



OPEN

Perceptually motivated loss functions for computer generated holographic displays

Fan Yang^{1,2}, Andrew Kadis¹, Ralf Mouthaan¹, Benjamin Wetherfield¹, Andrzej Kaczorowski² & Timothy D. Wilkinson¹✉

Understanding and improving the perceived quality of reconstructed images is key to developing computer-generated holography algorithms for high-fidelity holographic displays. However, current algorithms are typically optimized using mean squared error, which is widely criticized for its poor correlation with perceptual quality. In our work, we present a comprehensive analysis of employing contemporary image quality metrics (IQM) as loss functions in the hologram optimization process. Extensive objective and subjective assessment of experimentally reconstructed images reveal the relative performance of IQM losses for hologram optimization. Our results reveal that the perceived image quality improves considerably when the appropriate IQM loss function is used, highlighting the value of developing perceptually-motivated loss functions for hologram optimization.

Holography offers a unique ability to control light, which profoundly impacts various applications from optical telecommunications¹, data storage², microscopy³ to two- and three-dimensional displays^{4,5}. Advances in algorithms and computational capacity have enabled Computer-Generated Holograms (CGHs) to be numerically calculated by simulating light diffraction and interference. The obtained CGH is displayed on a spatial light modulator (SLM), which modulates coherent illumination to reproduce the desired scenes. The goal of CGH algorithms is to compute a hologram that can be displayed on an SLM and that produces an intensity distribution that best approximates the desired image.

CGHs are commonly displayed on nematic liquid crystal SLMs, which boost superior light efficiency but are restricted to modulating only the phase of the incident beam. To solve the phase-only restriction imposed by these SLMs, double phase^{4,6} and error diffusion methods^{7–9} directly encode complex-amplitude diffraction fields into phase-only holograms. Another approach, known as the one-step phase retrieval algorithm (OSPR)^{10,11}, displays multiple phase-only holograms within a short time interval to statistically average out errors in the replay field. Trained deep learning-based CGH algorithms are also employed as non-iterative solutions^{12–14}. Iterative CGH algorithms such as direct search (DS)¹⁵ and simulated annealing (SA)¹⁶ alter single pixels in the hologram to find the optimal hologram. Phase retrieval methods like the Gerchberg-Saxton algorithm (GS)¹⁷ and hybrid input–output (HIO)^{18,19} method have also been explored.

Recently, the gradient descent method has been applied to phase-only CGH optimization^{12–14,20–24}. The gradient of a predefined objective function is calculated and used to update the hologram phase at each iteration. This method can be further combined with a camera as a feedback optimization strategy to eliminate optical artifacts in experimental setups^{13,22}. The specific loss function selected is essential in these iterative optimization approaches to drive the hologram phase to its optimal state. A standard choice of the loss function is the mean squared error (MSE) due to its simplicity of use and clear physical meaning. Though MSE quantifies the per-pixel error in the reconstructed image, it is widely criticized for its poor correlation with perceptual quality^{25–28}.

A promising but relatively less exploited approach is to use image quality metrics (IQMs) in the phase-only CGH optimization process. The traditional role of IQMs in digital holography is to dynamically monitor the optimization process and to evaluate the perceptual quality of obtained images^{29–32}. Modern IQMs model assesses visual quality based on a priori knowledge regarding the human visual system or uses learned models trained with large datasets. They use image features in appropriate perceptual spaces^{28,33} for image quality evaluation but have not yet been fully exploited in the CGH optimization process. Here, we focus on the use of IQMs as an alternative to the ubiquitous MSE for the training loss, with the intention of using the gradient of these perceptual metrics to strive for a better CGH optimization algorithm. The use of perceptual motivated loss functions has recently gained attention in foveated CGH^{34,35}, focusing specifically on speckle suppression in the foveal

¹Centre of Molecular Materials, Photonics and Electronics, University of Cambridge, Cambridge, UK. ²Research Division, VividQ Ltd., Cambridge, UK. ✉email: tdw13@cam.ac.uk

