Displays

Holographic displays have been explored as elegant approaches to variable focus control and aberration correction, which are essential features for eyeglasses-style displays. Several recent near-eye display designs employ holographic projectors and/or holographic optical elements (HOE) to achieve a compact form factor. While most of the designs rely on a single phase-only spatial light modulator (SLM) [Chen and Chu 2015], configurations using two phase SLMs [Levin et al. 2016], or a combination of both amplitude and phase SLMs [Shi et al. 2017] have also been explored. Although the proposed holographic benchtop prototypes are large and bulky, many recent works have proposed ways forward to achieve compact form factors. For example, multi-functional HOEs [Jang et al. 2019, 2017; Li et al. 2016; Maimone et al. 2017] and waveguides [Yeom et al. 2015] have been explored as techniques that allow relaying the projected imagery into the eye. Two critical limitations of existing holographic near-eye displays are the tiny eyebox and poor image quality. Several recent works have achieved an increased eyebox size by using eyetracking [Jang et al. 2019, 2017]. Eyetracking has also been proposed for focus control and aberration correction [Maimone et al. 2017]. These existing display designs merely differ in hardware implementation details such as the approaches used for optical path folding, eyetracking and component design. In contrast, the algorithms used for generating the holograms, which restrict

the achievable image quality, remain the same or similar across the displays – low image quality remains a critical limitation of today’s holographic displays.

## 2.2 Holographic Display Setups and Limitations

In this section we briefly review holographic display configurations and the most important limitations that restrict image quality in holographic displays.

*SLM technology.* Achieving accurate, complex modulation of both phase and amplitude with a single device remains an open problem [Reichelt et al. 2012]. In recent work, phase-only SLMs are often preferred due to their higher diffraction efficiency. However, these phase-only SLMs require a trillion sub-wavelength sized pixels to display holograms comparable to conventional holograms [Reichelt et al. 2012]. Unfortunately, existing SLMs only have resolutions ranging up to  $3840 \times 2160$  (4K UHD) with pixel pitches limited to approximately  $4\ \mu\text{m}$  and fill factors less than 95%. The space for electronics between the active pixel areas further leads to zero-order undiffracted light which often causes severe artifacts in the holographic images.

*Phase wrapping and quantization errors.* Existing SLMs implement quantized phase modulation with limited bit-depth. Typically the phase values of the computer generated digital hologram are wrapped to lie within  $[0, 2\pi]$ , and are further quantized to the bit-depth of the SLM. The resulting approximation errors introduced by phase wrapping and quantization significantly deteriorate the holographic image quality [Dallas and Lohmann 1972]. Furthermore, the phase modulation and diffraction efficiency of the SLM pixels are dependent on voltage controlled birefringence made available via a calibrated lookup table (LUT). Any inconsistencies in the LUTs cause non-linear phase modulation which further degrade the display quality.

*Coherent Laser Speckle.* Coherent light sources such as single-longitudinal-mode lasers are sources whose coherence length is large. These sources produce coherent speckle noise. Recent work aims at reducing speckle by using rotating diffusers [Bianco et al. 2016], modulating or quickly repositioning the laser beam [Kang 2008] or superposition of multiple reconstructions [Golan and Shoham 2009]. Coherent laser speckle noise can also be mitigated by using partially coherent light sources, however, this results in blur or loss of depth perception due to reduced coherence length [Dainty 1977]. Although many techniques have been proposed to reduce noise in CGH [Bianco et al. 2018], effective holographic noise suppression still remains an open problem.

## 2.3 CGH Phase Retrieval Algorithms

Holography for displays relies on diffraction and interference of light for generating imagery. Based on the diffracted field, a hologram can be classified as a far-field Fourier hologram or a near-field Fresnel hologram. Using a phase-only SLMs requires computing phase-only holograms that are capable of producing the diffraction field that can closely mimic the target image. This phase retrieval problem is generally ill-posed and non-convex. Though introduced for Fourier phase retrieval, early methods such as error reduction

using iterative optimization [Gerchberg 1972; Lesem et al. 1969] and hybrid input-output (HIO) methods [Bauschke et al. 2003; Fienup 1982] are applicable for both Fourier and Fresnel holograms. Researchers have also explored phase-retrieval methods using first-order non-linear optimization [Fienup 1993; Gonsalves 1976; Lane 1991], alternative direction methods for phase retrieval [Marchesini et al. 2016; Wen et al. 2012], non-convex optimization [Zhang et al. 2017], and methods overcoming the non-convex nature of the phase retrieval problem by lifting, i.e. relaxation, to a semidefinite [Candes et al. 2013] or linear program [Bahmani and Romberg 2017; Goldstein and Studer 2018]. Recently, Chakravarthula et al. [2019; 2020] demonstrated an optimization approach using first-order gradient descent methods to solve for holograms with flexible loss functions. We refer the reader to Barbastathis et al. [2019] for an overview of learned phase retrieval methods.

All of these methods, and the following algorithms, have in common that they assume a perfect image formation model and ignore deviations from the perfect forward model and non-linearities that occur in the real hardware prototype. The proposed method addresses this limitation.

*Point and Polygonal Algorithms.* Successful scene representations in computer graphics model a scene as a collection of points (3D point cloud), an RGB-D image, a polygonal mesh or stacked layers of intensity modulation. Researchers have leveraged such scene representation for computing holograms [Leseberg and Frère 1988; Waters 1966], popularly known as Fresnel holograms. While iterative phase retrieval methods can be applied for computing Fresnel holograms, direct methods in combination with some form of amplitude-phase encoding are much faster [Maimone et al. 2017]. One can also use a look-up table of precomputed elemental fringes to speed up the computation [Lucente 1993; Shi et al. 2017]. Recent point-source based CGH computation methods leverage the parallelization of modern GPUs [Chen and Wilkinson 2009; Masuda et al. 2006; Petz and Magnor 2003]. Instead of computing the wave propagation for millions of points, a 3D object can be represented as a collection of tilted and shifted planes (polygonal mesh), whose diffraction patterns can be computed by the fast Fourier transform (FFT) algorithm [Matsushima 2005; Tommasi and Bianco 1993], also considering texture and shading [Ahrenberg et al. 2008; Matsushima 2005], which is not possible with the point-based hologram computation. Occlusion culling effects can also be provided using geometric facet selection by ray tracing [Kim et al. 2008], silhouette methods [Matsushima and Nakahara 2009; Matsushima et al. 2014] and inverse orthographic projection techniques [Jia et al. 2014]. These methods, however, do not support subtle view-dependent effects such as intra-pupil occlusion.

*Hogel-based Algorithms.* A hologram can be partitioned spatially into elementary hologram patches (hogels), each producing local ray distributions (images) that together reconstruct multiple views supporting intra-ocular occlusions [Lucente and Galyean 1995; Smithwick et al. 2010; Yamaguchi et al. 1993], similar to light field displays [Lanman and Luebke 2013]. These holograms which encode a light field are dubbed “holographic stereograms”. Although conventional stereograms suffer lack of focus cues and limited depth of field [Lucente and Galyean 1995], holographic stereograms

can be paired with point-source methods to enhance the image fidelity to provide improved resolution, accommodation and occlusion cues [Shi et al. 2017; Zhang et al. 2015]. Such light field holograms typically require choosing a specific hogel size, which, in turn, requires trading off spatial and angular resolution. An overlap-add approach was recently proposed to overcome this spatio-angular resolution tradeoff, making more efficient use of information encoded in the light field [Padmanaban et al. 2019].

In contrast to light field based holograms, a further line of research proposes to slice objects at multiple depths and superimpose the wavefronts from each slice on the hologram plane [Bayraktar and Özcan 2010; Zhao et al. 2015], similar to layer-based light field displays [Wetzstein et al. 2012]. Moreover, layer-based and light field methods can both be combined to produce view-dependent occlusion effects [Chen and Chu 2015; Zhang et al. 2016]. The proposed hardware-in-the-loop phase retrieval naturally facilitates layer-based displays due to their similarity in formulating an image loss function, in contrast to hogel-based display which requires explicitly incorporating depth in the loss for computing the holograms.

### 3 COMPUTATIONAL DISPLAY HOLOGRAPHY

Computational display holography aims to replace a real static hologram with a spatial light modulator (SLM) whose states are configurable. To this end, existing methods simulate the process of optical recording and reconstruction of the real hologram using numerical methods. Computing digital holograms can be interpreted as a coherent optical system producing a complex wave field, which is an image of the wave field originally reflected or refracted by the object. Generating such digital holograms for an SLM is possible as long as the sampling theorem is fulfilled: at least two pixels of the SLM sample a fringe period of the hologram. Owing to the limitations in existing SLM modulation, we focus on phase-only holograms in this work which have higher diffraction efficiency. Moreover, we adopt a Fresnel holography regime relevant to near-eye displays [Maimone et al. 2017], although the proposed method is not limited to Fresnel holography configurations.

For a hologram of complex amplitude  $H(\zeta, \eta)$  illuminated by a reference wave field  $E_R(\zeta, \eta)$ , the resulting field in the image plane at a distance  $d$  can be calculated using the scalar (Fresnel) diffraction integral [Goodman 2005]:

$$E_I(x, y) = \frac{1}{j\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(\zeta, \eta) E_R(\zeta, \eta) \frac{\exp(jk\rho)}{\rho} d\zeta d\eta, \quad (1)$$

where  $(\zeta, \eta)$  are the coordinates on the hologram plane,  $(x, y)$  are the coordinates on the image plane,  $k = \frac{2\pi}{\lambda}$  is the wave number and  $\rho = \sqrt{(\zeta - x)^2 + (\eta - y)^2 + d^2}$  is the Euclidean distance between the points on the hologram and image planes, respectively. The resulting complex wave field represents the reconstructed wave field on the image plane, containing both amplitude and phase. Note that a phase-only hologram modulates only the phase of light, and hence has a constant amplitude across the hologram plane  $H(\zeta, \eta) = c \exp(j\Phi(\zeta, \eta))$  where  $c$  is the constant amplitude, which we assume to be unity in the following.

