HoOPOE: A Holography-Optimized Perceptual ObjectivE function

ETHAN CHEN, University of Rochester, USA and Reality Labs Research, USA
XI PENG, University of North Carolina at Chapel Hill (UNC), USA
ALEXANDRE CHAPIRO, Reality Labs Research, USA
DARIO SEYB, Reality Labs Research, USA
MICHAEL GRUNDMANN, Reality Labs Research, USA
CARL MARSHALL, Reality Labs Research, USA
SUSHANT KONDGULI, Reality Labs Research, USA
PRANEETH CHAKRAVARTHULA, University of North Carolina at Chapel Hill (UNC), USA
YUHAO ZHU, University of Rochester, USA

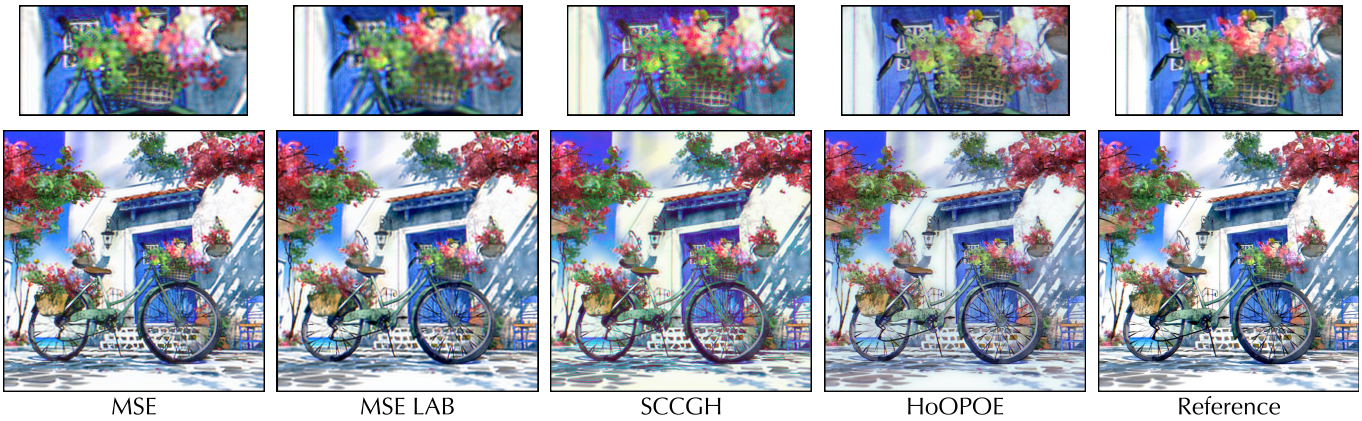

Fig. 1. A comparison of a HoOPOE-trained 3D hologram against those trained under three other loss functions: Mean Squared Error (MSE), Mean Squared Error in LAB space (MSE LAB) [International Commission on Illumination 2019], and SCCGH [Markley et al. 2023]. The results shown here simulate temporal integration at 480 fps. Note that the subjective differences are more significant to an in-person observer using our physical setup, due to the differences in viewing conditions (e.g., spatial/temporal resolution, display luminance). A user study shows that our loss yields significantly better results than those of the other approaches. Copyright for the Blender scene (licensed under CC-BY-SA) used to generate this focal stack goes to Tada [2023].

Holographic devices must reconstruct an entire light field given a finite number of spatial light modulator (SLM) pixels, and must do so using scalar-wave diffraction from coherent illumination alone. As a result, computer-generated holography (CGH) often has speckling artifacts. We propose HoOPOE (Holography-Optimized Perceptual ObjectivE function), a CGH-specific, perceptually-aware loss function that hides speckles from human vision. HoOPOE is effective for CGH optimizations because it accounts for the human visual system's sensitivity to spatio-temporal frequencies and color, and its parameters are tuned to maximize its performance as a loss function for CGH. We use HoOPOE to generate holograms for mesh-derived focal stacks. We then conduct a user study to benchmark HoOPOE-generated holograms against existing loss functions. HoOPOE yields the most visually pleasing results; users prefer HoOPOE 68% of the time over the next best objective function on average.

CCS Concepts: • **Computing methodologies** → **Perception**; **Mixed / augmented reality**; • **Hardware** → *Displays and imagers.*

Additional Key Words and Phrases: Computer-Generated Holography, Perceptual Loss Functions, Image Quality Metrics, Augmented Reality

ACM Reference Format:
Ethan Chen, Xi Peng, Alexandre Chapiro, Dario Seyb, Michael Grundmann, Carl Marshall, Sushant Kondguli, Praneeth Chakravarthula, and Yuhao Zhu. 2026. HoOPOE: A Holography-Optimized Perceptual ObjectivE function. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26), December 01–04, 2026, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3829340.3842313

Authors' Contact Information: Ethan Chen, echen48@ur.rochester.edu, University of Rochester, Rochester, New York, USA and Reality Labs Research, Sunnyvale, California, USA; Xi Peng, xipeng@cs.unc.edu, University of North Carolina at Chapel Hill (UNC), Chapel Hill, North Carolina, USA; Alexandre Chapiro, achapiro@meta.com, Reality Labs Research, Sunnyvale, California, USA; Dario Seyb, dario.seyb@gmail.com, Reality Labs Research, Redmond, Washington, USA; Michael Grundmann, mgrundmann@meta.com, Reality Labs Research, Redmond, Washington, USA; Carl Marshall, csmarshall@meta.com, Reality Labs Research, Redmond, Washington, USA; Sushant Kondguli, sushant.kondguli@gmail.com, Reality Labs Research, Sunnyvale, California, USA; Praneeth Chakravarthula, cpk@cs.unc.edu, University of North Carolina at Chapel Hill (UNC), Chapel Hill, North Carolina, USA; Yuhao Zhu, yzhu@rochester.edu, University of Rochester, Rochester, New York, USA.

This work is licensed under a Creative Commons Attribution 4.0 International License.
SA Conference Papers '26, Kuala Lumpur, Malaysia
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2842-6/2026/12
https://doi.org/10.1145/3829340.3842313

1 Introduction

Holographic displays have many advantages over traditional emissive displays for virtual and augmented reality. Conventional displays present flat images at a fixed focal length, relying only on binocular disparity to generate depth cues. In contrast, holographic displays can theoretically replicate the entire outgoing light field by using a spatial light modulator (SLM) to manipulate the wavefront of light emitted from coherent (laser) sources. This allows holographic displays to provide the correct depth cues to drive accommodation, a key factor in immersive 3D experience. Additionally, holographic displays have naturally wide color gamuts, can support larger eyeboxes, and can eliminate bulky waveguides and projective lenses.

However, the quality of images rendered on holographic displays is harmed by speckle noise, which arises from the constructive and destructive interference of coherent light emitted by the holographic display. Speckle noise obscures fine detail, reduces contrast, and introduces distracting visual artifacts that undermine the advantages that holographic displays would otherwise offer. Unlike many visual artifacts that can be solved through calibration or post-rendering correction, speckle is intrinsic to holographic displays.

The classic approach to speckle reduction is to increase the degrees of freedom in the system. This is achieved through either temporal multiplexing, which requires operating the SLM at a high frame rate, or through a high-resolution SLM. The increased degree of freedom allows us to more easily synthesize a proper spatiotemporal SLM pattern to match a given wavefront. Neither approach is desirable, however, as they significantly increase the computational cost and power consumption of the system.

We take the reverse perspective and ask: without increasing the system complexity, how can we most effectively hide speckles from human vision — guided by a perceptual loss function?

Prior perceptual metrics such as CVVDP [Mantiuk et al. 2024], MILO [Cogalan et al. 2025], LPIPS [Zhang et al. 2018], and DISTS [Ding et al. 2020], while effective as quality predictors, may not be effective as objective functions for CGH algorithms. These metrics are often trained and calibrated on image quality datasets [Ding et al. 2021], where natural images are corrupted with a specific set of artifacts (e.g., blur, brightness changes, color saturation, spatial nonuniformity). When using these metrics as loss functions for CGH, the optimizer will discover stimuli that were unseen during metric calibration. "Adversarial examples" can easily "hack" the perceptual metrics: trained stimuli have low loss but are visually objectionable (see Fig. 2 and Supplementary Material **??** for examples).