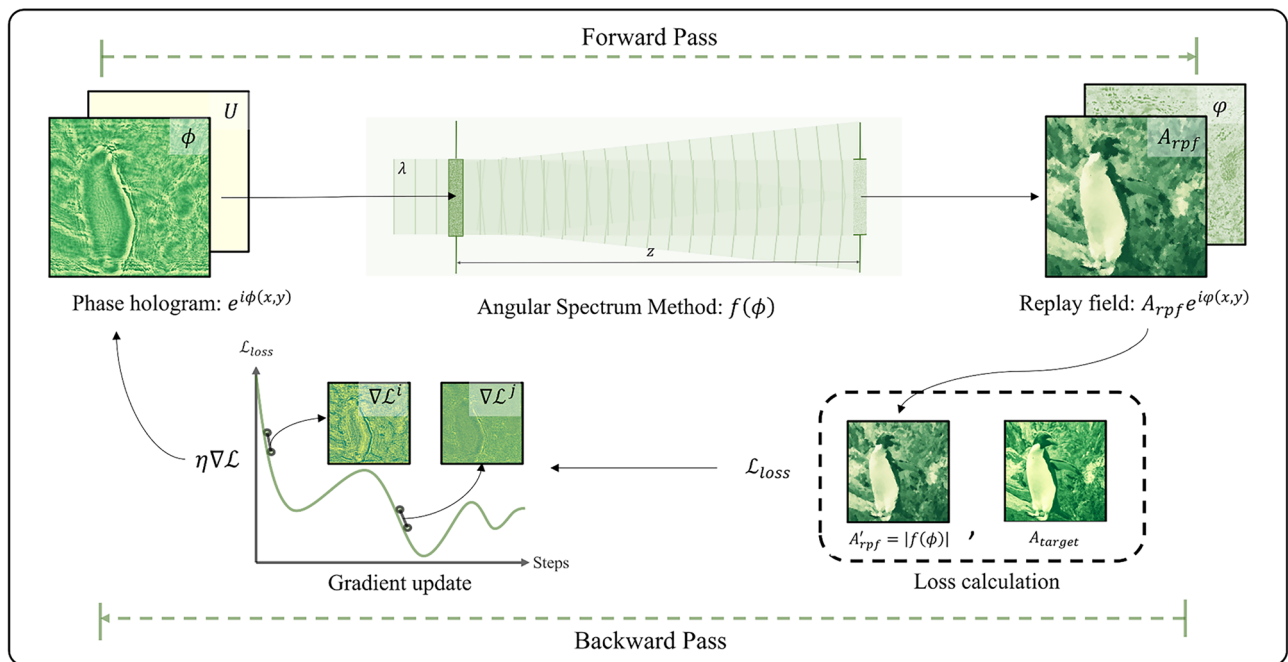


Figure 1. CGH optimization model based on the gradient descent method.

region and peripheral perception. Other non-holographic image restoration applications have also explored perceptual losses, though it is observed that there is no single loss function that outperforms all others across different applications^{36–38}.

In this paper, we present a comprehensive comparison of different IQMs as losses for CGH optimization using gradient descent. Specifically, we first choose ten optimization-suitable IQMs together with mean absolute error (MAE) and MSE to generate CGHs. These IQMs have not been applied to the hologram design, and are selected among the plethora of existing metrics due to their well establishment as well as their differentiability, a requirement for use in the gradient descent method. We build a holographic display prototype to acquire an optical reconstruction dataset of IQM optimization phase holograms. We use this dataset to perform an in-depth analysis of the relative performance of IQM losses based on extensive objective quality assessments as well as subjective comparisons informed by human perceptual judgments. Finally, we present a rigorous procedure for evaluating the perceptual quality of holographic images and highlight the value of developing perceptually-motivated loss functions for hologram optimization.

Background

CGH optimization model using the gradient descent method. CGH generation based on the gradient descent method can be generalized as an optimization model. In the forward pass, the model propagates a phase hologram to the replay plane to produce a reconstructed image, which is used to calculate the loss by comparing it to the target image. In the backward pass, the model traverses backward from the output, collecting the derivatives of the loss function with respect to the phase hologram and updating the hologram to minimize the loss. The model iteratively goes through the forward pass and the backward pass to obtain the optimized phase hologram. This process is illustrated in Fig. 1.