The diffraction integral in Eq. (1) can also be viewed as a superposition integral of waves, thus representing a linear shift-invariant

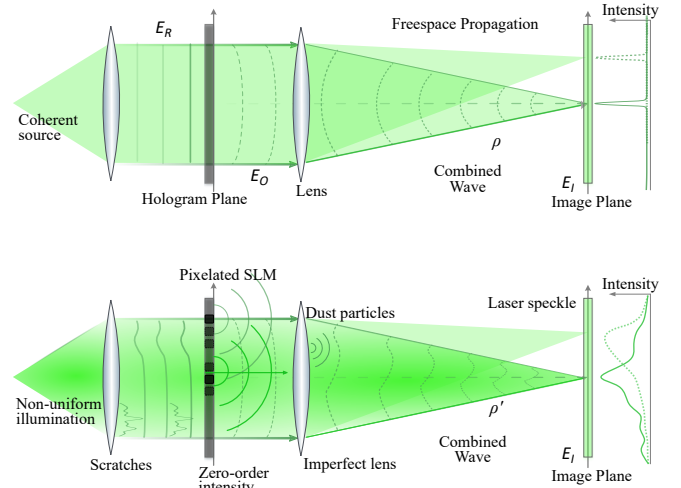


Fig. 2. Illustration of a representative (Fourier) holographic image formation model and its deviations on a real-world prototype. For computing the scalar diffraction integral, light from a coherent illumination source modulated by a continuous hologram aperture is simulated. Evaluating the integral for the given forward model results in the digitally propagated wave field on the image plane (top). However, the diffraction integral no longer holds in a real-world prototype where severe deviations occur due to the illumination source, imperfect optics, pixelated SLM and coherent laser speckle (bottom). Explicitly modeling and calibrating these deviations is often not feasible, see text.

system which can be written as a convolution

$$E_I(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(\zeta, \eta) E_R(\zeta, \eta) \cdot g(x - \zeta, y - \eta) d\zeta d\eta, \quad (2)$$

where the kernel

$$g(\zeta, \eta) = \frac{1}{j\lambda} \frac{\exp \left[ jk\sqrt{d^2 + \zeta^2 + \eta^2} \right]}{\sqrt{d^2 + \zeta^2 + \eta^2}} \quad (3)$$

is the impulse response of free space propagation. This definition is the point-source propagation model from Chakravarthula et al. [2019]. Note that generating the hologram is exactly the inverse process, i.e. propagating the wave field from the image plane to the hologram plane, or equivalently, convolving the conjugate kernel with the image field [Maimone et al. 2017]. One can invoke the convolution theorem to express Eq. (2) as

$$E_I = (H \circ E_R) * g = \mathcal{F}^{-1}(\mathcal{F}[H \circ E_R] \circ \mathcal{F}[g]) \quad (4)$$

where  $\circ$  is the Hadamard element-wise product and  $\mathcal{F}$  is the Fourier transform operator. The above convolution model from Eq. (4) can be reformulated to reflect Fresnel, Fraunhofer, and angular spectrum propagation by modifying the kernel  $g$  with necessary approximations. For these propagation models, please refer to the Supplementary Material.

While the above formulation allows for synthesizing phase-only digital holograms, it assumes a continuous hologram aperture and an aberration free-hardware optical system, all simulated on a computer. However, a real hardware prototype severely deviates from ideal coherent light transport. Next, we discuss several deviations



which affect the final holographic image quality, building towards the underlying *non-ideal* coherent light transport for a phase-only holographic display.

### 3.1 SLM Fill Factor

In contrast to a continuous aperture assumed in the scalar diffraction integral, an SLM is discretized into pixels. For a theoretical fill factor of 100% of the SLM, the pixel size equals the pixel pitch and the SLM would act as a continuous aperture. However, existing SLMs in CMOS technology contain small non-modulating zones between the individual phase modulating pixels, limiting the fill factor. As shown in Figure 2, we can characterize such physical SLMs by the number of pixels  $N$  and  $M$  in  $\zeta$  and  $\eta$  directions, respectively with the pixel pitches  $\Delta\zeta$  and  $\Delta\eta$  and with fill factors in the range  $[0, 1]$ . The transmittance of such an SLM can be modeled as follows:

$$t_{\text{SLM}} = \text{rect}\left(\frac{\zeta}{N\Delta\zeta}, \frac{\eta}{M\Delta\eta}\right) [t_{\text{ap}} + t_{\text{ds}}], \quad (5)$$

where  $t_{\text{ap}}$  is the transmission function of the *active pixel area* displaying the phase pattern of the computed phase-only hologram,  $t_{\text{ds}}$  is the transmission of the *dead space area* of the SLM pixels, and  $\text{rect}\left(\frac{\zeta}{N\Delta\zeta}, \frac{\eta}{M\Delta\eta}\right)$  is the total SLM aperture area. Please refer to the Supplementary Material for details. Inspecting Eq. (5), it becomes clear that the SLM introduces an extra complex amplitude to the hologram, which typically shows up as a zero-order intensity overlay, significantly distorting the reconstructed image pattern.

### 3.2 Non-Linear SLM, Phase Wrapping and Quantization

As discussed in Section 2, the phase modulation of an SLM is controlled by a voltage that is represented by a gray level. The mapping between gray levels and voltage levels can be non-linear and is typically represented by a Lookup Table (LUT). Approximation errors in this LUT results in erroneous phase modulations. Moreover, wrapping the phase from  $2\pi$  back to 0 and the further phase quantization into (often only 8-bit) gray levels causes quantization noise manifesting as severe aberrations. Such errors resulting from the SLM non-linearities and phase representation can be modeled as:

$$\Phi'_{\text{cdh}} = \Phi_{\text{cdh}} + \Delta\phi_{\text{non-lin}} + \Delta\phi_{\text{wrap}} + \Delta\phi_{\text{quant}}, \quad (6)$$

where  $\Phi_{\text{cdh}}$  is the computed display hologram phase pattern from ideal light transport model, whereas  $\Phi'_{\text{cdh}}$  is the represented phase pattern on the SLM effected by deviations due to non-linearities in LUT ( $\Delta\phi_{\text{non-lin}}$ ), phase wrapping ( $\Delta\phi_{\text{wrap}}$ ) and quantization ( $\Delta\phi_{\text{quant}}$ ) errors.

### 3.3 Illumination, Optics and Alignment Errors

A typical digital hologram, when illuminated by a real reference wave, eliminates the reference wave component from the hologram, leaving only the object wave which forms the image field. However, the illuminating reference wave often cannot be accurately replicated in simulation. Furthermore, inconsistencies in the angle between the real and digital reference waves result in tilt phase errors. Such deviations can occur due to misalignment of the SLM and several reflective and/or refractive optics in the display setup.

We express them as

$$H_{\text{illum}} = H \cdot A_{\text{err;illum}} \exp(j\phi_{\text{err;illum}}), \quad (7)$$

where  $A_{\text{err;illum}}$  and  $\phi_{\text{err;illum}}$  model the deviations in amplitude and phase.

### 3.4 Phase Aberrations from Model Approximations

Evaluating a full Fresnel integral (Eq. (1)) is desirable but is often computationally expensive. Relying on Fresnel or Fraunhofer approximations instead to compute digital holograms at reasonably large distances significantly reduces the computational cost and has been proposed before [Peng et al. 2017]. The approximation errors with respect to the full Fresnel integral cause phase aberrations ( $\phi_{\text{err;phs}}$ ) that can be modeled as

$$H_{\text{err;phs}} = H \cdot \exp(j\phi_{\text{err;phs}}). \quad (8)$$

Such phase aberrations can result in curvatures in the reconstructed images. For example, a Fresnel approximation results in a parabolic phase error.

### 3.5 Coherent Noise

A coherent source of light impinging on an SLM also results in the coherent noise, which can manifest itself as blur or grainy speckle. Any reflections within the holographic display setup cause grainy coherent speckle noise due to optically rough surfaces. Non-diffusing transparent objects introduce coherent noise from undesired diffraction and multiple reflections due to dust particles, scratches and defects in and on the optical elements. While such errors are challenging to model, their effect on the wave propagation and *ideal* holographic image formation can be described by the deviation in the computed hologram as

$$H_{\text{err;Cnoise}} = H \cdot A_{\text{Cnoise}} \exp(j\phi_{\text{Cnoise}}), \quad (9)$$

where  $A_{\text{Cnoise}}$  and  $\phi_{\text{Cnoise}}$  are the amplitude and phase of the coherent noise, respectively.

This behavior may result in severe deviations from the ideal coherent light transport from Eq. (1). As the sum and product of complex exponentials is another complex exponential, the effect of all the deviations in a real physical setup can be combined into a single complex exponential describing the aberration wave field  $E_{\text{err}} = A_{\text{err}} \exp(j\phi_{\text{err}})$ .

## 4 HARDWARE-IN-THE-LOOP IN-EXACT PHASE RETRIEVAL

In this section, we describe the proposed holographic phase retrieval method which incorporates hardware deviations from the ideal image formation. We start by modeling the *unknown* coherent light transport of a hologram as a function of the propagated wave field on the image plane. We derive an upper bound of the deviations from the *ideal* light transport (Eq. (1)) and represent this bound by a differentiable *learned aberration approximator* parameterized by a deep neural network. This formulation allows us to cast the holographic computation as a complex non-convex optimization problem, which we solve using first-order optimization methods [Chakravarthula et al. 2019]. We iteratively refine the aberration approximator in the region around the optimum via *hardware-in-the-loop* display

captures. Once learned, the aberration approximator can be used for all future hologram calculations.

#### 4.1 Compensating Phase Patterns

Expressing a complex hologram as a phase-only hologram, the conjugate complex amplitude of all deviations described in the previous section can be consumed into the computed hologram phase pattern as a *compensating phase perturbation*  $\Phi_{\text{dev}}$ , i.e.  $H' = \exp(j[\Phi(\zeta, \eta) + \Phi_{\text{dev}}(\zeta, \eta)])$ . This allows us to reformulate the light transport as

$$E'_I(x, y) = \frac{1}{j\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H'(\zeta, \eta) E_R(\zeta, \eta) E_{\text{err}}(\zeta, \eta) \frac{\exp(jk\rho)}{\rho} d\zeta d\eta \quad (10)$$

where  $H'(\zeta, \eta)$  is the phase-only hologram with *compensating phase* for setup deviations errors, i.e.  $E_{\text{err}}$  introduced in Section 3. In this work, we propose a method to efficiently estimate these model deviations and compute the compensating phase patterns.