Existing perceptual metrics, understandably, perform suboptimally on these types of stimuli, because these stimuli are only produced during CGH training (when the metrics are used as a loss function) and are not commonly present in existing image quality datasets used for training metrics. There are no large-scale subjective datasets available that would allow for a thorough recalibration.

In this work, we introduce HoOPOE (Holography-Optimized Perceptual ObjectivE), a loss function designed specifically for holographic image optimization. Our approach is guided by a key requirement: effective perceptual loss functions must not only correlate with human perception, but must also remain robust against adversarial stimuli encountered during optimization. HoOPOE accounts for the contrast sensitivity of human vision to stimuli of varying spatiotemporal frequencies and color, and is specifically optimized for stimuli that one might encounter during hologram synthesis.

Using a gradient-descent CGH algorithm [Liu and Takaki 2020; Peng et al. 2020; Zhang et al. 2017] as a case study, we show that HoOPOE generates more visually pleasing holograms compared to existing loss functions. The loss function, by being temporally aware, also informs the design of time multiplexing: the biggest perceptual gains occur when increasing the frame rate from 120 frames per second to 240 fps. The increase in perceptual quality for the jump between 240 fps and 480 fps was not statistically significant.

We make the following contributions:

- We introduce HoOPOE, a novel perceptual loss function for CGH that accounts for various aspects of human vision, while remaining robust during optimization.
- We evaluate the performance of state-of-the-art perceptual quality metrics, and demonstrate limitations when used as holographic optimization losses.
- We benchmark HoOPOE against existing loss functions using a pairwise comparison psychophysical test. We show that HoOPOE generates holograms that are significantly more visually pleasing.

Our project website is located at https://horizon-lab.org/hoopoe/.

2 Related Work

2.1 Perceptual Loss Functions for CGH

The shift from alternating-projection methods to gradient-based optimization [Chakravarthula et al. 2019; Peng et al. 2020] reframed CGH as a question of *what to optimize* rather than *how to optimize*. As it is impossible to remove all speckles from scalar wave diffraction alone using coherent illumination [Goodman 2007; Iwai and Asakura 1996], perceptual models can play a critical role in improving visual quality by hiding speckles where they are least visible.

Yang et al. [2022] conducted a comprehensive benchmark of image quality metrics as holographic objective functions. The authors evaluated 12 metrics using a two-alternative forced choice test, and found that MS-SSIM [Wang et al. 2003] outperformed all alternatives. However, their study was restricted to monochromatic image content. The metrics they benchmarked did not account for chromatic or temporal aspects of perception.

For color holography, Simultaneous Color Computer Generated Holography (SCCGH) [Markley et al. 2023] proposed a novel objective function that calculates loss in a cone opponent color space. SCCGH applies a low-pass filter to the image to eliminate theoretically invisible signals of high spatial frequency. While this approach models both color perception and spatial contrast sensitivity, it does not account for temporal contrast sensitivity.

Others explored gaze-contingent techniques to optimize holographic rendering. Chakravarthula et al. [2021] exploits differences in peripheral and foveal contrast sensitivity to suppress speckle more aggressively in foveal vision. Walton et al. [2022] extends Freeman and Simoncelli [2011] and Walton et al. [2021] to holography by formulating a loss function that accounts for ventral metamerism. Chakravarthula et al. [2022a] further improves holographic rendering by accounting for the pupil size and position of the viewer. These



(a) Reference image. JOD: 10 (b) Predicted JOD: 8.51 (c) Predicted JOD: 7.03

Fig. 2. CVVDP scores Fig. 2b (trained by CVVDP) over 1 JOD higher than Fig. 2c (trained by HoOPOE), even though Fig. 2c is visually preferable. CVVDP's standard full HD display model was used to score both images. See Supplementary Material **??** for more examples. Copyright for the original 3D model (licensed CC-BY-SA) goes to Tada [2023].

methods must be computed in real-time, as they rely on eye tracking data. While they do account for variations in spatial contrast sensitivity, these solutions are not temporally aware.

Not every CGH algorithm can accommodate perceptual loss functions, however. Closed-form solutions, like Maimone et al. [2017], Choi et al. [2025], and Padmanaban et al. [2019], sidestep optimization altogether. Because no objective is minimized, the choice of loss function does not arise in these pipelines.

2.2 Temporal Multiplexing in Holography

Temporal multiplexing has emerged as a practical approach for speckle reduction. Uncorrelated speckle noise displayed at a sufficiently high frame rate becomes perceptually invisible, as the human visual system integrates across time.

Takaki and Yokouchi [2011] and Mori et al. [2014] decomposed an image into a set of point groups, then displayed them sequentially to recreate the original image. Choi et al. [2022] further advanced temporal multiplexing by demonstrating that jointly optimizing hologram frames across time significantly increased the image reconstruction quality. The authors calculate loss by time-averaging the holographic image to simulate the temporal summation of the human eye. Sha et al. [2024] and Lee et al. [2022] also tabulate loss against the time-averaged holographic image.

However, existing works on temporal multiplexing in holography share a common limitation. They do not account for the human visual system's variable sensitivity across temporal frequencies. Prior work makes the implicit assumption that all noise that is above a specific temporal frequency threshold is equally imperceptible with an instantaneous cutoff. In reality, contrast sensitivity functions show that sensitivity declines gradually with increasing temporal frequency [Ashraf et al. 2024; Robson 1966]. This suggests that weighting errors across temporal frequency bands could yield perceptual benefits beyond what current methods achieve.

2.3 Perceptual Metrics vs. Objective Functions

Image quality metrics evolved from simple pixel-wise comparisons to sophisticated models of human perception. Early structural similarity methods like SSIM [Wang et al. 2004] and MS-SSIM [Wang et al. 2003] compare local luminance, contrast, and structure between images. NLPD [Laparra et al. 2017] computes perceptual distance through modeling contrast and gain control via Laplacian pyramid decomposition. More recent learned metrics like LPIPS [Zhang et al. 2018] are trained to be strong predictors of human perceptual judgments. Visible Difference Predictors (VDPs) [Mantiuk et al. 2021, 2024] are perceptual quality metrics that account for contrast sensitivity across different spatial and temporal frequencies, color channels, and eccentricities.

Many of these metrics perform well on image quality benchmark datasets like TID2013 [Ponomarenko et al. 2015]. However, a good predictor is not necessarily a good loss function, and optimizing for perceptual metrics may not yield visually pleasing images [Mustafa et al. 2022]. Ding et al. [2021] conducted a study demonstrating this gap, showing that many metrics with good evaluation performance were outperformed by MSE when used as loss functions for image generation tasks. Similarly, we observe that when used as an objective function, CVVDP generates "adversarial stimuli" that achieve high predicted quality scores, but results may still appear visibly degraded (Fig. 2 and Supplementary Material **??**).

Complex perceptual metrics can be highly non-convex, hurting convergence [Mantiuk et al. 2024, Sec. 6.2]. Interestingly, for speckle suppression, we find that the non-convexity of existing metrics is *not* the primary reason they perform poorly as loss functions. For example, in Supplementary Material **??** we show that although the optimization landscape of CVVDP contains many basins and the optimizer can be trapped in different ones depending on initialization, these basins tend to achieve similarly low loss values. We observe the same behavior for other metrics, consistent with prior observations in learning theory research [Allen-Zhu et al. 2019; Du et al. 2019]. Instead, the ineffectiveness of existing metrics such as CVVDP as a loss function for CGH is caused by "adversarial stimuli" whose quality is systematically misjudged by the metric.