In the forward pass, we consider the angular spectrum method^{39,40} with planar illuminating wave for modeling the diffraction propagation function:

$$f(\phi) = \mathcal{F}^{-1} \left\{ \mathcal{F} \left\{ e^{i\phi(x,y)} \right\} \times \exp \left[j2\pi z \sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2} \right] \right\}. \quad (1)$$

Here, $\phi(x, y)$ is the phase hologram that has been quantized so that it can be displayed on a binary or 8-bit SLM, λ is the wavelength, f_x, f_y are spatial frequencies, and z is the propagating distance between the hologram plane and the replay field plane. \mathcal{F} and \mathcal{F}^{-1} denote the Fourier transform and the inverse Fourier transform, respectively. The resulting field $f(\phi)$ is a complex replay field, whose amplitude is related to the reconstructed image intensity by $I(\mu, \nu) = |f(\phi)|^2$. To evaluate the perceived image quality, the amplitude of the replay field A_{rpf} is compared with the target amplitude A_{target} using a loss function \mathcal{L} . Though intensity-based objective functions can also be utilized for image quality evaluation, amplitude-based objective functions have been found to yield better algorithmic performance and are preferable in hologram optimization^{41,42}. Therefore, the CGH optimization algorithm aims to find the optimal quantized phase hologram $\hat{\phi}$ that minimizes the loss function \mathcal{L} describing the visual quality, calculated from the reconstructed image amplitude $|f(\phi)|$ and the intended target image amplitude A_{target} :

IQM losses	Library	Underlying principle
MAE	Pytorch	Pixel-based absolute error with average pooling
MSE	Pytorch	Pixel-based squared error with average pooling
NLPD	IQA	Root MSE in the weighted Laplacian pyramid decomposition domain
SSIM	IQA	A weighted combination of measures: luminance, contrast and structure
MS-SSIM	IQA	The multi-scale representation of the SSIM
FSIM	PIQ	A weighted combination of the phase congruency feature and the gradient magnitude feature
MS-GMSD	PIQ	The multi-scale representation of GMSD, measuring standard deviation based on pixel-wise gradient magnitude similarity map
VSI	PIQ	Similarities in the gradient magnitude and the visual saliency
HaarPSI	PIQ	local similarities and the relative importance of image areas based on Haar wavelet
VIF	PIQ	Model the image source using Gaussian scale mixtures on wavelet coefficients and quantify mutual information
LPIPS	IQA	Evaluate the Euclidean distance between image patches based on feature maps
DISTS	IQA	Combination of SSIM-like structure and texture similarity measurements based on the VGG network

Table 1. The utilized underlying principle of IQM losses for CGH optimization.

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \mathcal{L}(s \cdot |f(\phi)|, A_{\text{target}}), \quad (2)$$

where s is a scaling factor for normalization. The mean square error (MSE) for a m by n sampling points is commonly used as the loss function, computed by averaging the squared amplitude differences of reconstructed and target image pixels:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{mn} \sum_{m,n} [|f(\phi)| - A_{\text{target}}]^2. \quad (3)$$

In the backward pass, the model calculates the gradient $\partial \mathcal{L} / \partial \phi^{k-1}$ of the loss function with respect to the current estimate of the phase hologram ϕ^{k-1} to update the next estimate phase ϕ^k . The gradient can be calculated by the chain rule, which involves the calculation of complex derivatives:

$$\frac{\partial \mathcal{L}}{\partial \phi^{k-1}} = \frac{\partial \mathcal{L}}{\partial A'_{\text{rpf}}} \cdot \frac{\partial A'_{\text{rpf}}}{\partial f} \cdot \frac{\partial f}{\partial \phi^{k-1}}, f: \mathbb{C} \rightarrow \mathbb{C}. \quad (4)$$

In complex analysis, the holomorphic requirement for functions to be complex differentiable is very strict. Wirtinger calculus relaxes this requirement and allows approximate complex derivatives of nonholomorphic functions to be more easily calculated by using a conjugate coordinate system^{21,43,44}. Recently, Wirtinger calculus has been implemented in automatic differentiation packages in machine learning libraries such as TensorFlow and PyTorch. These automatic differentiation packages keep a record of all the data and operations that have been done in the forward pass in a direct acyclic graph and automatically compute gradients using the chain rule. For a learning rate η , the next estimate phase hologram $\phi^{(k)}$ is given by:

$$\phi^{(k)} = \phi^{(k-1)} - \eta \nabla \mathcal{L}(s \cdot |f(\phi^{(k-1)})|, A_{\text{target}}). \quad (5)$$

Several update strategies, such as Adagrad⁴⁵ and Adam⁴⁶, propose learning rate update rules to improve accuracy and convergence speed.