#### 4.2 In-exact Phase Retrieval

Consider a phase hologram  $H(\Phi)$  that is propagated using an *ideal* wave propagation function  $\mathcal{P}$ . This propagation results in an observed image wave field given by  $E_I = \mathcal{P}(H(\Phi))$ . However, in practice, the propagation in a real-world holographic display may deviate significantly from the ideal propagation model, see Eq. (10). The resulting image wave field, taking into account both the systemic and content-dependent deviations ( $E_{\text{err}}$ ), can be defined as

$$E'_I = \tilde{\mathcal{P}}(\exp(j\Phi)), \quad (11)$$

where  $\tilde{\mathcal{P}}$  is the deviated light propagation. To design a hologram, we penalize the distance between the reconstructed intensity image  $\tilde{I}(\Phi) = |E'_I(\Phi)|^2$  and target image  $I$  as described by a custom penalty function. For example, for an  $\mathcal{L}_2$  distance, the penalty would be

$$\Phi_{\text{opt}} = \underset{\Phi}{\text{minimize}} \quad \|\tilde{I}(\Phi) - I\|^2. \quad (12)$$

Alternatively, we can formulate the *wave field on the image plane* (Eq. (11)) as a combination of the *ideal non-deviated* field  $\mathcal{P}(H(\Phi))$  and the *aberration* field  $\mathcal{R}(\Phi)$  originating from the real world aberrations. The image intensity then is

$$\begin{aligned} \tilde{I} &= |\mathcal{P}(H(\Phi)) + \mathcal{R}(\Phi)|^2 \\ &= |\mathcal{P}(H(\Phi))|^2 + |\mathcal{R}(\Phi)|^2 + 2|\mathcal{P}(H(\Phi))||\mathcal{R}(\Phi)|\cos(\Delta\theta), \end{aligned} \quad (13)$$

where  $\Delta\theta$  is the phase difference between the *ideal* wave field and the *aberration* wave field on the image plane. Note that the interference of the two wave fields can be constructive, destructive or an intermediary. However,

$$\cos(\Delta\theta) \leq 1 \quad (14)$$

allows us to bound the image intensity as

$$\tilde{I} = |\mathcal{P}(H(\Phi)) + \mathcal{R}(\Phi)|^2 \leq (|\mathcal{P}(H(\Phi))| + |\mathcal{R}(\Phi)|)^2, \quad (15)$$

Assuming a given phase retrieval method for the ideal forward propagation model is accurate within  $\epsilon$ -error ( $\epsilon > 0$ ) for a local neighborhood  $\Phi^*$  around  $\Phi$  which fulfills  $\|\mathcal{P}(H(\Phi^*)) - I\|^2 \leq \epsilon$ ,

the bound on image intensity from Eq. 15 further allows us to bound the objective in Eq. (12) as

$$\|\tilde{\mathcal{P}}(H(\Phi^*)) - I\|^2 \leq \|( |\mathcal{P}(H(\Phi^*))| + |\mathcal{R}(\Phi^*)| )^2 - I\|^2 + \epsilon. \quad (16)$$

With this bounded objective function in hand, we can now solve for an aberration compensating hologram phase  $\Phi^*$  with an *ideal* propagation image wave field and a wave field error, to produce an intended target image on the real prototype, that is

$$\Phi_{\text{opt}} = \underset{\Phi^*}{\text{minimize}} \quad \underbrace{\|( |\mathcal{P}(H(\Phi^*))| + |\mathcal{R}(\Phi^*)| )^2 - I\|^2}_{\text{upper bound error}}. \quad (17)$$

To this end, we learn the aberration errors using a generative adversarial network (GAN), which acts as an *Aberration Approximator*  $\mathcal{D}$  of the real display. We refer to this network as the *Aberration Approximator* in the remainder of the manuscript. While the wave field itself is difficult to measure, modeling the upper bound error allows us to use the intensity instead, which can be measured by a conventional camera and is used as an the input to the network. The new *aberrated* intensity on the image plane of a real hardware prototype is given by the updated propagated field intensity

$$|\tilde{\mathcal{P}}(H(\Phi))|^2 \leq (|\mathcal{P}(H(\Phi))| + |\mathcal{R}(\Phi)|)^2 \leq \mathcal{D}(|\mathcal{P}(H(\Phi))|)^2. \quad (18)$$

We update the *Aberration Approximator* iteratively by learning the aberration wave error and optimize for a compensating hologram phase pattern. In the subsequent sections, we describe our method of constructing this approximator for the real hardware display prototype and our hologram phase optimization strategy. We note that vision aberrations including astigmatism compensation can be accurately modeled and implemented using an additional Zernike-phase [Maimone et al. 2017]. This additional phase can be directly added to our holograms to compensate for vision-induced aberrations.

#### 4.3 Learned Aberration Approximator

As described in Section 3, it is challenging to formally model all deviations from the ideal forward model in a real holographic display. Furthermore, the individual steps in the physical display chain that contribute to aberrations are often non-differentiable (e.g. non-linear SLM response) which prevents using efficient first-order gradient solvers. We tackle this problem by modeling the upper bound Eq. (18) as a *differentiable* function  $\mathcal{D}$  that acts as an *approximation* to the real hardware, and with it we can solve the upper bound Eq. (17) optimization problem. Formally, we model real-world deviations using

$$\mathcal{D} : \mathbb{I} \mapsto \tilde{\mathbb{I}}, \quad (19)$$

where  $\mathbb{I}$  denotes the set of images computed using ideal phase hologram propagation and  $\tilde{\mathbb{I}}$  is the set of real world captures.

Deep neural networks are powerful models for non-linear and non-trivial image-to-image mappings and hence are a natural choice for our differentiable hardware approximator. Existing deep learning approaches have been proposed for transferring styles (e.g. Monet to Van Gogh) [Zhu et al. 2017] or hallucinating realistic images from under-constrained semantic labels (e.g. segmentation maps, sketches) [Isola et al. 2016]. In contrast, our task is to approximate captured holographic aberrations precisely down to minute details.

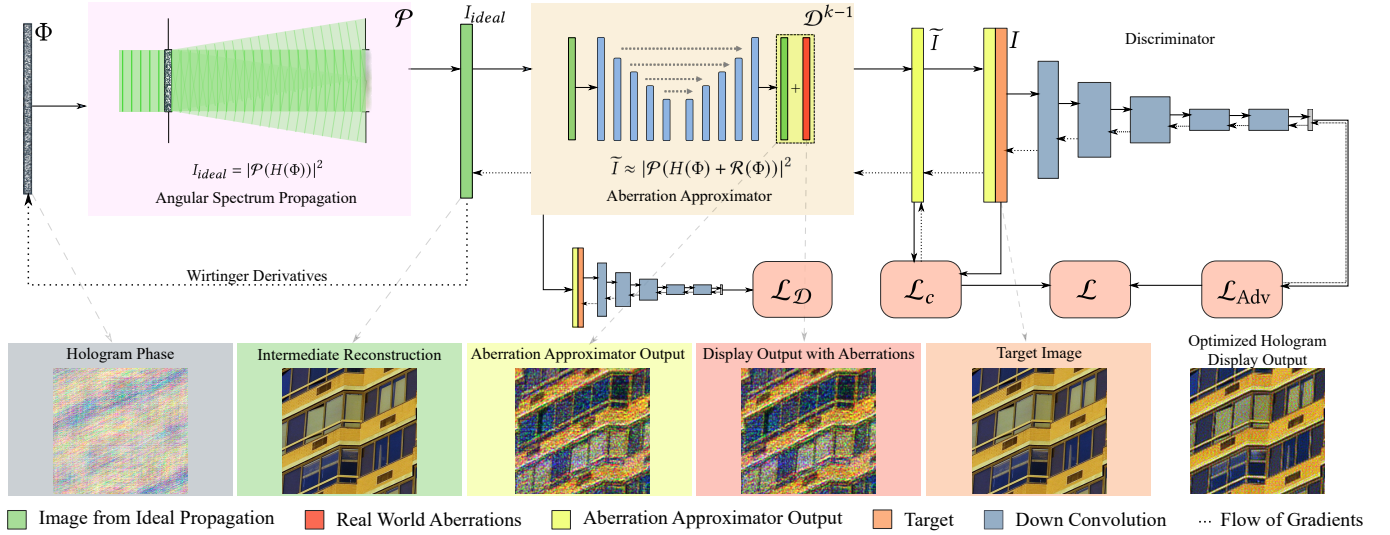


Fig. 3. Overall pipeline of our hardware-in-the-loop phase retrieval method. We compensate for real world errors in holographic image reconstruction by a hardware-in-the-loop optimization method. In contrast to existing approaches, we use a trained neural network that acts as a differentiable mapping function from the *ideal* simulated reconstruction to the *aberrated* real world display. Using this network as an approximator to the hardware, we perform a hardware-in-the-loop optimization to generate *aberration-compensating* holograms that eliminate diverse hardware deviations.

Departing from existing work, our aberration approximator *must not* hallucinate plausible, natural features involved and the mapping to be learned needs to be constrained. As such, we propose a conditional GAN conditioned on the ideal reconstruction instead of a semantic or style guide that can be found in Pix2Pix [Isola et al. 2016]. Specifically, we train a substantially modified U-Net [Ronneberger et al. 2015] to map *ideal* simulated reconstructions to those captured from a real holographic display. We observe that conditioning the discriminator on the ideal simulated reconstruction provides better training stability and performance. Our training data is captured in a *hardware-in-the-loop* fashion as discussed in Section 5.1. We discuss the network architecture and training details in the following section.

#### 4.3.1 Network Architecture.

**Generator.** Our generator network is a modified U-Net with a base structure of 8 downsampling operations using stride 2 convolutions with  $5 \times 5$  kernel window followed by symmetric upsampling using stride 2 transposed convolutions with  $4 \times 4$  kernel window. However, departing from the architecture proposed in Pix2Pix [Isola et al. 2016], we cater the network towards our application through several modifications. First, we remove dropout as we found that it provides excessive regularization. Second, the last layer of the Pix2Pix U-Net consists of directly upsampling an  $H/2 \times W/2 \times 64$  feature map to the final  $H \times W \times 3$  RGB image, while we instead choose to first upsample to a high-resolution  $H \times W \times 32$  feature map before using an additional stride 1 convolution to produce the final  $H \times W \times 3$  RGB image. Third, we remove instance normalization from the first two encoding layers and the last two decoding layers to allow for better tone matching. Lastly, our skip connections connect the encoding layers after LeakyReLU activation instead of before as is done in Pix2Pix. For more details, please see the Supplementary Material.

**Discriminator.** Using a discriminator during the training phase helps in constructing a robust loss function for improving the perceptual quality of predictions from our aberration approximator network. We use a  $94 \times 94$  PatchGAN discriminator [Isola et al. 2016], which consists of three downsampling stride 2 convolutions followed by three stride 1 convolutions. The last layer uses a Sigmoid activation to convert the output of the discriminator into a probability score between 0 and 1 predicting whether the given image is a real holographic capture or a generated image. Although conditional GANs typically provide auxiliary information such as segmentation maps or edge sketches for the discriminator to be “conditioned” on, we do not utilize these in our application. Nevertheless, we found that conditioning the discriminator on the ideal simulated reconstruction improved our learned aberration approximator. In this setting, our discriminator learns to recognize realistic transformations from simulations to the actual hardware capture. Please refer to the Supplementary Material for additional details.

**Loss Function.** We use a weighted combination of  $\ell_1$  loss  $\mathcal{L}_{\ell_1}$ , perceptual loss  $\mathcal{L}_{Perc}$  [Johnson et al. 2016], and adversarial loss  $\mathcal{L}_{Adv}$  to train the *Aberration Approximator*:

$$\mathcal{L}_D = \lambda_{\ell_1} \mathcal{L}_{\ell_1} + \lambda_{Perc} \mathcal{L}_{Perc} + \lambda_{Adv} \mathcal{L}_{Adv}. \quad (20)$$

The perceptual loss compares the image features from activation layers in a pre-trained VGG-19 neural network, that is,

$$\mathcal{L}_{Perc}(x, y) = \sum_l v_l \|\phi_l(x) - \phi_l(y)\|_1, \quad (21)$$

where  $\phi_l$  is the output of the  $l$ -th layer of the VGG-19 pre-trained network and  $v_l$  is the corresponding loss balancing weight. Specifically, we use the outputs of ReLU activations just before the first two maxpool layers, i.e. relu1\_2 and relu2\_2. The combined  $\ell_1$  loss and

perceptual loss acts as a *content loss* for learning most of the aberrations in the holographic display, while the adversarial loss provides additional direction for learning any remaining errors that were missed by the content loss. In experiments, we set  $v_{\text{relu}_2} = 1.5$ ,  $v_{\text{relu}_2} = 1.0$ ,  $\lambda_1 = 10.0$ ,  $\lambda_{\text{Perc}} = 0.05$ , and  $\lambda_{\text{Adv}} = 0.5$  which produces a 1:1 ratio of content loss to adversarial loss.