3 HoOPOE

HoOPOE is a loss function grounded in human vision and tailored for computer-generated holography. Our loss function takes as input a reference video (ground truth) and a test video (generated by a

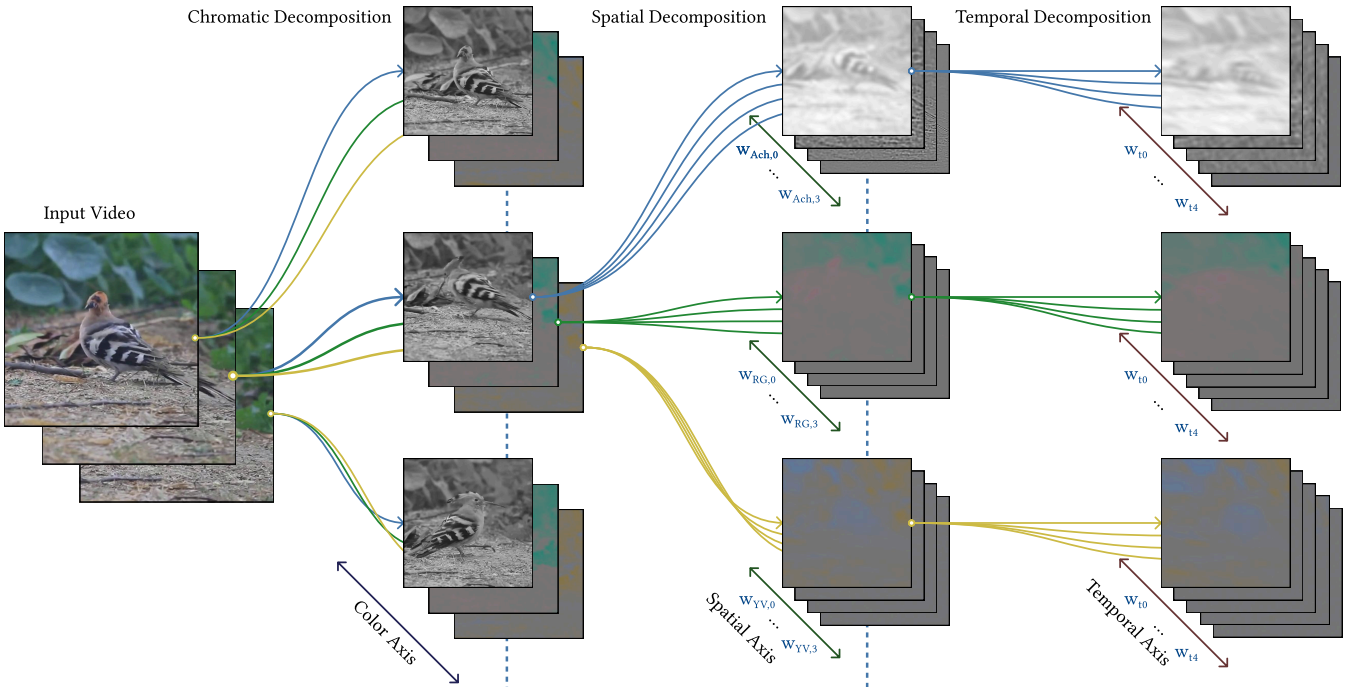

Fig. 3. Visualization of HoOPOE's architecture. We decompose each video frame into three chromatic channels. Each channel is then decomposed into a set of N spatial frequency bins through a Laplacian pyramid. Each combination of color channel c and spatial-frequency s is associated with a trainable weight $w_{c,s}$. For each c and s bin, we further decompose the content into a set of M temporal-frequency bins through a Laplacian pyramid. Each temporal-frequency bin t is associated with a trainable weight w_t. The final loss is a weighted MSE across the bins between the reference and test videos. To avoid visual clutter, only one spatial band per channel is temporally decomposed in this graphic. Copyright for the original video (licensed CC-BY) goes to Ram [2018].

CGH algorithm), and outputs a single loss value representing the perceptual difference between the two. We provide an overview first, followed by elaborations on the design details.

3.1 Overview

Our loss function is designed with three goals in mind. First, it must be perceptually grounded. To that end, HoOPOE converts both the reference and test videos into an opponent color space approximating the opponent process of color vision. Each color channel (achromatic, yellow-blue, and red-green) is then decomposed into a set of bins consisting of varying spatial and temporal frequency bands. Operating in the spatio-temporal-chromatic space rather than the RGB space allows us to, through learned weighting later, account for the human contrast sensitivity to varying spatial and temporal frequencies across color channels.

Second, the loss must facilitate the smooth, gradient-based optimization that CGH algorithms usually rely on. To that end, we compute the loss as a simple weighted mean squared error (MSE) across the spatio-temporal-chromatic bins between the reference and test videos. Specifically, each spatio-temporal-chromatic bin is assigned a trainable weight; we calculate the difference between the reference and test videos on each weighted bin, and pool the per-bin difference together to form the loss using MSE. A visual representation of our loss function's pipeline can be found in Fig. 3.

Formally, the HoOPOE loss function is defined as

$$l_{\text{HoOPOE}} = \sum_{c,s,t} \frac{w_{c,s} \cdot w_t}{\|P_{s,t,c}\|} \sum_{p \in P_{s,t,c}} (L_{\text{ref}}(p,c,s,t) - L_{\text{test}}(p,c,s,t))^2. \quad (1)$$

where $L(p,c,s,t)$ denotes the magnitude at pixel p in color channel c, spatial Laplacian pyramid band s, and temporal Laplacian pyramid band t. $P_{s,t,c}$ is the set of all pixels across all spatio-temporal bands and chromatic channels. $w_{c,s}$ and w_t are trainable weights representing the perceptual importance of each spatio-chromatic bin and temporal bin, respectively.

The loss formulation produces a quadratic (convex) objective function in the prediction error, with well-conditioned gradients that support efficient optimization. At the same time, the trainable parameters $w_{c,s}$ and w_t permit perceptually motivated and CGH-specific weighting across bins, which leads us to our third point.

Finally, the loss must be robust against adversarial stimuli encountered during optimization. These stimuli minimize the loss while being perceptually objectionable. The key idea is that the spatio-temporal-chromatic weights in the loss function should be tuned with the CGH algorithm in the loop. In this way, the loss function can correctly score the perceptual quality of content with CGH-specific artifacts. This CGH-in-the-loop loss function tuning amounts to a challenging bi-level optimization problem. Solving that is the goal of this work.

3.2 Color Space Conversion

We first convert the input video data into the CIELAB color space [International Commission on Illumination 2019], used as the basis for our loss function. The CIELAB color space was derived with the goal of perceptual uniformity. Calculating color difference in a perceptual color space is critical for a loss function to produce perceptually accurate images.

Compared to more recent color spaces designed for perceptual uniformity, such as CAM16 UCS [Li et al. 2017], ICtCp [International Telecommunication Union 2025], and OKLab [Ottosson 2020], CIELAB uses a simpler transformation from and to display-referred color spaces, such as sRGB. In practice, we found that this reduces the complexity of the optimization landscape and leads to better convergence. Alternative physiologically relevant color spaces such as the color-opponent DKL space were not employed. As they do not aim to be perceptually uniform, the gradients in this space are less useful for guiding optimization.

3.3 Spatial and Chromatic Weighting

We then decompose each color channel into a set of N spatial-frequency bands through a Laplacian pyramid. Each pyramid level acts as a band-pass filter, representing a particular spatial frequency band. We associate each of the $3 \times N$ pyramid levels with a weight $w_{c,s}$, which determines the relative importance of each spatial frequency band s and color channel c in the perceptual loss. We fit the weights specifically for computer-generated holograms according to their perceptual quality using the following procedure.

Optimization Space. We consider $N = 4$ to reduce the optimization space. As such, the set of weights $\boldsymbol{w} = \{w_{c,s}\}$ is composed of 12 scalars, corresponding to four frequency bands and three color channels such that each $w_{c,s} \in (0, 1]$. We also tune two learning parameters: learning rate and step decay.

Optimization Problem. At a high level, we perform a bi-level optimization. The goal is to optimize a set of weights $\boldsymbol{w}$ of HoOPOE (outer problem) such that the perceptual quality is maximized when a holographic image is trained under HoOPOE with $\boldsymbol{w}$ (inner problem). This bi-level optimization problem is formulated as follows:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{\operatorname{argmax}} \left[l_{\text{scoring}} \left(r, \text{CGH}(\theta^*_{\boldsymbol{w}}) \right) \right], \tag{2a}$$

$$\text{where} \quad \theta^*_{\boldsymbol{w}} = \underset{\theta}{\operatorname{argmin}} \left[l_{\text{HoOPOE}} \left(r, \text{CGH}(\theta); \boldsymbol{w} \right) \right]. \tag{2b}$$

In the outer optimization, $\boldsymbol{w}$ is the set of all 12 $w_{c,s}$ parameters, and l_{scoring} is a scoring function (see below) that scores the perceptual quality, with respect to the reference image r, of a test image generated by the CGH algorithm using the SLM pattern $\theta^*_{\boldsymbol{w}}$. In the inner optimization problem, we are optimizing the SLM pattern θ to minimize the HoOPOE loss parameterized by the parameters $\boldsymbol{w}$.