IQM as loss functions. IQMs play a vital role in the development and optimization of image processing and restoration algorithms. Generally, IQMs can be classified into full-reference methods, reduced-reference methods, and no-reference methods according to the availability of the original reference image. Since the target image is available in the CGH optimization model, we only consider full-reference methods as loss functions. IQMs are a function of a number of parameters, and different IQM implementations can yield significantly different results, impacting the performance of CGH optimization. We therefore consider ten differentiable full-reference IQMs from existing libraries IQA³⁷ and PIQ⁴⁷, benchmarked on common databases, which we believe include a wide range of state-of-art full-reference IQMs. We also include MAE and MSE as standards for comparison. Therefore, this IQM collection includes three error visibility methods: MSE, MAE and NLPD³³, six structural similarity methods: SSIM²⁶, MS-SSIM⁴⁸, FSIM⁴⁹, MS-GMSD⁵⁰, VSI⁵¹, HaarPSI⁵², one information-theoretical method: VIF⁵³, and two learning-based methods: LPIPS²⁵ and DISTS⁵⁴. Error visibility methods calculate the image error on a pixel-by-pixel basis. Structural similarity methods consider the perceived variation, including luminance, contrast, and structure, to assess image distortion. Information-theoretic methods quantify the amount of information loss in the distorted images with respect to the target images. Learning-based methods propose neural networks trained with numerous pictures to assess image quality. Table 1 summarizes the library of the IQMs considered as well as the underlying principle. The IQM is reformulated where neces-



Figure 2. Monochrome target image amplitudes from the DIV2K dataset.

sary so that a lower score indicates higher predicted quality. For example, if the selected IQM is *SSIM*, then \mathcal{L} is rewritten as $\mathcal{L}_{SSIM} = 1 - SSIM$.

Methods

Hologram generation. We generate CGHs for 100 high-resolution images in the DIV2K dataset^{55,56} pre-processed to give a monochrome target amplitude shown in Fig. 2. This is done for each IQM, and we therefore generate a dataset with a total of 1200 holograms. In each case we forward propagate, compare to the target, and then backward propagate to obtain the gradient for the IQM loss, which is used by the Adam optimizer to iteratively find the optimal phase hologram. In all cases we use the Adam optimizer with a 0.05 stepsize and default exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The total number of iterations is empirically set to 1000 with the initial 15 iterations using MSE as the loss function. We apply this basic preprocessing step since initial predictions can have a significant impact on the performance of some IQMs. This step is necessary to yield acceptable optimization results and reduce the training time for learning-based IQMs. During each iteration, we normalize the amplitude of the replay field since several IQMs require input data within the range $[0, 1]$.

The CGH generation is done on a machine with an Intel i7-8700 CPU @ 3.20 GHz and a GeForce GTX 1080 GPU. PyTorch 1.9.0 and CUDA 10.2 are used to implement complex-amplitude gradient descent optimization on the GPU. Computation takes roughly 190 GPU hours to generate the 1200 holograms to assess all 12 IQMs. Training details and computational time for each IQM loss are included in the supplementary material.

Optical reconstruction setup. In order to verify our image quality from simulation, we develop a physical optical display system. We display the holograms on a SLM and optically reconstruct the replay fields captured using a camera. The proposed holographic projection system is shown in Fig. 3. Our system uses an 8-bit phase-only SLM (FSLM-2K55-P) with a pixel pitch of $6.4 \mu\text{m}$ and a resolution of 1920×1080 . The SLM is made by the Xi'an Institute of Optics and Precision Mechanics company and is factory pre-calibrated in reflection mode. The first arm consists of a 532 nm laser source (Thorlabs CPS532), a half waveplate, a 4F lens system, and a polarizer. The 4F lens system comprises two lenses (lens 1 and lens 2) with focal lengths of 13 mm and 75 mm respectively, used to expand the beam. The expanded beam is then linearly polarized and illuminates the SLM. The second arm comprises a beam splitter and a 4F lens system with a spatial filter to reduce the DC component of the replay field and other unwanted higher diffraction orders. The focal lengths of these two lenses (lens 3 and lens 4) are 30 mm and 50 mm. The second arm is adjusted to project the reconstructed images onto the camera sensor. A neutral density filter can be inserted in the second arm to reduce the replay field intensity.