#### 4.4 Learned Hardware-in-the-loop Hologram Optimization

In lieu of differentiable real hardware, we use a trained neural network that estimates the experimental wave errors to compensate for any aberrations, directly in the computed phase holograms. To this end, we train our *Aberration Approximator* network in a two-step alternating optimization approach, listed in Algorithm 1. Specifically, we alternate between optimizing for phase patterns (with fixed forward model) and optimizing for local aberration errors by refining the network with real holographic display aberrations (from fixed phase patterns). We describe both steps in the following.

*Training the aberration approximator.* We initialize the proposed training scheme assuming an ideal wave field at the image plane, i.e.  $\mathcal{R}(\Phi) = 0$  and optimize for the underlying phase holograms for the input dataset of target images  $\mathbb{T}$ . When these holograms are displayed on a real prototype, severe artifacts are present which we model with the learned aberration approximator  $\mathcal{D}$ . To this end, we acquire repeatedly captured datasets of real aberrated images  $\tilde{\mathbb{I}}$  from a prototype holographic display and learn  $\mathcal{D}$  by training a GAN as described in Section 4.3.1. With the new aberration approximator  $\mathcal{D}$ , we then recompute the phase holograms, updating the propagation model to contain the aberrations modeled by the network, as described by Algorithm 1. Each iteration alternates between the phase computation, capture step, and aberration approximator network refinement steps using the most recent captures, thus learning to model finer residual errors than the previous iterations. Once learned, we use the frozen aberration approximator network model for all future computations of aberration compensated phase holograms.

*Hardware-in-the-loop phase retrieval.* For the phase retrieval stage, we use the trained frozen network  $\mathcal{D}$  to solve the aberration compensating phase optimization problem as described in Eq. (18). Our objective function is a combination of several penalty functions: 1) the *learned perceptual image similarity metric* (LPIPS) [Zhang et al. 2018] as a deep neural network perceptual quality metric based on human judgments, 2) multi-scale structural similarity index (MS-SSIM) [Wang et al. 2003] as a perceptual quality metric, 3)  $\mathcal{L}_2$  loss for pixel-wise accuracy. We note that our aberration approximator network learns to predict the real holographic display images which typically are prone to noise, spatial and radiometric inaccuracies compared to the *digital target image*. While the LPIPS loss is trained to be robust to such inaccuracies, training an additional adversarial loss, as shown in Figure 3, allows for a more robust loss function that better generalizes to real-world measured holographic images. Our overall loss function  $\mathcal{L}$  is a weighted combination of the above mentioned content and adversarial loss functions given by

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{LPIPS}} + \lambda_{\text{ms-ssim}} \mathcal{L}_{\text{ms-ssim}} + \lambda_{\ell_2} \mathcal{L}_{\ell_2}}_{\text{content loss } \mathcal{L}_c} + \lambda_{\text{Adv}} \mathcal{L}_{\text{Adv}} \quad (22)$$

---

#### Algorithm 1 Training the aberration approximator network $\mathcal{D}$ .

---

```

1: Inputs:
    $\mathbb{T}$  // Training image set
2: Outputs:
    $\mathcal{D}$ 
3: // Refine aberration approximator  $\mathcal{D}$  for  $K$  iterations
4:  $\mathcal{D}^0 = \text{Id}$ 
5: for  $k = 1$  to  $K$  do
6:   // Iterate over training dataset
7:   for  $I \in \mathbb{T}$  do
8:     // Optimize for each  $\Phi_I$  with a fixed  $\mathcal{D}$ 
9:      $\Phi_I^{k-1} \leftarrow \underset{\Phi}{\text{minimize}} \underbrace{\mathcal{L}(\mathcal{D}^{k-1}(|\mathcal{P}(H(\Phi))|^2) - I)}_{\text{upper bound error}}$ 
10:  end for
11:  // Compute ideal phase hologram images
12:   $\mathbb{I} \leftarrow \mathcal{P}(H(\Phi^{k-1}))$ 
13:  // Capture real hardware aberrated images
14:   $\tilde{\mathbb{I}} \leftarrow \tilde{\mathcal{P}}(H(\Phi^{k-1}))$ 
15:  // Refine  $\mathcal{D}$  with new captures
16:   $\mathcal{D}^k \leftarrow \text{refine}(\mathcal{D}^{k-1}, \mathbb{I}, \tilde{\mathbb{I}})$ 
17: end for
18: // Freeze  $\mathcal{D}$  for future use
19:  $\mathcal{D} \leftarrow \mathcal{D}^K$ 

```

---

Although the optimization loss function is similar to the loss function used to train our *aberration approximator* network, note that training of the *aberration approximator* is done over an entire dataset of images, whereas the phase-only hologram optimization is done for a single image, in the future, that is *unseen* during network training. We perform the phase retrieval by solving the above equation using the Wirtinger derivatives as described by Chakravarthula et al. [2019]. We repeat the phase optimization process until the aberration error outputs fall below a user-defined threshold. We derive the corresponding Wirtinger derivatives and offer insights into termination criteria in the Supplementary Material.

Please note that the latest implementations of machine learning libraries, such as TensorFlow, have *Wirtinger derivatives* built into their auto-differentiation packages, easing the use of traditional stochastic gradient descent (SGD) methods. While TensorFlow's implementation of gradients is similar to ours, PyTorch computes the complex derivatives via a Jacobian, which implicitly computes the Wirtinger derivatives. We discuss both implementations in detail in the Supplementary Material.

#### 4.5 Extension to Online-Camera Phase Optimization

In the approach discussed above, we use a learned *aberration approximator* which acts as a substitute for real display and camera hardware, to compute hologram phase patterns that compensate for real hardware aberrations. This approach uses the actual display and camera hardware only for acquiring training data for the aberration approximator network (and learning the unknown light transport).

Holograms can also be actively optimized in an online fashion, using images from a camera that sees the holographic projections at



each iteration as reference. In other words, holograms can be optimized to compensate for real hardware aberrations directly, without substituting the hardware with the trained aberration approximator network. Given that the light transport in a real holographic display is unknown, one can assume an ideal propagation model for the purpose of computing gradients for iterative hologram phase refinement. Alternatively, we can easily extend our network training strategy as described by Algorithm 1 to an online learning method for holographic phase retrieval of unknown propagation models. Here, the test images are now *seen* during training. Specifically, the alternating optimization for phase retrieval (with fixed forward model) and network refinement (with fixed holographic phase) can be used for computing aberration compensating holograms for each individual image. However, we note that such an approach takes several minutes for optimizing a single phase hologram as a result of the  $K$  sequential iterations, including display, capture, fine-tuning and training, making it impractical. We envision future training hardware to be capable of online phase refinement at fast rates. Although efficient online phase optimization is an exciting area for future research, we find that our learned hardware-in-the-loop phase retrieval method already compensates for severe real-world aberrations. We show holographic images obtained from online compensation in the Supplementary Material.

## 5 SETUP AND IMPLEMENTATION

We assess the holograms generated by the proposed method in simulation and experimentally on a hardware display prototype. We discuss the specific hardware setup and software implementation details in the following.

### 5.1 Hardware Prototype

Our prototype holographic display uses a HOLOEYE LETO-I liquid crystal on silicon (LCoS) reflective phase-only spatial light modulator with a resolution of  $1920 \times 1080$  and a pixel pitch of 6.4 nm. The SLM is controlled as an external monitor and the hologram phase patterns are transferred and displayed on it via HDMI port of a graphics card. This SLM is illuminated by a collimated and linearly polarized beam from a single optical fiber that is coupled to three laser diodes. The laser diodes emit at wavelengths 446 nm, 517 nm and 636 nm and are controlled using a ThorLabs LDC205C laser diode controller in a color field sequential manner.

The illuminated beam that is modulated by the phase-only SLM is focused on an intermediate plane where an iris is placed to discard unwanted diffraction orders and conjugate images. Specifically, we use a linear phase ramp on the computed phase hologram to physically shift the holographic image away from the zero-order undiffracted light. We then block this zero-order undiffracted light and the conjugate (ghost) images using an iris, allowing primarily the modulated light to form an image. We relay this light onto the camera sensor plane. We use a Canon Rebel T6i DSLR camera body, without the camera lens attached, to capture images for the assessment of the display's image quality. The camera has an output resolution of  $6000 \times 4000$  and a pixel pitch of  $3.72 \mu\text{m}$ , well above the pitch of our SLM. This oversampling on the image plane allows us to effectively capture high frequency content, including noise.

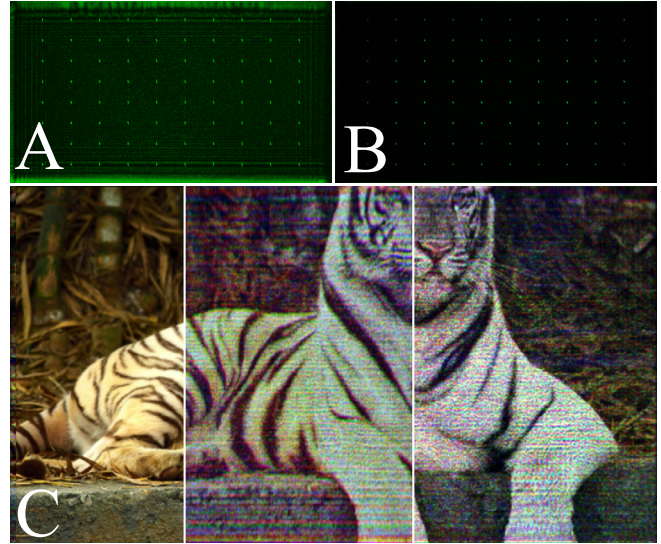


Fig. 4. Zero-order elimination. (A) Zero-order undiffracted light from the dead-space between SLM pixels result in a bright ringing pattern over the image that severely affects the image quality with poor contrast. (B) We eliminate the zero-order light by adding a linear phase ramp to the phase patterns and filtering the unwanted light using an iris. (C) Compared to the reference image (left), holographic projections with zero-order light show severe ringing artifacts (middle). Eliminating zero-order light improves the image contrast significantly but results in additional artifacts due to SLM modulation limitations (right). The holograms for (C) are computed using Wirtinger Holography [Chakravarthula et al. 2019], with a linear phase ramp added to it.

Figure 4 validates that we have almost completely eliminated the zero-order undiffracted light. However, the additional phase ramp added to the holograms to filter out the undiffracted and conjugate light results in severe artifacts manifesting as horizontal streaks as can be seen in Figure 4-C. Residual artifacts resulting from the several real-world deviations such as laser speckle, dust and scratches in the optical system, recall Section 3, further degrades the image quality noticeably. Our method successfully eliminates such severe artifacts as discussed in Section 6.