Reference Image Selection. We select a random set of images from Segment Anything 1B [Kirillov et al. 2023] as hologram optimization targets. Five images in total were selected and cropped to 1080p.

Optimization Strategy. While the inner optimization problem in Eqn. 2b can be solved using first-order, gradient-based methods, the outer optimization problem in Eqn. 2a only has the final value of the inner optimization problem, and cannot compute gradients of the outer loss back to $\boldsymbol{w}$.

Therefore, we use a zeroth-order optimization algorithm CMA-ES [Nomura and Shibata 2024], implemented in the popular framework Optuna [Akiba et al. 2019], to optimize our spatial-color weights $\boldsymbol{w}$. The low dimensionality of $\boldsymbol{w}$ also makes the search space friendly to zeroth-order techniques. It is theoretically possible to solve Eqn. 2a and Eqn. 2b in tandem using hypergradient descent [Maclaurin et al. 2015], which we leave for future work.

Specifically, we repeatedly alternate between adjusting the loss function weights via Optuna and performing CGH optimization via gradient descent in a loop. For each set of weights $\boldsymbol{w}$ proposed by Optuna, we form a loss function, using which we run a complete gradient descent hologram optimization pass using HoOPOE as a loss function. This produces a holographic image. We ran 1300 tuning trials in total.

Perceptual Scoring. We have two options to form $l_{scoring}$ in Eqn. 2a, which scores the quality of CGH content: 1) human-in-the-loop, where humans score CGH stimuli interactively as they are being generated, or 2) leveraging existing perceptual quality metrics such as CVVDP. The former is intractable; we estimate 200 hours per participant for human-in-the-loop optimization.

Although CVVDP was not suitable as a direct loss function to optimize SLM pixels, motivating the development of HoOPOE, it performs well as a perceptual scoring function for optimizing $w_{c,s}$. When CVVDP is used as a loss function, the optimizer will necessarily encounter "adversarial stimuli" that maximize CVVDP's predicted quality while remaining subjectively objectionable (see Fig. 2 and Supplementary Material **??** for examples). As a consequence, we found the resulting gradients unreliable in guiding the optimization. In contrast, we employ CVVDP to generate the 12 weights in the HoOPOE loss. By reducing the degrees of freedom of the optimization problem (from millions of pixels to only 12 parameters), we reduce the possibility of adversarial configurations appearing in the parameter search space.

Interpretation. Recent perceptually-based methods use models of human contrast sensitivity function (CSF) to generate weights for local contrast bands [Mantiuk et al. 2021, 2024; Tariq et al. 2023; Tursun et al. 2019], which has been shown to produce good results when comparing test and reference stimuli at similar luminance levels. We show that our weighting method can be interpreted as a *learned* approach toward contrast weighting. We discuss in Supplementary Material **??** why existing CSF-based weighting does not directly apply.

Using the model of Peli [1990], the local Weber contrast of a pixel p can be expressed by dividing the pixel value in a given Laplacian pyramid level by the corresponding pixel value in the Gaussian pyramid two levels below it in resolution:

$$C(p, s) = \frac{L(p, s)}{G(p, s + 2)}, \tag{3}$$

where $C(p, s)$ is the local contrast of pixel p at level s, $G(p, s + 2)$ is the pixel value at the Gaussian pyramid level $s + 2$, which represents the local adapting/mean luminance, and $L(p, s)$ is the pixel value at

the Laplacian pyramid level s, which represents the local variation from the local mean. C, G, and L are each defined per color channel c, but from this point on we drop the channel index c in the notation for simplicity's sake.

Given that $L(p,s) = G(p,s) - G(p,s+1)$ and $L(p,s+1) = G(p,s+1) - G(p,s+2)$, we know $L(p,s) + L(p,s+1) = G(p,s) - G(p,s+2)$. Therefore, Eqn. 3 can be re-written as:

$$C(p,s) = \frac{L(p,s)}{G(p,s) - (L(p,s) + L(p,s+1))}. \tag{4}$$

Now that each Laplacian level is associated with a weight, we can express a form of weighted contrast as:

$$C_w(p,s) = \frac{w_{c,s} L(p,s)}{G(p,s) - (w_{c,s} L(p,s) + w_{c,s+1} L(p,s+1))} \tag{5a}$$

$$= \frac{G(p,s) - (L(p,s) + L(p,s+1))}{G(p,s) - (w_{c,s} L(p,s) + w_{c,s+1} L(p,s+1))} \cdot C(p,s). \tag{5b}$$

Therefore, weighting the Laplacian pyramid levels has the effect of weighting local contrast of pixels, where the contrast weights are dependent on both the local image statistics (e.g., $G(p,s)$ and $L(p,s)$) as well as the learned weights assigned to each Laplacian level (e.g., $w_{c,s}$). Compared to conventional contrast weighting (Eqn. **??** in Supplementary Material **??**), our weighting provides two degrees of freedom to weight the contrast of each pixel[1] learned in accordance with perceptual quality.

3.4 Temporal Weighting

We can also extend HoOPOE to account for the temporal dimension of human vision. For each combination of spatial-frequency band and chromatic channel, we build a temporal Laplacian pyramid with M layers, each of which is associated with a weight w_t. We use a value of $M = 3$ to obtain a fine-grained decomposition of temporal frequency, which is needed to explore the effect of temporal multiplexing commonly used in CGH. The set of weights $\boldsymbol{w}_t = \{w_t\}$ determines the relative importance of errors in different temporal frequency bands of the reference and test videos.

The temporal weighting is considered in isolation from the spatio-chromatic weighting to reduce the parameter space. This is supported by the fact that spatial and temporal filtering in retinal processing are approximately separable, or are modeled as such (for chromatic channels) [Ashraf et al. 2024]. As a result, there are $3 \times N \times M$ bins, but only $3 \times N + M$ trainable weights.

We determine w_t through a human-in-the-loop procedure. Specifically, we first use a Sobol sampler to draw 50 different temporal weight configurations, each of which is used to train the CGH algorithm and produce a corresponding set of stimuli. We then apply a single-elimination tournament approach to select the weighting that yields the most preferred video output. Unlike spatial weighting, human-in-the-loop selection is viable here, as $\boldsymbol{w}_t$ is composed of a set of 3 weights, as opposed to the 12 weights of $\boldsymbol{w}_{c,s}$. The parameter search space grows exponentially as a function of parameter count, so $\boldsymbol{w}_t$'s search space is 9 orders of magnitude smaller than $\boldsymbol{w}_{c,s}$.

[1]These two degrees of freedom are shared by all the pixels at the same pyramid layer.

4 Validation

We conduct an IRB-approved user study to compare HoOPOE against prior loss functions for CGH.

4.1 Stimuli

We use HoOPOE to generate 5 CGH stimuli for subjective validation. We generate a focal stack of images derived from Blender's demo scene files. For each demo scene, we generated 5 images from a Blender camera with 5 different focal distances. These images have no overlap with the images used to fit the HoOPOE weights, and can be found in Supplementary Material **??**.

4.1.1 CGH Stimuli Generation. We use a forward model based on the Angular Spectrum Method (ASM) optimized through gradient descent [Chakravarthula et al. 2019, 2022b; Peng et al. 2020]. Specifically, we use the ASM propagation model from Peng et al. [2020]'s repository, modified to incorporate a Fourier aperture of 0.5. More details can be found in Supplementary Material **??**. We use SGD rather than neural network-based generation to isolate the effect of loss function choice on final image quality. Neural architectures like Shi et al. [2021], Peng et al. [2020], and Liu et al. [2025] could potentially introduce inductive biases like network capacity that would obscure the effect of loss function on visual quality.

We make the following assumptions:

- We optimize a focal stack for 5 propagation distances between 0.5 cm and 2.5 cm, uniformly spaced
- Phase-only SLM with a resolution of 1920×1080
- Pitch size of 8 μm
- RGB laser wavelengths of 638 nm, 520 nm, and 450 nm

For each RGB frame of the ground truth image, we generate three independent SLM patterns, one per laser wavelength. We do not perform any ROI cropping.