Reconstructed images are captured using a Canon EOS 6D camera without a camera lens attached. The camera output resolution is 5472×3648 with a gain setting of ISO 125 to minimize amplifier noise. For a fair comparison, we perform a camera calibration using a reconstructed circle grid pattern hologram and adjust the mean of captured image amplitude values to match the target image amplitude values. The target images are cropped to 1680×960 pixels to match the experimentally captured images. All reconstructed images are averaged out across three captured images which are captured in sRGB, the camera's native color space. We further applied the image linearization process that converts the captured image from sRGB intensity into monochromatic linear space amplitude^{13,57}.

Subjective testing. To subjectively differentiate quality variations of tested models, we gather human perceptual judgments by employing a 2-alternative forced choice (2AFC) method. The experiment asks subjects to indicate which one of two distorted images is perceptually closer to the reference image. Figure 4 illustrates the interface for this experiment: an image triplet with a pair of experimentally captured images and the corresponding reference image are simultaneously presented. Subjects are asked to select the better image between two distorted ones. After the selection, two new experimentally captured images, optimized according to different IQM losses, appear on the upper screen in randomized left-right order. Progress is indicated and a pause function

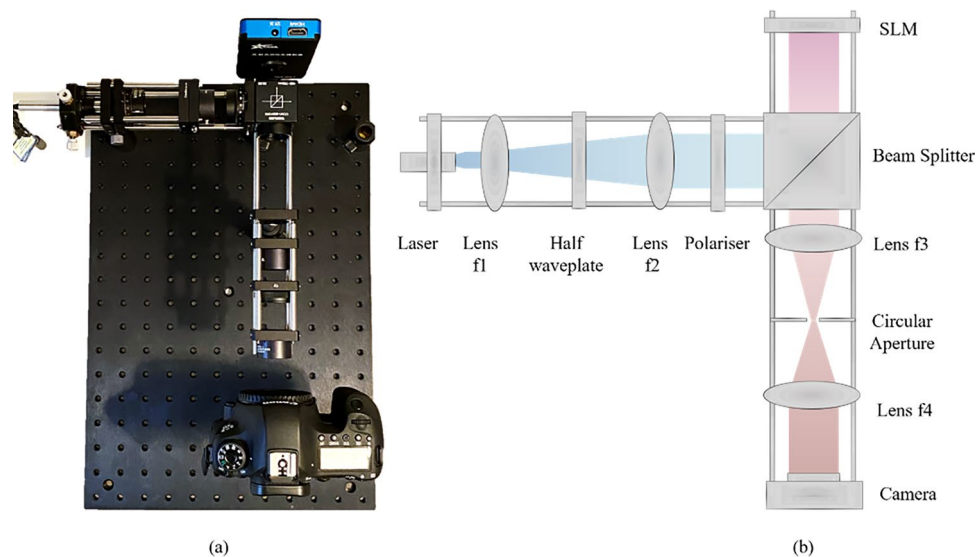


Figure 3. Holographic display system setup. (a) Our hardware display prototype with a Canon camera for image acquisition. (b) Optical system schematic diagram.