**Data Acquisition.** Learning the aberrations of the proposed hardware prototype requires capturing a dataset of aberrated real-world display outputs and the corresponding sharp reference images. Acquiring such data manually is often tedious and prone to errors. Specifically, misalignment between various captured images results in additional inconsistencies in the training data which are not due to the non-ideal behavior of the hardware.

To address these challenges, we use the display-camera setup shown in Figure 5 to acquire large training datasets robustly without human intervention. This is achieved by sequentially displaying phase holograms on the SLM and simultaneously capturing the corresponding holographic projections with the camera. We repeat this process for each refinement step of our *aberration approximator*, as described in Section 4.4.

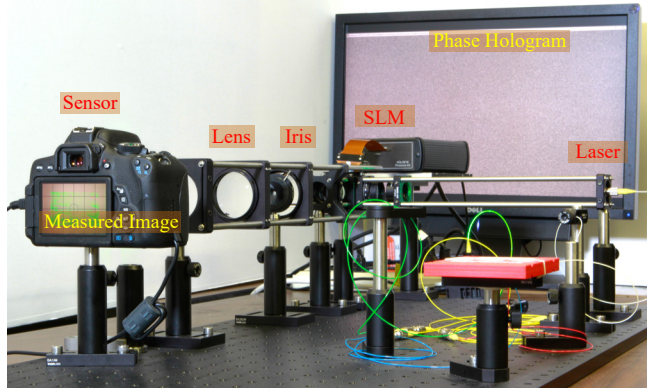


Fig. 5. *Hardware-in-the-loop display-capture setup*. Top: The display-prototype setup to generate the training data for our aberration approximator network. Bottom: Setup schematic, with RGB lasers that are coupled into a single-mode fiber which illuminates the reflective SLM displaying the phase pattern. The modulated wave is measured using a conventional intensity camera.

## 5.2 Implementation

We implement the *aberration approximator* and optimization procedure for the phase-only holograms using TensorFlow 1.14. We train the model on a GPU cluster providing 16 GB of memory required for handling the trained network as well as the high resolution  $1080 \times 1920$  images. We do not use data augmentation techniques and instead simply train on full resolution images. For divisibility, we zero-pad the input images to  $1280 \times 2048$  and then we crop the  $1080 \times 1920$  region-of-interest from the output of the network. We use the Adam optimizer for both the generator and discriminator with learning rate 0.0002 and an exponential decay rate of  $\beta_1 = 0.5$  for the first moment and  $\beta_2 = 0.999$  for the second moment, and we train for 40000 iterations. Our batch size is one. The overall training process takes around one day for a dataset of 120 images. The source images are from our custom dataset of images randomly picked from the DIV2K dataset [Agustsson and Timofte 2017] which contains high resolution images of natural scenes.

The proposed *hardware-in-the-loop* optimization is implemented using TensorFlow eager execution mode to allow for Wirtinger gradient updates. We obtain the gradients of the aberration approximator and the losses from auto-differentiation using TensorFlow and use those to compute Wirtinger gradients. We optimize the phase holograms using Adam optimizer with a learning rate of 0.001 and exponential decay rates of  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$  for

Table 1. Quantitative comparison against state-of-the-art deep learning techniques for image mapping. We outperform existing encoder-decoder and generative adversarial approaches in predicting the aberrations on the holographic display output.

	PSNR (dB)	SSIM	1 - LPIPS
Proposed	<b>29.6</b>	<b>0.831</b>	<b>0.943</b>
Proposed (unconditional)	24.5	0.628	0.825
Pix2PixHD [2017]	24.1	0.565	0.813
Pix2Pix [2016]	24.3	0.614	0.824
U-Net [2015]	24.7	0.595	0.786
U-Net [2015] with $\mathcal{L}_2$	24.4	0.563	0.425
U-Net [2015] with $\mathcal{L}_1$	25.1	0.590	0.497

the first and second moments, respectively. The optimization of the phase holograms for a given aberration approximator takes about 80 sec for 300 iterations on a consumer GPU.

## 6 ANALYSIS

To compute holograms that compensate for experimental aberrations, the aberration approximator network has to reliably replicate the hardware artifacts. In this section, we analyze the modeling capability of the learned aberration approximator. We first compare the performance of our network architecture against existing image-to-image translation approaches. Next, we perform an ablation study to demonstrate that our architecture modifications and loss function produce meaningful improvements in modeling real aberrations. Quantitative metrics for network evaluations are calculated between the ground truth “noisy” images from the holographic display and the outputs of the approximator networks. Lastly, we demonstrate in simulation that the aberration approximator provides useful gradients for phase retrieval optimization.

### 6.1 Baseline Comparisons

We compare our proposed aberration approximator architecture and loss function against state-of-the-art image-to-image translation approaches. Specifically, we compare against the vanilla U-Net [Ronneberger et al. 2015], Pix2Pix [Isola et al. 2017], Pix2PixHD [Wang et al. 2017]. We train the U-Nets using  $\ell_1$ ,  $\ell_2$ , and perceptual loss. For Pix2Pix and Pix2PixHD we use their default settings and loss functions. All methods were trained on full resolution images without augmentation. Due to memory constraints, we halve the number of filters in the U-Net at each layer and for Pix2PixHD we halve the number of filters in the first layer of the generator and discriminator.

Table 1 shows PSNR, SSIM, and LPIPS evaluation results. The proposed method outperforms the state-of-the-art networks by at least 5 dB in PSNR. The SSIM results also indicate that our approach produces more perceptually accurate reconstructions than other baseline methods. Figure 6 shows qualitative examples for a few selected images. The vanilla U-Nets with  $\ell_1$  and  $\ell_2$  losses fail to learn the diverse noise patterns. Training a U-Net with a perceptual loss allows for better replication of noise but the predictions are still far from the actual aberrations and severe checkerboarding artefacts can be observed. Pix2Pix is capable of learning the aberrations to an extent, however, the color tones and patterns do not match as closely as ours. Pix2PixHD learns to generate high frequency aberrations that look plausible, but these aberrations are often misaligned with



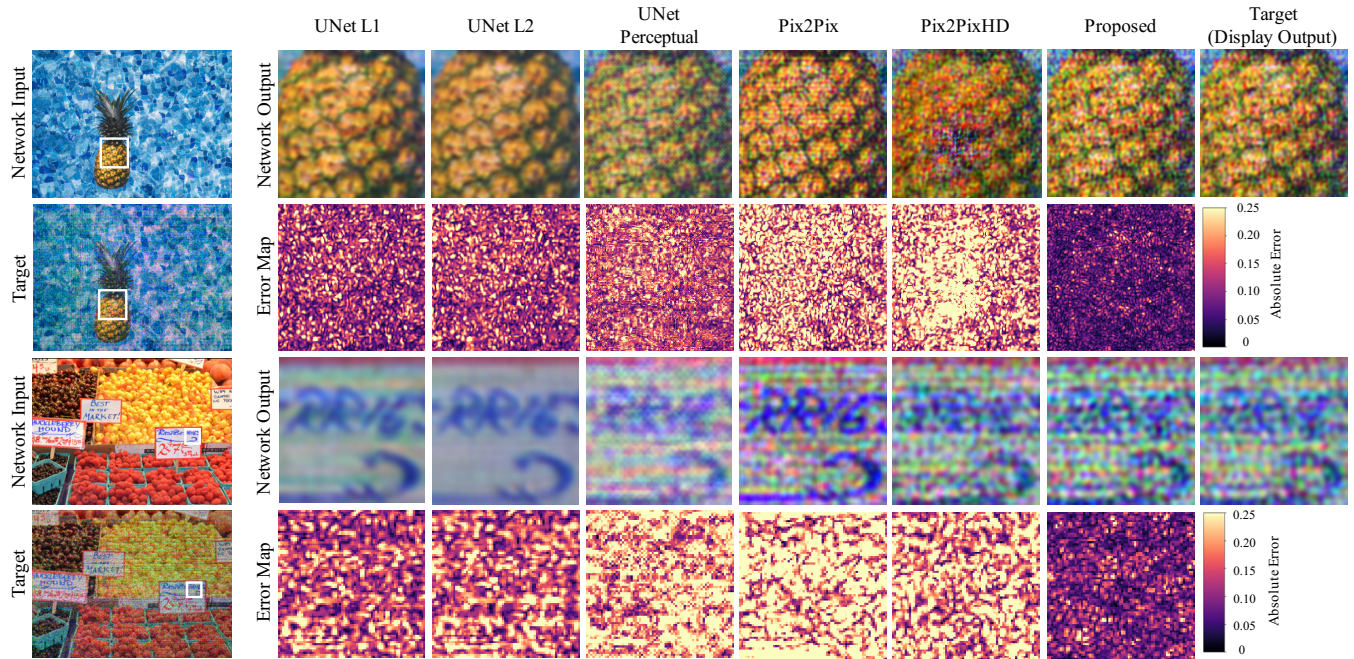


Fig. 6. Qualitative comparison of different deep learning techniques for holographic aberration prediction. We compare aberration predictions using our method against state-of-the-art image mapping techniques. In addition to example outputs from each method, we show absolute error maps between each method and the target display output. All U-Net approaches fail to match the aberrations of the display. Pix2Pix is able to coarsely match the aberrations but not the color tones. Pix2PixHD learns to generate aberrations that look believable but do not match that actual display captures, which is particularly evident in the corresponding error maps. Furthermore, Pix2PixHD sometimes produces additional artefacts; an example of which can be seen in the first row. The proposed approach models the aberrations in a real holographic display with greatest fidelity.

the actual display aberrations as can be seen in the error maps. Furthermore, Pix2PixHD occasionally fails and introduces additional artefacts. We found that deeper networks such as the U-Net and Pix2PixHD were more difficult to train for our task compared to shallower networks such as ours and Pix2Pix. We account this to the fact that the small loss gradients from the high frequency aberrations are difficult to backpropagate through these deeper networks.

## 6.2 Ablation Study

We performed an ablation study to demonstrate how our architecture design and loss function choices impact the aberration prediction quality. We found that training with an unconditional GAN was less stable and prone to falling into undesirable local minima, see Table 1 and the Supplementary Material. Intuitively, conditioning the discriminator on the ideal reconstructions allowed the discriminator to focus on distinguishing the aberrations only, which provides better training signal for our generator. We refer to the Supplementary Material for additional comparisons on the loss function components.

## 6.3 Simulated Phase Retrieval with Aberration Approximator

We demonstrate in simulation that using our aberration approximator provides useful gradients which allow for phase retrieval. Figure 7 shows simulated phase retrieval results with the proposed

Table 2. Quantitative results for phase retrieval with the aberration approximator in simulation, starting from hardware captures obtained using Wirtinger holography [2019] with DC component removed. The quantitative improvement demonstrates that our aberration approximator provides useful gradients that allow for optimization within the Wirtinger framework.