4.1.2 Baselines. In addition to HoOPOE, we benchmark three other loss functions in our user study. All loss functions, including HoOPOE, had their learning rate and step decay parameters tuned using Optuna [Akiba et al. 2019] to ensure that any difference in quality stems from the loss function and not from hyperparameter selection. Latency and memory overhead for training under each loss function are reported in Supplementary Material **??**.

Simultaneous Color Computer Generated Holography. While not the central focus of their paper, Markley et al. [2023] (SCCGH hereafter) proposes a perceptual loss function. The key idea of SCCGH's loss function is to filter out theoretically invisible high spatial frequency features before using MSE loss on the filtered stimuli. The original implementation in the paper uses empirical models of human vision. We implement an enhanced version using more psychophysically accurate models of human vision. See Supplementary Material **??** for more details.

Mean Squared Error. We also benchmark the conventional MSE metric, which is a simple yet common metric used in prior CGH literature [Choi et al. 2022; Peng et al. 2020; Yang et al. 2022]. In particular, we find that forming the MSE in the holographic display's color space (as opposed to other color spaces) performs the best.

Ablation. We perform an ablation study on HoOPOE by testing each addition in the HoOPOE pipeline in our validation study. We test the following losses as part of our ablation study:

- MSE tabulated in CIELAB [International Commission on Illumination 2019]. This represents HoOPOE in the absence of any frequency weighting (labeled MSE LAB).
- HoOPOE with spatial-chromatic weightings only, without temporal frequency weighting (labeled Spatial HoOPOE).

Temporal multiplexing. All tested conditions except HoOPOE optimize the eight frames of the stimulus independently, relying on stochastic decorrelation for speckle reduction. HoOPOE is the exception, as HoOPOE's mathematical formulation considers temporal frequency when tabulating loss (Sec. 3.4), and thus jointly optimizes all eight frames together. We do not adopt the time-averaged optimization technique of Choi et al. [2022] for the baseline loss functions as it is incompatible with the conventional display we perform validation on. When loss is tabulated on a time-averaged image, individual frames are free to leave our OLED's gamut, so long as the average image is close to the reference stimulus. These out-of-gamut pixels, when gamut-clipped and displayed sequentially, result in visible discoloration. This discoloration is an artifact of our display, not of the hologram itself. For reference, 31% of pixels are out-of-gamut for jointly optimized 8-frame MSE, as opposed to 1% of pixels when the frames are independently optimized.

Other Losses. We chose these loss functions as they empirically performed the best compared to a range of other alternatives that we have experimented with, including SSIM [Wang et al. 2004], MS-SSIM [Wang et al. 2003], MILO [Cogalan et al. 2025], ColorVideoVDP [Mantiuk et al. 2024], MSE in CAM16 UCS [Li et al. 2017], and MSE in OKLAB [Ottosson 2020]. We include examples of stimuli generated with these loss functions in Supplementary Material **??**.

Non-iterative CGH. We chose to exclude non-iterative holography techniques (like those described in Maimone et al. [2017], Padmanaban et al. [2019]), as these frameworks are not compatible with perceptual loss functions. Subsequent work [Chakravarthula et al. 2019, 2022c; Peng et al. 2020] has reported higher reconstruction quality for iterative, gradient-based methods.

4.2 Psychophysical Task

Each loss function was evaluated at four frame rates: 60, 120, 240, and 480 fps. This yields a total of 21 conditions per image (5 loss functions × 4 frame rates, plus ground truth). Each stimulus consisted of an 8-frame holographic video. Frames were looped continuously.

Procedure. We use a 2-Interval Forced Choice (2IFC) task for the psychophysics; a visualization of the task is in Fig. 4. Participants are presented with three stimuli: a central reference stimulus and two comparison stimuli. The stimuli are rendered on the center 1920×1080 pixels of the screen. Participants switch between which stimuli they are viewing using keyboard inputs. After the participant views all three stimuli, they are prompted to input which one of the two comparison stimuli they perceive to be closer to the reference stimulus. This psychophysical task is similar to Chen et al. [2024]. The two comparison stimuli are generated by randomly selecting two of the 21 possible conditions, which are then randomly assigned to either side. Despite optimizing the hologram for five planes, only the simulated image at 1.5 cm is shown to the user. We perform this simplification to reduce the complexity of the 2IFC task. The entire set of generated stimuli can be found in Supplementary Material **??**.

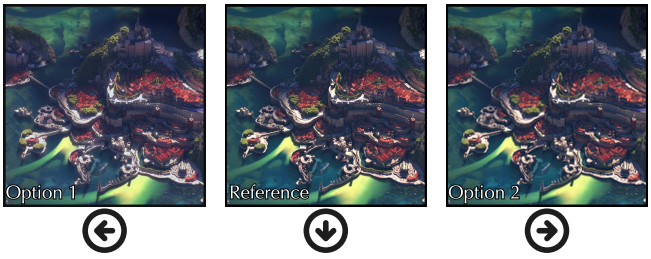

Fig. 4. An example of a 2IFC task. Participants use arrow keys to switch between three stimuli – two CGH renders and one ground truth stimulus. After viewing all three stimuli, the participant is allowed to select which one of the two options they perceive to be a more faithful recreation of the ground truth. This methodology is similar to Chen et al. [2024]. Images are cropped to 1080×1080 here, but are full HD in the actual study. Copyright for the original 3D model (licensed CC-BY-SA) goes to Krynski [2022].

Adaptive Sampling. Since we are benchmarking 5 scenes over 21 conditions, 6,300 pairwise comparisons are required per participant, which, if done naively, would take roughly 9 hours per participant. We utilize the ASAP [Mikhailiuk et al. 2021] method to sample pairwise comparisons to maximize information gain. This cuts down the time needed to obtain a good fit to under an hour per participant.

Participants. We recruit 12 participants for our validation study (ages 20–30, 1:1 male female ratio). All have normal or corrected-to-normal vision. None had been involved in the research project.

Hardware setup. Stimuli were presented on a Sony INZONE M10S OLED display, which supports refresh rates up to 480 Hz. The display was configured to render SDR sRGB images, with a white point luminance of 275 nits. Participants were comfortably seated 0.7 meters from the display. Head position was stabilized using a chin rest to ensure consistent viewing distance.

4.3 Results

A total of 4,362 pairwise comparisons were performed across 12 participants. Pairwise comparison results were converted into Just Objectionable Difference (JOD) scores via the pwcmp library [Perez-Ortiz and Mantiuk 2017]. 1 unit of JOD corresponds to a 75% chance of a participant preferring one stimulus over another. We used 1,000 bootstrap samples to generate the fit. One participant was classified as an outlier by pwcmp and excluded from the data analysis (dist_L of 2.0, which exceeds the customary threshold of 1.5).

Fig. 5 compares the JOD value across all methods (reference has a JOD of 10) across all stimuli; Supplementary Material **??** shows the JOD results for each individual stimulus. Fig. 7 visually compares different methods using one stimulus; Supplementary Material **??** shows more results of other stimuli. Sample SLM patterns for the benchmarked loss functions are in Supplementary Material **??**.

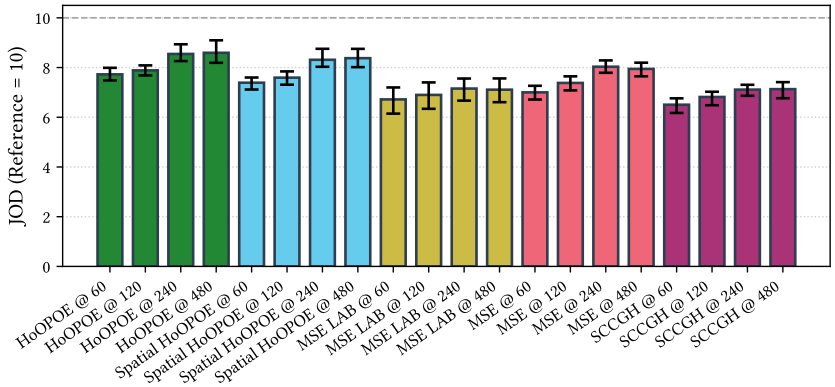

Fig. 5. JOD values tabulated with pwcmp [Perez-Ortiz and Mantiuk 2017] over 4,362 comparisons. The Y axis represents JOD values. A difference of 1 JOD corresponds to a 75% chance of preferring one stimulus over another. Higher JOD values are better. JOD values by stimuli are in Supplementary Material **??**.