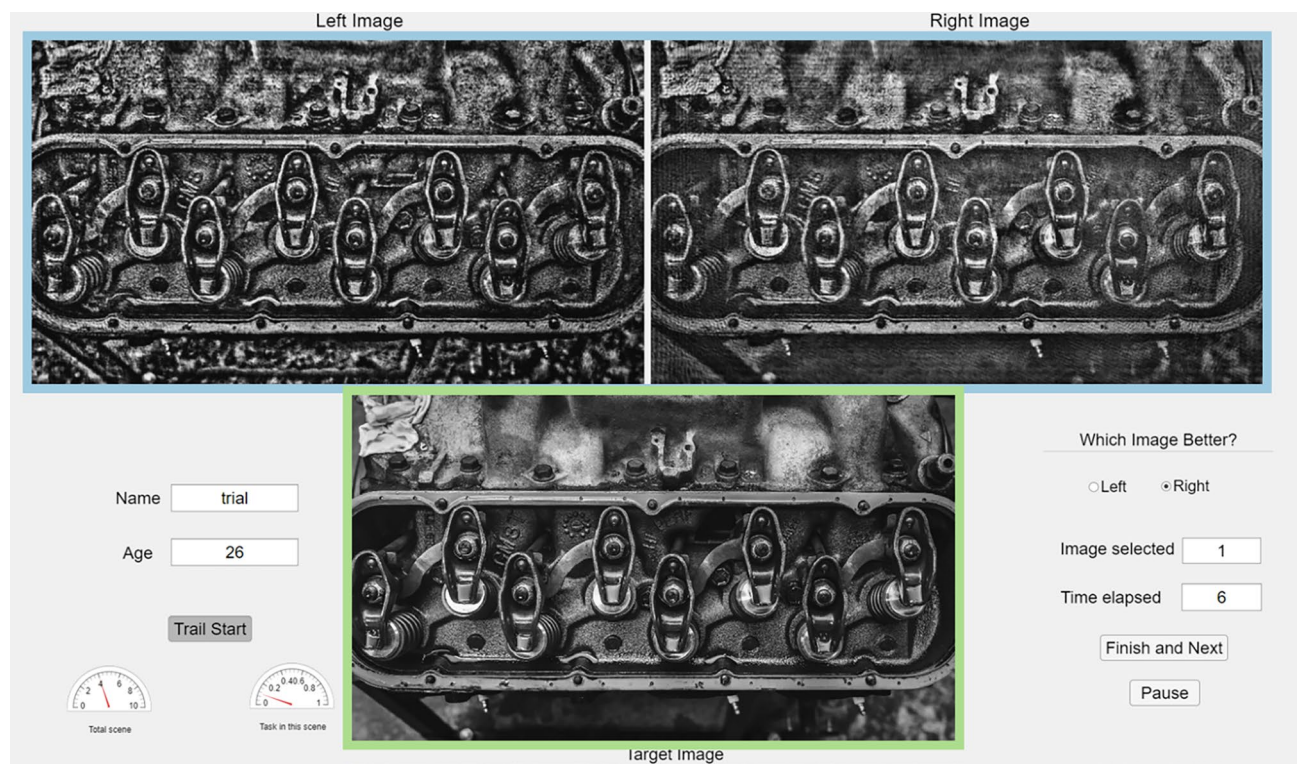


Figure 4. The user interface for collecting human judgments on IQM-based CGH optimization. The experimentally captured image pair from two IQM losses and the corresponding reference image are shown in the blue and the green box respectively.

is available to reduce visual fatigue. The screen has a resolution of 1920×1080 pixels, with the displayed image resolution at 875×500 . The user interface supports a zoom function for careful inspection of image details.

Participants are mainly university students and are provided with appropriate instructions, including an explanation of the experimental procedure as well as a demonstration session. To avoid fatigue, we pause the user interface every 15 min and allow subjects to take a break at any time during the experiment. Experiments are performed at a normal indoor light level with reasonable varying ambient conditions according to the recommendations of ITU-R BT 500³⁸. This subjective experiment was approved by the Cambridge Engineering Research Ethics committee and carried out according to the Declaration of Helsinki. We obtained informed consent and gathered paired comparisons from 20 subjects. Each subject responded to all possible combinations of generated

images for a pair of target images, doing so for ten pairs of target images, yielding $\binom{12}{2} \times 10 = 660$ stimuli. Data including time spent for each judgment, the paired-image display order and results of pairwise comparisons, is saved for analysis. The preferred image of the displayed pair contributes one point to the score of its IQM loss. Therefore, for the selected 10 sample images, each paired comparison could receive 0 to 10 points as the subjective score from the subject. In order to exclude abnormal results, we check several sentinels in each observation data that consist of pairs with obvious visual quality contrast. Overall, we received 13,200 judgments across 12 IQM losses, and each loss is ranked 1100 times. The average time for one judgment is approximately 3 s.