	PSNR (dB)	SSIM	1 - LPIPS
Proposed	<b>30.4</b>	<b>0.937</b>	<b>0.949</b>
Wirtinger Holography [2019]	19.9	0.571	0.505

aberration approximator included as described in Section 4.4. Only instead of displaying the optimized phase pattern on the physical display, we propagate the wavefront through the ideal forward model and the aberration estimator, effectively simulating the imperfect hardware system. These synthetic results validate the effectiveness of our proposed method (Algorithm 1). Table 2 shows quantitative improvement after our optimization process in simulation. The significant improvement across all image quality metrics demonstrates that the proposed technique indeed compensates for the adequately modeled hardware deviations. We note, since our aberration approximator is trained to map ideal Wirtinger reconstructions to hardware captures, it is unsuitable for evaluating other methods such as double phase encoding [2017] in the same simulation framework. As such, we defer to the next section for comparisons on the hardware prototype.



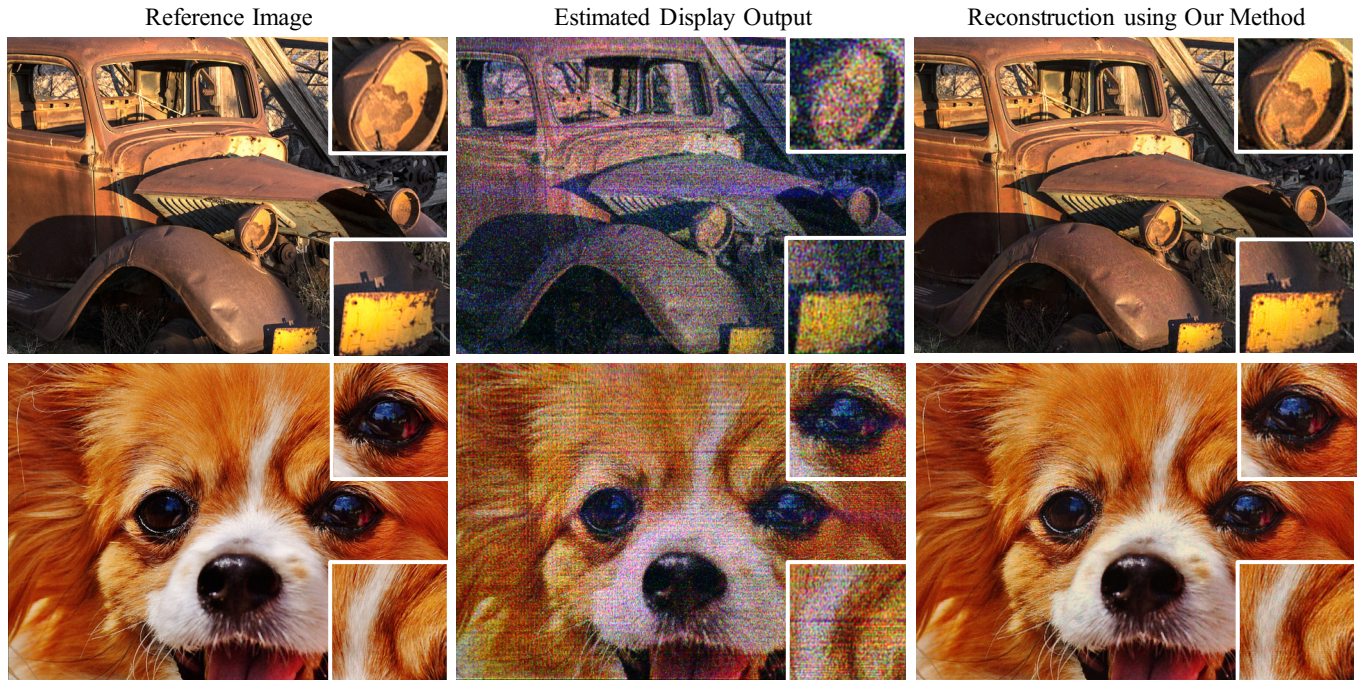


Fig. 7. *Simulated results for phase retrieval using the trained aberration approximator.* For a given target image (left), our aberration approximator network estimates the “aberrated” hardware display output (middle) due to real world deviations from an ideal light propagation model. We compensate for these real world deviations through our hologram optimization method. The simulated reconstructions of our holograms eliminate severe predicted aberrations and show a significant improvement in image quality (right).

Table 3. Quantitative results for holographic reconstructions using different phase retrieval methods. Metrics are computed on *hardware captures* compared against the aligned target image. For fair comparison we adjust intensities to match the target image intensity [Yoshikawa et al. 2016]. The proposed method not only outperforms other methods by 2.5 dB PSNR, but also greatly improves the SSIM and LPIPS perceptual metrics, and thus quantitatively demonstrates improved perceptual quality.

Real display output	PSNR (dB)	SSIM	1-LPIPS
Proposed	<b>20.5</b>	<b>0.625</b>	<b>0.541</b>
Wirtinger Holography [2019]	17.6	0.475	0.417
Double Phase Encoding [2017]	15.24	0.342	0.208

## 7 ASSESSMENT

We validate and evaluate our approach for real-world use by comparing it to the state-of-the-art holographic phase retrieval methods, Double phase encoding [Maimone et al. 2017] and Wirtinger Holography [Chakravarthula et al. 2019], in full color. Figure 8 shows experimentally acquired results from the holographic display setup described in Section 5.1. The proposed method outperforms existing state-of-the-art methods quantitatively and qualitatively for real-world display captures.

For the purpose of acquiring these results, we covered our hardware prototype with blackout curtains to mitigate the effect of ambient light on the final captures. We maintained the SLM look-up-tables (LUTs), camera settings and laser settings constant throughout the various experiments. In particular, we computed holograms of a resolution test pattern showing TV lines of varying frequencies

and finetuned the camera sensor plane position until all the frequencies in the TV chart can be seen, to minimize any loss of frequencies due to camera defocus errors. We kept the camera ISO at 100 and exposure at 10 ms. We used the SLM look up tables (LUTs) provided by the manufacturer to display the *calibration TV-hologram* but tuned the voltages by a small amount until a maximum diffraction efficiency is noticed. The LUTs and the corresponding voltages are then kept constant for all the experiments.

We found that the holograms for other methods needed an extra gamma correction to the phase patterns, before displaying on the hardware, for improved performance on our aged SLM with inaccurate look-up-tables, aiming for best hand-tuned display outputs. In contrast, we *do not* apply such gamma correction to the phase patterns from our hardware-in-the-loop method and the proposed approach directly compensates for the phase errors. Also, since the double phase encoding leads to the unwanted effect of a noticeable portion of light escaping the designated image window, we increase the laser power for these holograms to match the intensities for a fair comparison. We note that our optical setup is optimized to relay a real projected image directly to the sensor of the DSLR camera (Section 5.1) without lens distortions, eliminating the need for additional calibration.

The real captured results reported in Figure 8 show that the proposed approach compensates for most of the severe aberrations occurring in existing methods. We eliminate the zero-order undiffracted light that is caused by the dead pixel space in the SLM as





Fig. 8. *Experimental Validation.* We computed aberration compensated phase holograms using the proposed hardware-in-the-loop approach. We captured RGB color images of these aberration compensated holograms in a color sequential manner. The same camera settings were used for all the image capture experiments while the output laser power was adjusted to white-balance the illumination. Note that the laser power was adjusted (increased) for double phase encoding holograms due to the unwanted loss of light in the double phase encoding approach.



discussed in Section 5.1. This significantly improves the contrast and can be seen in the results reported, e.g. the black regions between the flowers in second row and the doll in the last row of results. The proposed approach eliminates the ringing at the edges of the holographic projections and reconstruction noise that is present in existing methods. This can be observed in the image patches that are selected from the peripheries of the real captures. As a result, fine details such as the skin of the star fish in the first row and texture on the porcelain in the last row are revealed by the proposed method in contrast to Wirtinger holography or double phase encoding methods. Similarly, the fine details on the flowers and the colored glass window in the middle rows is made visible with the proposed method at high contrast and resolution.

We further validate our method by computing several image quality metrics such as PSNR, SSIM and LPIPS perceptual similarity on real holographic images from a custom test dataset consisting of images randomly chosen from the DIV2K dataset [Agustsson and Timofte 2017] and that are not seen by the aberration approximator before during the training stage. The resulting quality metrics for the real captured holographic images for our method compared against the state-of-the-art double phase encoding and Wirtinger holography are reported in Table 3. The target display images were first gamma-corrected and the holograms of these gamma-corrected intensity images were computed. These holograms were displayed on a real holographic display and captured with a linear camera sensor to compute the corresponding image quality metrics between the captured images and the gamma-corrected targets. The proposed method demonstrates 2.5 dB improvement over the prior methods and significant increases in SSIM and LPIPS performance – validating the improved quality of the real holographic images shown in Figure 8. For further evaluation, we provide additional real captured results in Supplementary Material.

### 7.1 3D Holographic Display

We also extend the proposed hardware-in-the-loop phase retrieval method to tackle 3D volumetric scenes via dynamic global scene focus. Computing full Fresnel holograms of 3D scenes using point-source integration methods generally result in holographic imagery with continuous per-pixel focus cues. However, these methods require computing and superposing the underlying lens phase patterns individually for several million points that make the 3D scene, which is a computationally expensive process. Moreover, such Fresnel holograms are typically not *phase-only* and require either a complex modulation of both amplitude and phase [Shi et al. 2017], or a heuristic encoding of the complex hologram [Maimone et al. 2017]. Each of these approaches results in reduction of the holographic projection quality. As an alternative, the 3D volume can be discretized into multiple focal planes and a superposition of holograms corresponding to only those depth planes can be computed to render a 3D hologram. The holograms for each focal plane can be independently or jointly computed using our optimization framework, however, at the cost of a reduction in speed by  $N$  for  $N$  discrete depth planes.

Therefore, we provide 3D holograms by sweeping holographic projections of 2D scenes through the 3D volume in space, via dynamic global scene focus. Specifically, since the human eye can focus only at a single depth at any given instant of time, we display

a 2D hologram of the 3D scene whose focus is changed to match the focal depth of the user’s eye [Maimone et al. 2017]. The depth of field blur in the scenes is rendered in image space [Chakravarthula et al. 2018; Cholewiak et al. 2017]. While correcting for the aberrations at various depths as described by Maimone et al. [2017] is an option, that requires training the aberration approximator with holographic images generated and captured at various discretely sampled depths. Instead, we train our aberration approximator to predict deviations at a single depth but add additional lens phase functions to the aberration compensated holograms to move the projections across the continuous 3D volume. As shown in Figure 9, the global focus change of the 3D scenes is accurately displayed on our holographic display prototype.

### 7.2 Robustness to Eye Motion and Display Variability

In this section, we evaluate the robustness of our method to spatial, temporal and systemic changes in the display-capture setup, and its inherent hardware limitations.