Our method significantly outperforms other benchmarks at their respective frame rates ($p < 0.05$, with some $p < 0.001$). Supplementary Material **??** shows the hypothesis test results. Spatial HoOPOE outperforms MSE LAB, which demonstrates the effectiveness of spatial and chromatic weighting. HoOPOE outperforms Spatial HoOPOE, which demonstrates the value of temporal weighting.

While more performance can be gained from increasing the frame rate of the holographic display, there are diminishing returns. The jump from 120 fps to 240 fps represents the greatest improvement in visual quality. Users prefer the 240 fps rendering over the 120 fps rendering 74% of the time. However, users are only 50.7% likely to prefer the 480 fps rendering over the 240 fps rendering. Since display power consumption is a function of refresh rate, our results point to opportunities to make a calculated power-vs-quality trade-off.

To confirm that HoOPOE's advantage is not dependent on temporal multiplexing, we conduct a user study (N=18) with single-frame stimuli. HoOPOE continues to significantly outperform baseline losses under these conditions (Supplementary Material **??**).

5 Discussion

5.1 Generalizability of HoOPOE

The fundamental methodology of HoOPOE generalizes to different hardware setups and viewing conditions. Weights are tuned once for each setup, which takes about 6 hours. Given that hardware and viewing distances are fixed for an XR device, this one-time overhead is negligible. We demonstrate that HoOPOE generalizes to different hardware setups and data distributions, even without tuning.

First we show that HoOPOE generalizes to a different ASM model [Matsushima and Shimobaba 2009]. HoOPOE still outperforms all baselines (Supplementary Material **??**). Additionally, we re-tune HoOPOE's spatial-chromatic weights for two different SLM setups and propagation distances (4.5 μm with 0.25 cm to 1.25 cm distances, 12.5 μm with 2 cm to 10 cm distances). HoOPOE still outperforms all other loss functions in both scenarios, even without temporal weight tuning (Supplementary Material **??**). We also generate holographic images under the aforementioned two SLM hardware setups without re-tuning HoOPOE. A user study (N=18) demonstrates that users still prefer HoOPOE-generated stimuli over those of other losses. Results can be found in Supplementary Material **??**. HoOPOE also generalizes beyond the validation dataset. HoOPOE empirically outperforms other losses in optical-see-through UI content (Supplementary Material **??**).

HoOPOE can generalize across different hardware configurations and viewing conditions because the tuning procedure represents images in device-independent units (e.g., retinal frequency) using CVVDP. Specifically, CVVDP serves as the scoring function used to optimize the HoOPOE weights (Sec. 3.3); it takes as input a display profile that specifies the characteristics of the display, and uses this information to convert the HoOPOE-generated stimuli into display-independent measures of frequency, luminance, and color.

In our current tuning procedure (Sec. 3.4), we utilize humans-in-the-loop to set HoOPOE's temporal weights. In a pilot study, we found that multiple judges converged to the same set of temporal weights, given the same set of Sobol-sampled temporal weight configurations. Despite being tuned by a small number of participants on SA-1B data [Kirillov et al. 2023], the temporal weights generalize to the validation study performed on a different dataset (Blender renders) and on a large number of (different) participants (N=30 in Sec. 4 and Supplementary Material **??**). This suggests participant- and content-wise generalization for the temporal weighting scheme.

5.2 Validation on a Conventional Display

We chose to perform validation on a conventional display rather than a holographic prototype. The goal of HoOPOE is to mask artifacts from the propagation model itself. As is well-known in the literature [Chakravarthula et al. 2020; Peng et al. 2020], an actual CGH display would introduce confounding display-specific artifacts. Validating on a conventional display isolates the contribution of the loss function itself, and is a well-established approach in literature [Chakravarthula et al. 2021; Georgiou et al. 2023; Yang et al. 2022]. HoOPOE is also complementary with prior techniques that address display-specific artifacts (e.g. CITL training [Peng et al. 2020]): HoOPOE can serve as the loss function for such pipelines, where the learned forward model can absorb display-specific distortions.

Simulation-based evaluation on a calibrated conventional display additionally makes our results display-agnostic and reproducible.

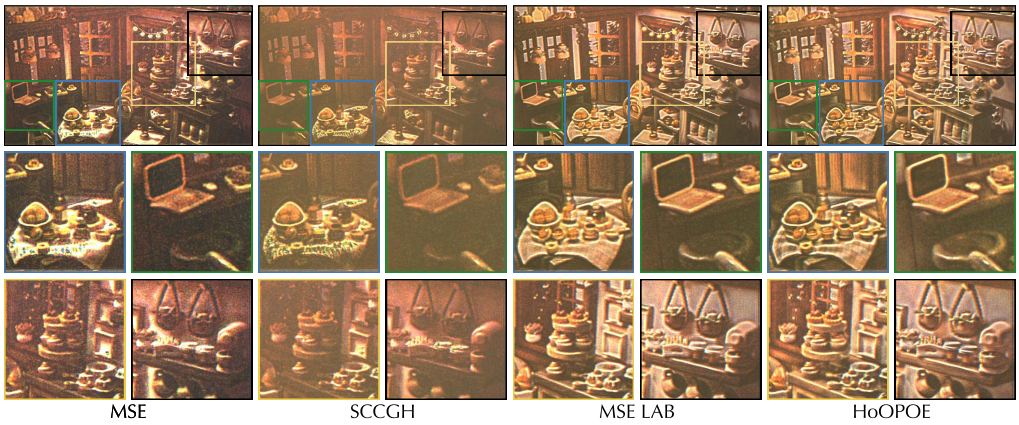

Fig. 6. Camera captures for MSE, SCCGH, MSE LAB, and Spatial HoOPOE. MSE LAB is more color-accurate than MSE. HoOPOE preserves contrast, especially in dark regions. Only 1 frame (3 sequential RGB frames) was optimized and displayed. SLM patterns rendered in Fig. **??**. More details are discussed in Supplementary Material **??**. Copyright for the original 3D model (licensed under CC-BY-SA) goes to Morena [2023].

That said, we performed real hardware studies on a CGH display using the HOLOEYE PLUTO-2.1 SLM. We empirically find that HoOPOE still outperforms all baselines judging by camera captures of the display output. An example is shown in Fig. 6, and more examples and details are discussed in Supplementary Material **??**.

5.3 Tone Enhancement in HoOPOE

Because HoOPOE weights loss across different Laplacian pyramid layers, HoOPOE-reconstructed images may over- or under-emphasize high frequency edges if the weights are improperly tuned. This could make images appear blurred or tone enhanced. However, our 2IFC study allows direct comparisons of image regions against the reference. Artifacts like tonal enhancement and blurring are immediately visible under 2IFC. Users still prefer our method over alternatives, suggesting that the artifacts brought by HoOPOE do not negatively impact user experience. In Supplementary Material **??**, we also analyze the randomness of HoOPOE's SLM patterns.

5.4 HoOPOE as a Perceptual Metric

HoOPOE was designed as a loss function, not as a perceptual metric. As observed by Mustafa et al. [2022], good domain-specific loss functions are not necessarily good visual quality predictors. HoOPOE adopts the same task-specificity principle for CGH: rather than reusing existing metrics [Cogalan et al. 2025; Mantiuk et al. 2024], we instead fit HoOPOE weights against CGH-specific artifacts.

That said, HoOPOE still performs well as a perceptual metric when used on a video quality dataset, XR-DAVID [Chapiro et al. 2024], *without* any retuning. We outperform SCCGH, MSE LAB, MS-SSIM, SSIM, and LPIPS [Zhang et al. 2018] in predicting stimuli quality. A table of Spearman's rank correlation coefficients can be found in Supplementary Material **??**.

5.5 Modeling of Other Visual Phenomena

HoOPOE aims to maximize performance as a loss function by only modeling aspects of human vision that are beneficial for optimization. We experimented with visual masking, local contrast calculation, and contrast sensitivity in HoOPOE. These additions degraded results, and were excluded from the user study. A comparison of these HoOPOE variants is in Supplementary Material **??**.