We employ the Bradley-Terry model^{59,60} to aggregate pairwise comparisons and obtain a global ranking of IQM losses for CGH optimization based on the subjective data. From partial orderings provided in the data, we wish to infer not only the ranking order of tested losses but also the subjective visual quality scores associated with the losses themselves. If we denote $s = [s_1, s_2, s_3, \dots, s_m]$ as subjective scores of the evaluated IQM losses, the Bradley-Terry model assumes that the probability of choosing loss i over loss j is:

$$p_{ij} = \frac{e^{s_i}}{e^{s_i} + e^{s_j}}. \quad (6)$$

Given the observed number of times that IQM loss i is favored over IQM loss j as w_{ij} , We then can obtain the likelihood of i over j as $p_{ij}^{w_{ij}}$. Thus, assuming outcomes of each paired comparison are statistically independent, the likelihood function of all (i, j) pairs is defined by:

$$P = \prod_{i=1}^M \prod_{\substack{j=1 \\ j \neq i}}^M p_{ij}^{w_{ij}}. \quad (7)$$

The subjective score for IQM loss s_i can then be jointly estimated by maximizing the log-likelihood of all pairwise comparison observations:

$$\ell(s_i) = \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M (w_{ij}s_i - w_{ij} \log(e^{s_i} + e^{s_j})). \quad (8)$$

Results and discussion

Hologram generation results. The simulated reconstruction results based on IQM optimization models are shown in Fig. 5. Corresponding phase holograms, as well as the experimental captured results in sRGB space, are shown in the second and third rows respectively.

Qualitative interpretation. We first make a qualitative comparison across all IQM-optimized methods for experimental results. As shown in Fig. 6, most IQM-based optimization models converge on a reasonable visual quality. We observe that MAE, MSE, NLPD, SSIM, and MS-SSIM perform well but have undesirable local noise, which can be observed in the image patches selected from the reconstructed images. FSIM and VIF amplify high-frequency information, leading to structural over-enhancement. VSI, MS-GMSD and HaarPSI preserve the overall structures with a smooth appearance, but artificially reduce local contrast with noticeable artifacts. Models based on deep-learning methods such as LPIPS and DISTs can recover the target image details but superimpose textures on the image.

The optically reconstructed images exhibit laser speckle noise and are subject to optical aberrations, resulting in some noticeable common artifacts across all IQMs, including ghost and ripple effects. The dynamic range of the camera is limited and captured images are prone to photometric distortions, including reduced contrast and saturation.

Objective interpretation. We use the proposed IQMs as quality measures to evaluate the performance of gradient descent based CGH optimization using different IQM losses. All IQMs are used to objectively evaluate the captured results. Scores are averaged over all 100 images for each metric and for each IQM-based loss shown in Table 2. Each element indicates the score of an IQM loss evaluated using another IQM as a quality predictor.

By inspecting each row of the metric table, we find MAE, NLPD, SSIM, and MS-SSIM maintain the best performance among all IQM losses as previously predicted by the qualitative comparison. MS-SSIM loss produces superior reconstruction quality and objectively ranks as the best performing IQM-based CGH optimization model on most evaluation metrics, while FSIM ranks as the least preferred method. Several other IQM losses, including NLPD, MAE, SSIM, HaarPSI and MS-GMSD, also outperform the MSE loss, which objectively validates the use of IQMs for CGH optimization.

Since the PIQ library implements its own SSIM and MS-SSIM metrics for image quality assessment, we can further evaluate our top-performing models by using these metrics, as shown in Table 3. Though both the IQA and PIQ libraries have been benchmarked on a set of common databases and have nearly consistent ranking results in model evaluation, there is disagreement with the actual values of performance evaluation, with the IQM library generally obtaining lower scores. Hence, in the absence of a standard IQM implementation, it becomes more challenging to compare the performance of different algorithms.