*Robustness to Eye Motion.* To evaluate the quality of aberration-compensated holographic images with eye motion, we compare the quality of display outputs captured for various *unseen* horizontal and vertical translations of a Canon T6i DSLR camera. Note that the holograms for this experiment are generated using an aberration approximator network that is trained on a dataset of holographic images captured with a fixed camera position. A subset of images shown in the top row of Figure 10 shows that the overall quality of holographic images remain stable with the camera motion up to 4 cm.

*Robustness to Temporal Variability.* To assess the performance of the trained network to temporal variations of the hardware, we compare the captured images from the hardware display prototype over a *time window of three months* with a Canon T6i DSLR camera, as shown in the bottom left of Figure 10. It can be noticed that our aberration approximator predicts real world deviations consistently and there is no significant change in the image quality of aberration-compensated holographic projections over time. The disparities in the brightness and color is due to the differences in the laser power and we note that these are not fundamental to the proposed method.

*Robustness to Acquisition Device.* While the dataset of real hardware captures that we use for training the aberration approximator network inherently contains the nonlinearities of the camera, we tailor our holographic phase optimization loss functions to be perceptually consistent but agnostic to radiometric variations as discussed in Section 4.4. Also note that the optical layout used in our setup is designed to relay the holographic image directly on the camera sensor without any lens distortion or chromatic aberration, eliminating the need for additional lens calibration. To validate this, we capture and compare the display output with two different cameras: a Canon Rebel T6i DSLR camera of resolution  $6000 \times 4000$  and a Fujifilm FinePix S5 Pro DSLR camera of resolution  $4256 \times 2848$ . The Canon T6i has a 24 MP APS-C ( $22.3 \times 14.0$  mm) sized CMOS sensor and features DIGIC 6 processor. On the other hand, Fujifilm S5 Pro has a 6 MP APS-C ( $23 \times 15.5$  mm) sized CCD sensor. Despite

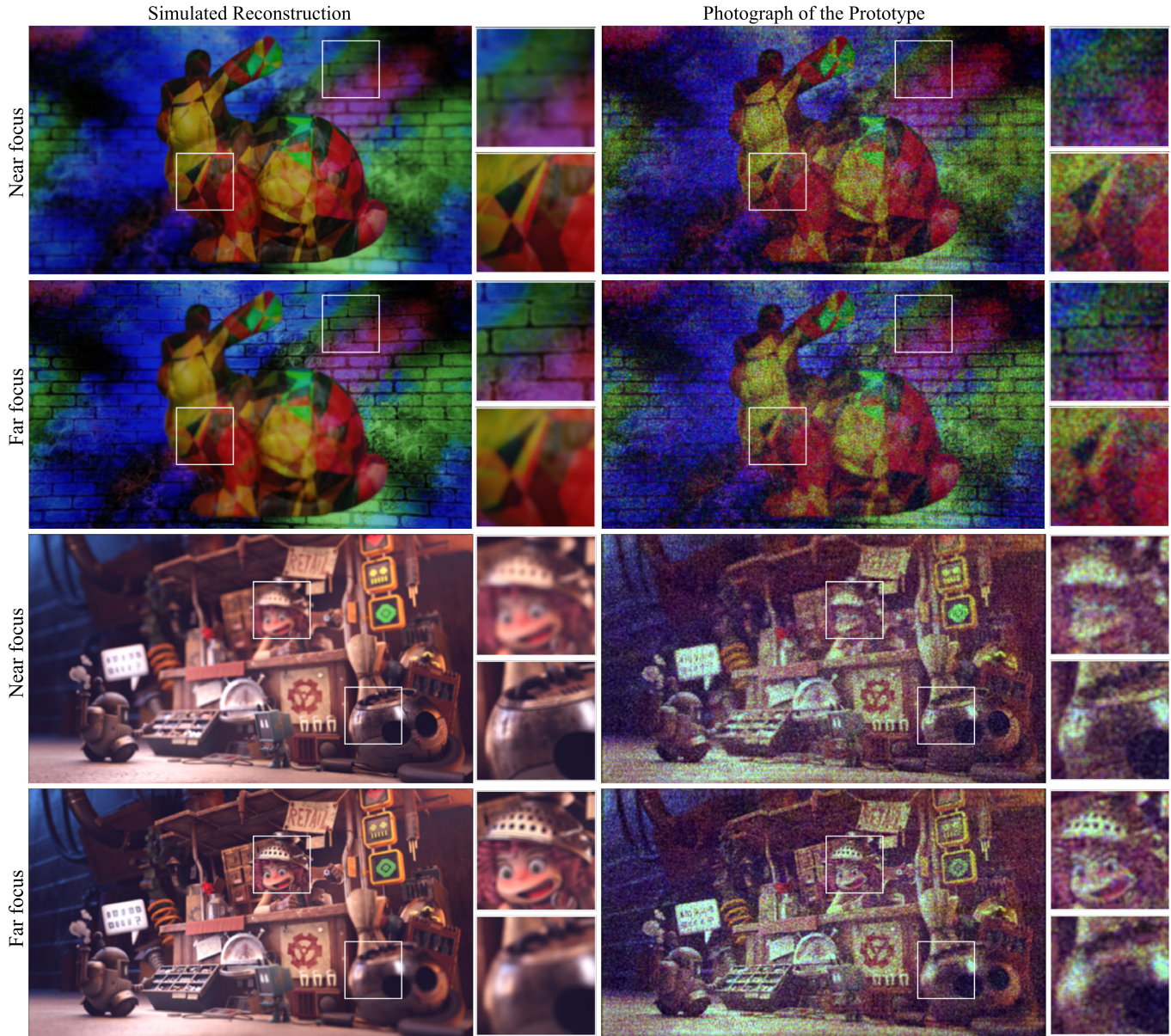


Fig. 9. 3D holograms via dynamic global scene focus. We acquire RGB color images of 3D volumetric scenes projected at different depths via global scene focus and captured in a color sequential manner. The images were captured with the camera settings for all the experiments, an ISO of 100 and an exposure time of 10 ms, while the output laser power was adjusted to white-balance the illumination.

the two devices having entirely different imaging sensor properties, the display captures from either cameras are consistent, as shown in lower right part of the Figure 10, and validate the robustness of our optimization framework against radiometric variability of acquisition devices.

## 8 DISCUSSION AND CONCLUSION

We introduce a *machine learned hardware-in-the-loop* phase retrieval method to estimate the unknown forward model in a real holographic display. The proposed method allows us to compensate

hardware deviations from the ideal forward model which are non-linear and difficult to model. To this end, we learn an *aberration approximator* from hardware captures, that is parameterized by a deep neural network, and effectively emulates the real hardware. Using this aberration approximator function allows us to formulate the holographic phase retrieval problem as an optimization problem that can be solved using first-order optimization methods. The proposed approach iteratively refines the upper bound on the estimated objective function which we have derived in this work.



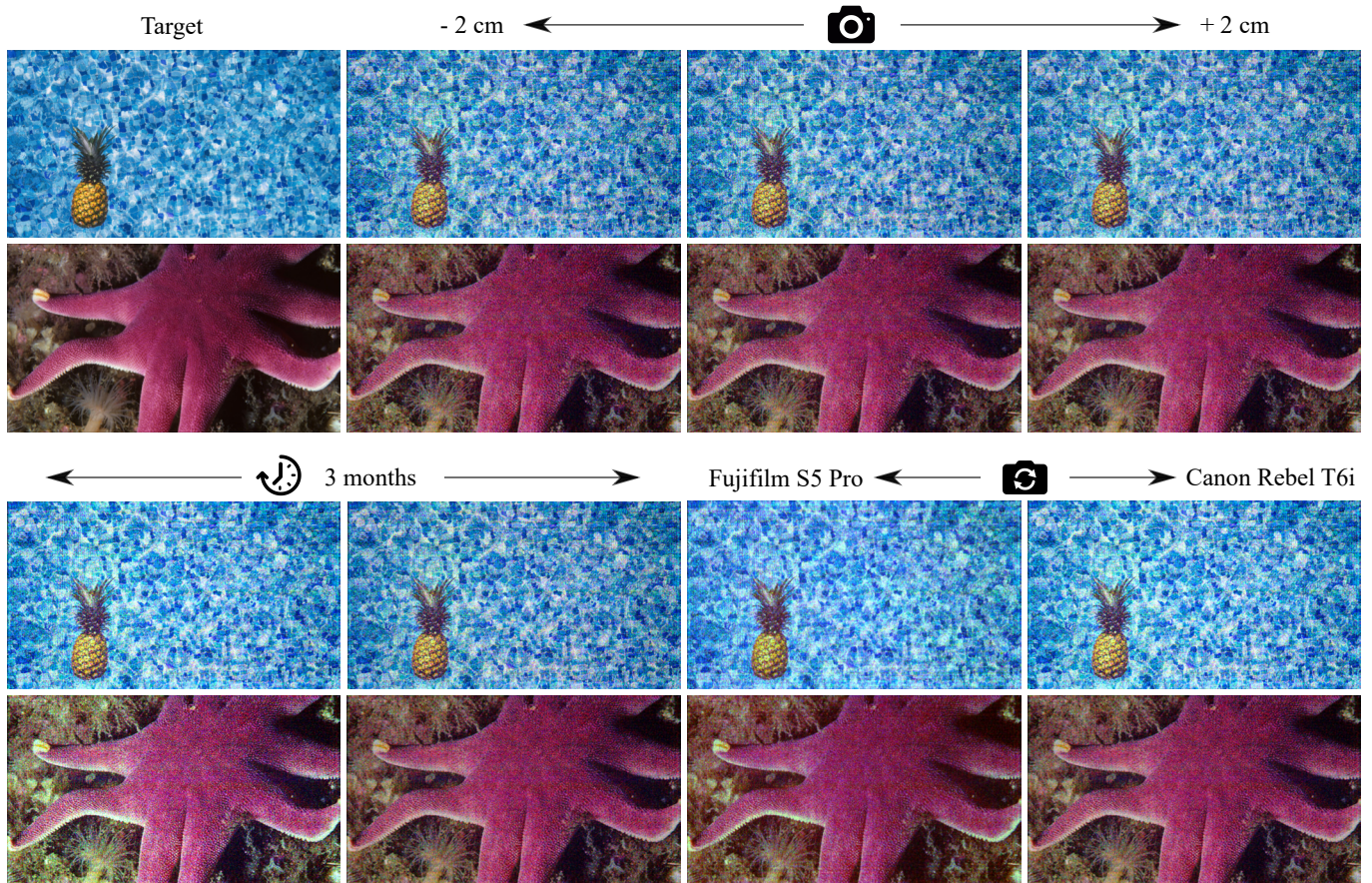


Fig. 10. *Robustness to spatial, temporal and acquisition device.* We evaluate the robustness of our method by acquiring and comparing the quality of images with spatial, temporal and camera variability. In specific, we evaluate the quality of measured holographic images spatially by displacing the camera up to 4 cm (top), temporally over a period of 3 months (bottom left) and with different imaging devices (bottom right). The proposed method produces consistent stable results for all setup variations. Spatial and temporal variability experiments are performed with a Canon Rebel T6i DSLR camera.