5.6 HoOPOE-Applicable Optimization Targets

HoOPOE compares reference and test frame data in image space. It is agnostic to how image data is generated. In Sec. 4.1 the focal stack was derived from 3D mesh data. However, HoOPOE is applicable to any CGH pipeline that optimizes an image-space objective, including 3D [Choi et al. 2022; Lee et al. 2022], pupil-aware [Chakravarthula et al. 2022a], and light field based [Schiffers et al. 2023] approaches. HoOPOE is independent of the scene representation and propagation model that produce the images.

In our user studies we benchmark HoOPOE's performance using static reference images, but HoOPOE is applicable to video content as well. The temporal decomposition in Eqn. 1 is performed on both the test and reference stimuli. We demonstrate that HoOPOE is applicable to video reference data generated from a different underlying 3D representation format in Supplementary Material **??**.

6 Conclusion

We propose HoOPOE, a loss function for computer-generated holography. The main contribution is a methodology that is robust against "adversarial stimuli" encountered during CGH optimization that have low loss under existing perceptual metrics but are subjectively objectionable. HoOPOE addresses this by 1) parameterizing the loss function across the spatio-temporal-chromatic dimensions that are perceptually relevant, and 2) tuning HoOPOE parameters with a CGH algorithm in the loop. We show that CGH content generated using HoOPOE is visually preferable to content generated using existing perceptual metrics.

Acknowledgments

The work is partially supported by NSF Awards #2225860, #2126642, and #2044963 as well as a Meta research grant.

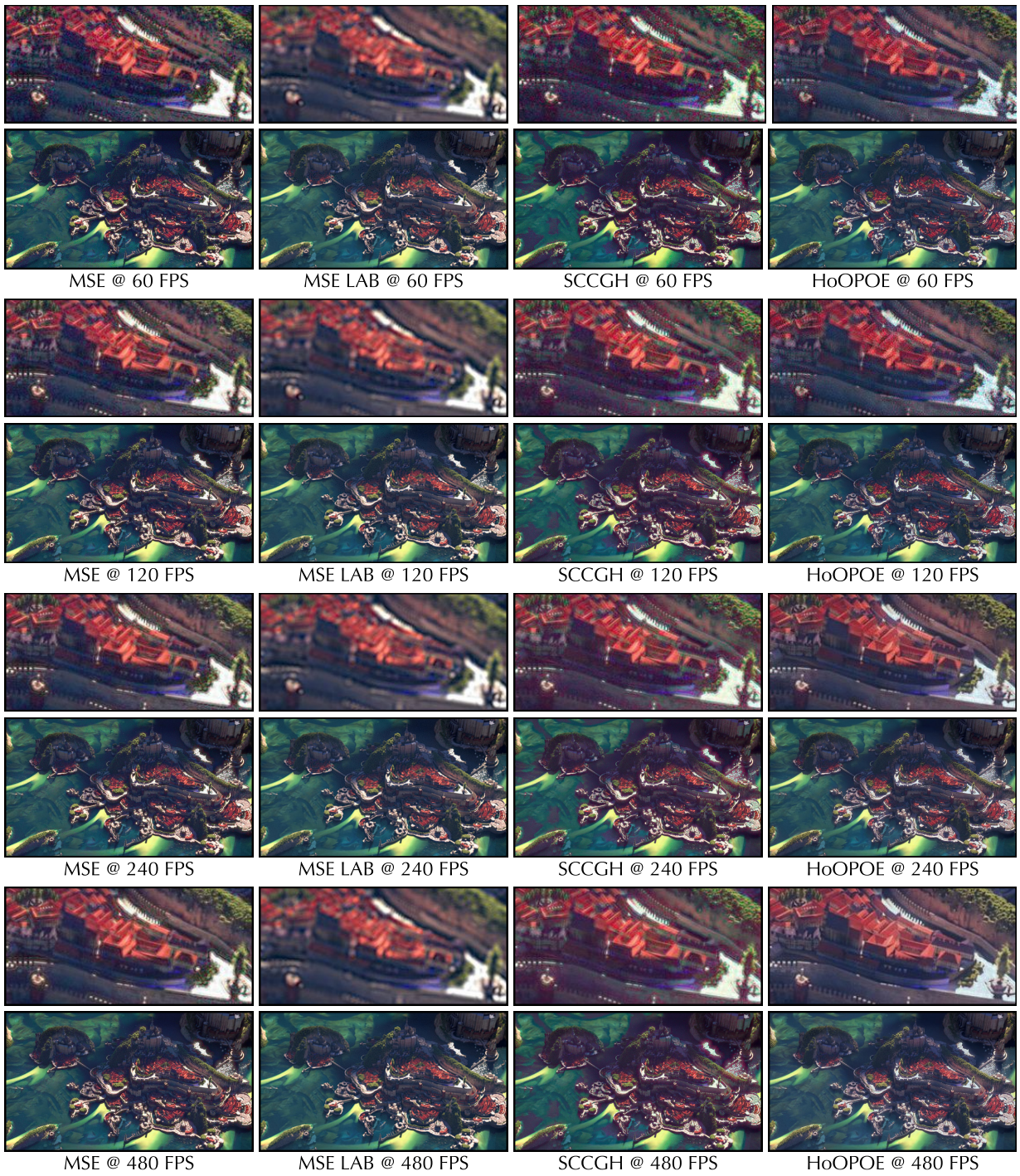

Fig. 7. Images generated by HoOPOE, MSE, MSE LAB, and SCCGH at various frame rates for stimulus 1. The ground truth of this stimulus is Fig. **??**. Other examples can be found in Supplementary Material **??**. Note that this visualization of time multiplexing via time averaging is only an approximation of real-life viewing. Temporal artifacts like flicker are invisible in this render. Copyright for the original 3D model (licensed under CC-BY-SA) goes to Krynski [2022].

References

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK, USA) *(KDD '19)*. Association for Computing Machinery, New York, NY, USA, 2623–2631. doi:10.1145/3292500.3330701

Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. 2019. A Convergence Theory for Deep Learning via Over-Parameterization. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 242–252. https://proceedings.mlr.press/v97/allen-zhu19a.html

Maliha Ashraf, Rafał K Mantiuk, Alexandre Chapiro, and Sophie Wuerger. 2024. castleCSF—A contrast sensitivity function of color, area, spatiotemporal frequency, luminance and eccentricity. *Journal of vision* 24, 4 (2024), 5–5.

Praneeth Chakravarthula, Seung-Hwan Baek, Florian Schiffers, Ethan Tseng, Grace Kuo, Andrew Maimone, Nathan Matsuda, Oliver Cossairt, Douglas Lanman, and Felix Heide. 2022a. Pupil-Aware Holography. *ACM Trans. Graph.* 41, 6, Article 212 (Nov. 2022), 15 pages. doi:10.1145/3550454.3555508

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), 1–13.

Praneeth Chakravarthula, Florian Schiffers, Oliver Cossairt, Ethan Tseng, Seung-Hwan Baek, and Felix Heide. 2022b. Differentiable cameras and displays. In *ACM SIGGRAPH 2022 Courses* (Vancouver, British Columbia, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 213 pages. doi:10.1145/3532720.3535685

Praneeth Chakravarthula, Ethan Tseng, Henry Fuchs, and Felix Heide. 2022c. Hogel-free holography. *ACM Transactions on Graphics* 41, 5 (2022), 1–16.

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 (Nov. 2020), 18 pages. doi:10.1145/3414685.3417846

Praneeth Chakravarthula, Zhan Zhang, Okan Tursun, Piotr Didyk, Qi Sun, and Henry Fuchs. 2021. Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4194–4203.

Alexandre Chapiro, Yuta Asano, Rafal Mantiuk, Param Hanji, and Maliha Ashraf. 2024. XR-DAVID: XR Display Artifact Video Dataset. (2024). doi:10.17863/CAM.108210

Kenneth Chen, Thomas Wan, Nathan Matsuda, Ajit Ninan, Alexandre Chapiro, and Qi Sun. 2024. Pea-pods: Perceptual evaluation of algorithms for power optimization in xr displays. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–17.