We *validate* that our aberration approximator accurately models images acquired from our prototype display. We *assess* the proposed phase retrieval approach by solving for phase-only holograms that compensate for severe errors. In particular, our approach eliminates severe non-linear artifacts in the real holographic reconstructions. Without modeling deviations, the approach allows us to eliminate zero-order undiffracted light, non-ideal and non-linear phase response of the SLM device and severe ringing and chromatic artifacts that are not tackled by existing phase retrieval methods. The proposed method outperforms existing state-of-the-art methods quantitatively and qualitative for real-world display captures.

We envision the proposed *hardware-in-the-loop* phase retrieval method to enable research towards high-quality artifact-free holographic near-eye displays of the future.

## ACKNOWLEDGMENTS

The authors are grateful to Joel S Kollin, Bernard Kress and Brian Guenter for lending the SLMs, Roarke Horstmeyer for lending the laser diode controller and Jim Mahaney for hardware support during

the COVID-19 pandemic. This research is supported by NSF grants 1840131 and 1405847.

## REFERENCES

- Eirikur Agustsson and Radu Timofte. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Lukas Ahrenberg, Philip Benzie, Marcus Magnor, and John Watson. 2008. Computer generated holograms from three dimensional meshes using an analytic light transport model. *Applied optics* 47, 10 (2008), 1567–1574.
- Sohail Bahmani and Justin Romberg. 2017. Phase Retrieval Meets Statistical Learning Theory: A Flexible Convex Relaxation. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Vol. 54. PMLR, 252–260.
- George Barbastathis, Aydogan Ozcan, and Guohai Situ. 2019. On the use of deep learning for computational imaging. *Optica* 6, 8 (2019), 921–943.
- Heinz H Bauschke, Patrick L Combettes, and D Russell Luke. 2003. Hybrid projection–reflection method for phase retrieval. *JOSA A* 20, 6 (2003), 1025–1034.
- Muharrem Bayraktar and Meriç Özcan. 2010. Method to calculate the far field of three-dimensional objects for computer-generated holography. *Applied optics* 49, 24 (2010), 4647–4654.
- Vittorio Bianco, Pasquale Memmolo, Marco Leo, Silvio Montresor, Cosimo Distanto, Melania Paturzo, Pascal Picart, Bahram Javidi, and Pietro Ferraro. 2018. Strategies for reducing speckle noise in digital holography. *Light: Science & Applications* 7, 1 (2018), 1–16.



- Vittorio Bianco, Pasquale Memmolo, Melania Paturzo, Andrea Finizio, Bahram Javidi, and Pietro Ferraro. 2016. Quasi noise-free digital holography. *Light: Science & Applications* 5, 9 (2016), e16142.
- Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. 2013. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics* 66, 8 (2013), 1241–1274.
- Praneeth Chakravarthula, David Dunn, Kaan Aksit, and Henry Fuchs. 2018. Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery. *IEEE transactions on visualization and computer graphics* 24, 11 (2018), 2906–2916.
- Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. 2019. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 213.
- Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Felix Heide, and Henry Fuchs. 2020. Computing high quality phase-only holograms for holographic displays. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)*, Vol. 11310. International Society for Optics and Photonics, 1131006.
- JS Chen and DP Chu. 2015. Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications. *Optics express* 23, 14 (2015), 18143–18155.
- Rick H-Y Chen and Timothy D Wilkinson. 2009. Computer generated hologram from point cloud using graphics processor. *Applied optics* 48, 36 (2009), 6841–6850.
- Steven A Cholewiak, Gordon D Love, Pratul P Srinivasan, Ren Ng, and Martin S Banks. 2017. ChromaBlur: Rendering chromatic eye aberration improves accommodation and realism. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 210.
- J Christopher Dainty. 1977. I The statistics of speckle patterns. In *Progress in optics*. Vol. 14. Elsevier, 1–46.
- WJ Dallas and AW Lohmann. 1972. Phase quantization in holograms—depth effects. *Applied optics* 11, 1 (1972), 192–194.
- James R Fienup. 1982. Phase retrieval algorithms: a comparison. *Applied optics* 21, 15 (1982), 2758–2769.
- James R Fienup. 1993. Phase-retrieval algorithms for a complicated optical system. *Applied optics* 32, 10 (1993), 1737–1746.
- Ralph W Gerchberg. 1972. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik* 35 (1972), 237–246.
- Lior Golan and Shy Shoham. 2009. Speckle elimination using shift-averaging in high-rate holographic projection. *Optics express* 17, 3 (2009), 1330–1339.
- Tom Goldstein and Christoph Studer. 2018. PhaseMax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory* (2018).
- RA Gonsalves. 1976. Phase retrieval from modulus data. *JOSA* 66, 9 (1976), 961–964.
- Joseph W Goodman. 2005. *Introduction to Fourier optics*. Roberts and Company Publishers.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 5967–5976.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Changwon Jang, Kiseung Bang, Gang Li, and Byoungcho Lee. 2019. Holographic near-eye display with expanded eye-box. *ACM Transactions on Graphics (TOG)* 37, 6 (2019), 195.
- Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoungcho Lee. 2017. Retinal 3D: augmented reality near-eye display via pupil-tracked light field projection on retina. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 190.
- Jia Jia, Juan Liu, Guofan Jin, and Yongtian Wang. 2014. Fast and effective occlusion culling for 3D holographic displays by inverse orthographic projection with low angular sampling. *Applied optics* 53, 27 (2014), 6287–6293.
- Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision (ECCV)*.
- Xin Kang. 2008. An effective method for reducing speckle noise in digital holography. *Chinese Optics Letters* 6, 2 (2008), 100–103.
- Hwi Kim, Joonku Hahn, and Byoungcho Lee. 2008. Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography. *Applied optics* 47, 19 (2008), D117–D127.
- RG Lane. 1991. Phase retrieval using conjugate gradient minimization. *Journal of Modern Optics* 38, 9 (1991), 1797–1813.
- Douglas Lanman and David Luebke. 2013. Near-eye light field displays. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 220.
- Detlef Leseberg and Christian Frère. 1988. Computer-generated holograms of 3-D objects composed of tilted planar segments. *Applied optics* 27, 14 (1988), 3020–3024.
- LB Lesem, PM Hirsch, and JA Jordan. 1969. The kinoform: a new wavefront reconstruction device. *IBM Journal of Research and Development* 13, 2 (1969), 150–155.
- Anat Levin, Haggai Maron, and Michal Yarom. 2016. Passive light and viewpoint sensitive display of 3D content. In *2016 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–15.
- Gang Li, Dukho Lee, Youngmo Jeong, Jaebum Cho, and Byoungcho Lee. 2016. Holographic display for see-through augmented reality using mirror-lens holographic optical element. *Optics letters* 41, 11 (2016), 2486–2489.
- Mark Lucente and Tinsley A Galyean. 1995. Rendering interactive holographic images. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 387–394.
- Mark E Lucente. 1993. Interactive computation of holograms using a look-up table. *Journal of Electronic Imaging* 2, 1 (1993), 28–34.
- Andrew Maimone, Andreas Georgiou, and Joel S Kollin. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 85.
- Stefano Marchesini, Yu-Chao Tu, and Hau-tieng Wu. 2016. Alternating projection, ptychographic imaging and phase synchronization. *Applied and Computational Harmonic Analysis* 41, 3 (2016), 815–851.
- Nobuyuki Masuda, Tomoyoshi Ito, Takashi Tanaka, Atsushi Shiraki, and Takashige Sugie. 2006. Computer generated holography using a graphics processing unit. *Optics Express* 14, 2 (2006), 603–608.
- Kyoji Matsushima. 2005. Computer-generated holograms for three-dimensional surface objects with shade and texture. *Applied optics* 44, 22 (2005), 4607–4614.
- 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.
- Kyoji Matsushima, Masaki Nakamura, and Sumio Nakahara. 2014. Silhouette method for hidden surface removal in computer holography and its acceleration using the switch-back technique. *Optics express* 22, 20 (2014), 24450–24465.
- Nitish Padmanaban, Yifan Peng, and Gordon Wetstein. 2019. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 214.
- Yifan Peng, Xiong Dun, Qilin Sun, and Wolfgang Heidrich. 2017. Mix-and-match holography. *ACM Transactions on Graphics* (2017).
- Christoph Petz and Marcus Magnor. 2003. Fast hologram synthesis for 3D geometry models using graphics hardware. In *Proc. SPIE*, Vol. 5005. 266–275.
- Stephan Reichelt, Ralf Häussler, Gerald Fütterer, Norbert Leister, Hiromi Kato, Naru Usukura, and Yuuichi Kanbayashi. 2012. Full-range, complex spatial light modulator for real-time holography. *Optics letters* 37, 11 (2012), 1955–1957.
- 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 Transactions on Graphics (TOG)* 36, 6 (2017), 236.
- Quinn YJ Smithwick, James Barabas, Daniel E Smalley, and V Michael Bove. 2010. Interactive holographic stereograms with accommodation cues. In *Practical Holography XXIV: Materials and Applications*, Vol. 7619. International Society for Optics and Photonics, 761903.
- Tullio Tommasi and Bruno Bianco. 1993. Computer-generated holograms of tilted planes by a spatial frequency approach. *JOSA A* 10, 2 (1993), 299–305.
- Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2017. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), 8798–8807.
- Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, Vol. 2. Ieee, 1398–1402.
- James P Waters. 1966. Holographic image synthesis utilizing theoretical methods. *Applied physics letters* 9, 11 (1966), 405–407.
- Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini. 2012. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems* 28, 11 (2012), 115010.
- Gordon Wetstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. 2012. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. (2012).
- Masahiro Yamaguchi, Hideshi Hoshino, Toshio Honda, and Nagaaki Ohyama. 1993. Phase-added stereogram: calculation of hologram using computer graphics technique. In *Proc. SPIE*, Vol. 1914. 25–31.
- 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. *Optics express* 23, 25 (2015), 32025–32034.
- Hiroshi Yoshikawa, Takeshi Yamaguchi, and Hiroki Uetake. 2016. Image quality evaluation and control of computer-generated holograms. In *Practical Holography XXX: Materials and Applications*, Vol. 9771. International Society for Optics and Photonics, 97710N.
- Hao Zhang, Yan Zhao, Liangcai Cao, and Guofan Jin. 2015. Fully computed holographic stereogram based algorithm for computer-generated holograms with accurate depth

- cues. *Optics express* 23, 4 (2015), 3901–3913.
- Hao Zhang, Yan Zhao, Liangcai Cao, and Guofan Jin. 2016. Layered holographic stereogram based on inverse Fresnel diffraction. *Applied optics* 55, 3 (2016), A154–A159.
- Jingzhao Zhang, Nicolas Pégard, Jingshan Zhong, Hillel Adesnik, and Laura Waller. 2017. 3D computer-generated holography by non-convex optimization. *Optica* 4, 10 (2017), 1306–1313.
- Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Yan Zhao, Liangcai Cao, Hao Zhang, Dezhao Kong, and Guofan Jin. 2015. Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method. *Optics express* 23, 20 (2015), 25440–25449.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision (ICCV)* (2017), 2242–2251.