Suyeon Choi, Brian Chao, Jacqueline Yang, Manu Gopakumar, and Gordon Wetzstein. 2025. Gaussian Wave Splatting for Computer-Generated Holography. *ACM Trans. Graph.* 44, 4, Article 57 (July 2025), 13 pages. doi:10.1145/3731163

Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, Matthew O'Toole, and Gordon Wetzstein. 2022. Time-multiplexed Neural Holography: A Flexible Framework for Holographic Near-eye Displays with Fast Heavily-quantized Spatial Light Modulators. In *ACM SIGGRAPH 2022 Conference Proceedings* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 32, 9 pages. doi:10.1145/3528233.3530734

Ugur Cogalan, Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Colin Groth. 2025. MILO: A Lightweight Perceptual Quality Metric for Image and Latent-Space Optimization. *ACM Transactions on Graphics (TOG)* 44, 6 (2025), 1–11.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. 2020. Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence* 44, 5 (2020), 2567–2581.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. 2021. Comparison of full-reference image quality models for optimization of image processing systems. *International Journal of Computer Vision* 129, 4 (2021), 1258–1281.

Simon S. Du, Xiyu Zhai, Barnabas Poczos, and Aarti Singh. 2019. Gradient Descent Provably Optimizes Over-parameterized Neural Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=S1eK3i09YQ

Jeremy Freeman and Eero P Simoncelli. 2011. Metamers of the ventral stream. *Nature neuroscience* 14, 9 (2011), 1195–1201.

Andreas Georgiou, Joel Kollin, Charlie Hewitt, Praneeth Chakravarthula, and Brian Guenter. 2023. Visual perception of noise in a simulated holographic display—a user study. *Displays* 76 (2023), 102333.

Joseph W Goodman. 2007. *Speckle phenomena in optics: theory and applications*. Roberts and company Publishers.

International Commission on Illumination. 2019. ISO/CIE 11664-4:2019. https://www.iso.org/standard/74166.html

International Telecommunication Union. 2025. BT.2100 : Image parameter values for high dynamic range television for use in production and international programme exchange. https://www.itu.int/rec/R-REC-BT.2100

Toshiaki Iwai and Toshimitsu Asakura. 1996. Speckle reduction in coherent information processing. *Proc. IEEE* 84, 5 (1996), 765–781.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4015–4026.

Piotr Krynski. 2022. Blender 3.3 LTS - Scanlands. https://www.blender.org/download/demo/splash/blender-3.3-splash.blend

Valero Laparra, Alexander Berardino, Johannes Ballé, and Eero P Simoncelli. 2017. Perceptually optimized image rendering. *Journal of the Optical Society of America A* 34, 9 (2017), 1511–1525.

Byounghyo Lee, Dongyeon Kim, Seungjae Lee, Chun Chen, and Byoungho Lee. 2022. High-contrast, speckle-free, true 3D holography via binary CGH optimization. *Scientific reports* 12, 1 (2022), 2811.

Changjun Li, Zhiqiang Li, Zhifeng Wang, Yang Xu, Ming Ronnier Luo, Guihua Cui, Manuel Melgosa, Michael H Brill, and Michael Pointer. 2017. Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. *Color Research & Application* 42, 6 (2017), 703–718.

Ninghe Liu, Kexuan Liu, Yixin Yang, Yifan Peng, and Liangcai Cao. 2025. Propagation-adaptive 4K computer-generated holography using physics-constrained spatial and Fourier neural operator. *Nature communications* 16, 1 (2025), 7761.

Shujian Liu and Yasuhiro Takaki. 2020. Optimization of phase-only computer-generated holograms based on the gradient descent method. *Applied Sciences* 10, 12 (2020), 4283.

Dougal Maclaurin, David Duvenaud, and Ryan Adams. 2015. Gradient-based hyperparameter optimization through reversible learning. In *International conference on machine learning*. PMLR, 2113–2122.

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), 1–16.

Rafał K Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–19.

Rafal K. Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVDP: A visual difference predictor for image, video and display distortions. *ACM Trans. Graph.* 43, 4, Article 129 (July 2024), 20 pages. doi:10.1145/3658144

Eric Markley, Nathan Matsuda, Florian Schiffers, Oliver Cossairt, and Grace Kuo. 2023. Simultaneous Color Computer Generated Holography. In *SIGGRAPH Asia 2023 Conference Papers* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 22, 11 pages. doi:10.1145/3610548.3618250

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

Aliaksei Mikhailiuk, Clifford Wilmot, Maria Perez-Ortiz, Dingcheng Yue, and Rafal Mantiuk. 2021. Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization. In *2020 IEEE International Conference on Pattern Recognition (ICPR)*.

Nicole Morena. 2023. Blender 3.5 - Cozy Kitchen. https://www.blender.org/download/demo/splash/blender-3.5-splash.blend

Yutaka Mori, Takahiko Fukuoka, and Takanori Nomura. 2014. Speckle reduction in holographic projection by random pixel separation with time multiplexing. *Appl. Opt.* 53, 35 (Dec 2014), 8182–8188. doi:10.1364/AO.53.008182

Aamir Mustafa, Aliaksei Mikhailiuk, Dan Andrei Iliescu, Varun Babbar, and Rafał K Mantiuk. 2022. Training a task-specific image reconstruction loss. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2319–2328.

Masahiro Nomura and Masashi Shibata. 2024. cmaes : A Simple yet Practical Python Library for CMA-ES. *arXiv preprint arXiv:2402.01373* (2024).

Björn Ottosson. 2020. A perceptual color space for image processing. https://bottosson.github.io/posts/oklab/

Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. 2019. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

Eli Peli. 1990. Contrast in complex images. *Journal of the Optical Society of America A* 7, 10 (1990), 2032–2040.

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.

Maria Perez-Ortiz and Rafal K Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686* (2017).

Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal processing: Image communication* 30 (2015), 57–77.

M S Ram. 2018. Eurasian hoopoe (Upupa epops) dust-bathing. https://www.youtube.com/watch?v=xm93eB8aXHE

J. G. Robson. 1966. Spatial and Temporal Contrast-Sensitivity Functions of the Visual System. *J. Opt. Soc. Am.* 56, 8 (Aug 1966), 1141–1142. doi:10.1364/JOSA.56.001141

Florian Schiffers, Praneeth Chakravarthula, Nathan Matsuda, Grace Kuo, Ethan Tseng, Douglas Lanman, Felix Heide, and Oliver Cossairt. 2023. Stochastic light field holography. In *2023 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–12.

Jinze Sha, Antoni Wojcik, Benjamin Wetherfield, Jianghan Yu, and Timothy D Wilkinson. 2024. Multi frame holograms batched optimization for binary phase spatial light modulators. *Scientific Reports* 14, 1 (2024), 19380.

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

Gaku Tada. 2023. Blender 4.0 - Gaku. https://www.blender.org/download/demo/splash/blender-4.0-splash.blend

Yasuhiro Takaki and Masahito Yokouchi. 2011. Speckle-free and grayscale hologram reconstruction using time-multiplexing technique. *Optics express* 19, 8 (2011), 7567–7579.

Taimoor Tariq, Nathan Matsuda, Eric Penner, Jerry Jia, Douglas Lanman, Ajit Ninan, and Alexandre Chapiro. 2023. Perceptually adaptive real-time tone mapping. In *SIGGRAPH Asia 2023 Conference Papers*. 1–10.

Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.

David R Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan Akşit, Anthony Steed, and Tobias Ritschel. 2021. Beyond blur: Real-time ventral metamers for foveated rendering. *ACM Transactions on Graphics* 40, 4 (2021), 1–14.

David R Walton, Koray Kavaklı, Rafael Kuffner Dos Anjos, David Swapp, Tim Weyrich, Hakan Urey, Anthony Steed, Tobias Ritschel, and Kaan Akşit. 2022. Metameric varifocal holograms. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 746–755.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612. doi:10.1109/TIP.2003.819861

Z. Wang, E.P. Simoncelli, and A.C. Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Vol. 2. 1398–1402 Vol.2. doi:10.1109/ACSSC.2003.1292216

Fan Yang, Andrew Kadis, Ralf Mouthaan, Benjamin Wetherfield, Andrzej Kaczorowski, and Timothy D Wilkinson. 2022. Perceptually motivated loss functions for computer generated holographic displays. *Scientific reports* 12, 1 (2022), 7709.

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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 586–595. doi:10.1109/CVPR.2018.00068