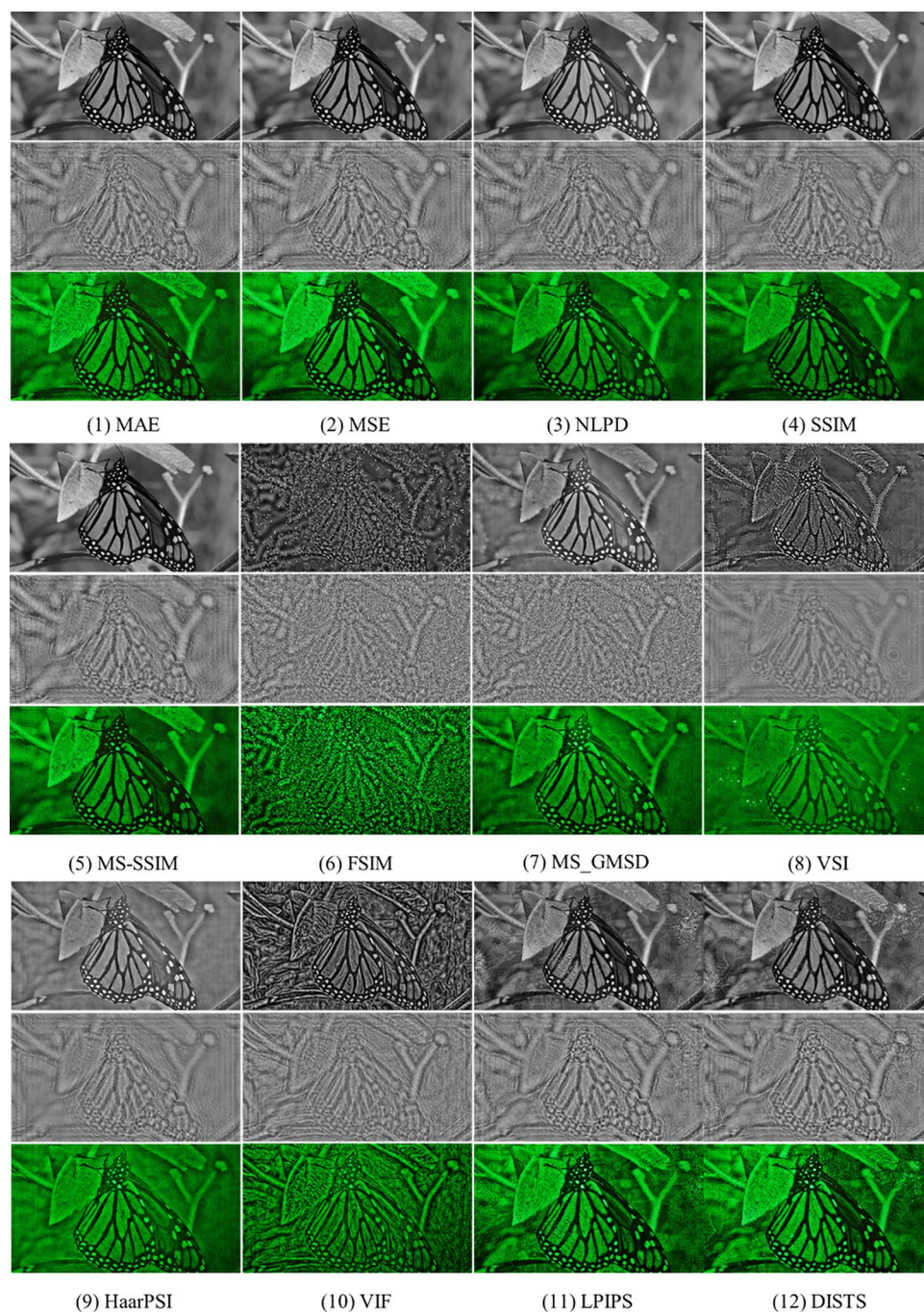


Figure 5. Simulated and captured results for CGH optimization using twelve different IQM losses. We show the reconstructed image at the top for each loss with the phase hologram in the middle and its corresponding captured results at the bottom.

Subjective interpretation. We implement the Bradley-Terry model in R to iteratively solve the given equation Eq. (8) and obtain the optimal estimate s_i for each model. The Bradley-Terry model scores are normalized by shifting to zero means, resulting in a global ranking of perceptual optimization performance. We further conduct independent two-sample two-tailed t-tests to investigate whether the differences between the subjective performance of IQM losses are statistically significant. Specifically, we consider that the obtained observations from participants are normally distributed under the null hypothesis and compare the ranking scores for any of the two losses. If the comparison cannot reject the null hypothesis of no difference at the standard significance level $\alpha = 0.05$, we put the evaluated losses in the same group as they are statistically indistinguishable. Figure 7 shows the scatter plot of the combined subjective and objective performance of tested IQM losses for CGH optimization. Scatter points with the same color are in the same statistical significance group for subjective tests. The objective global ranking score for each IQM loss can be obtained by adding ranking orders from all quality

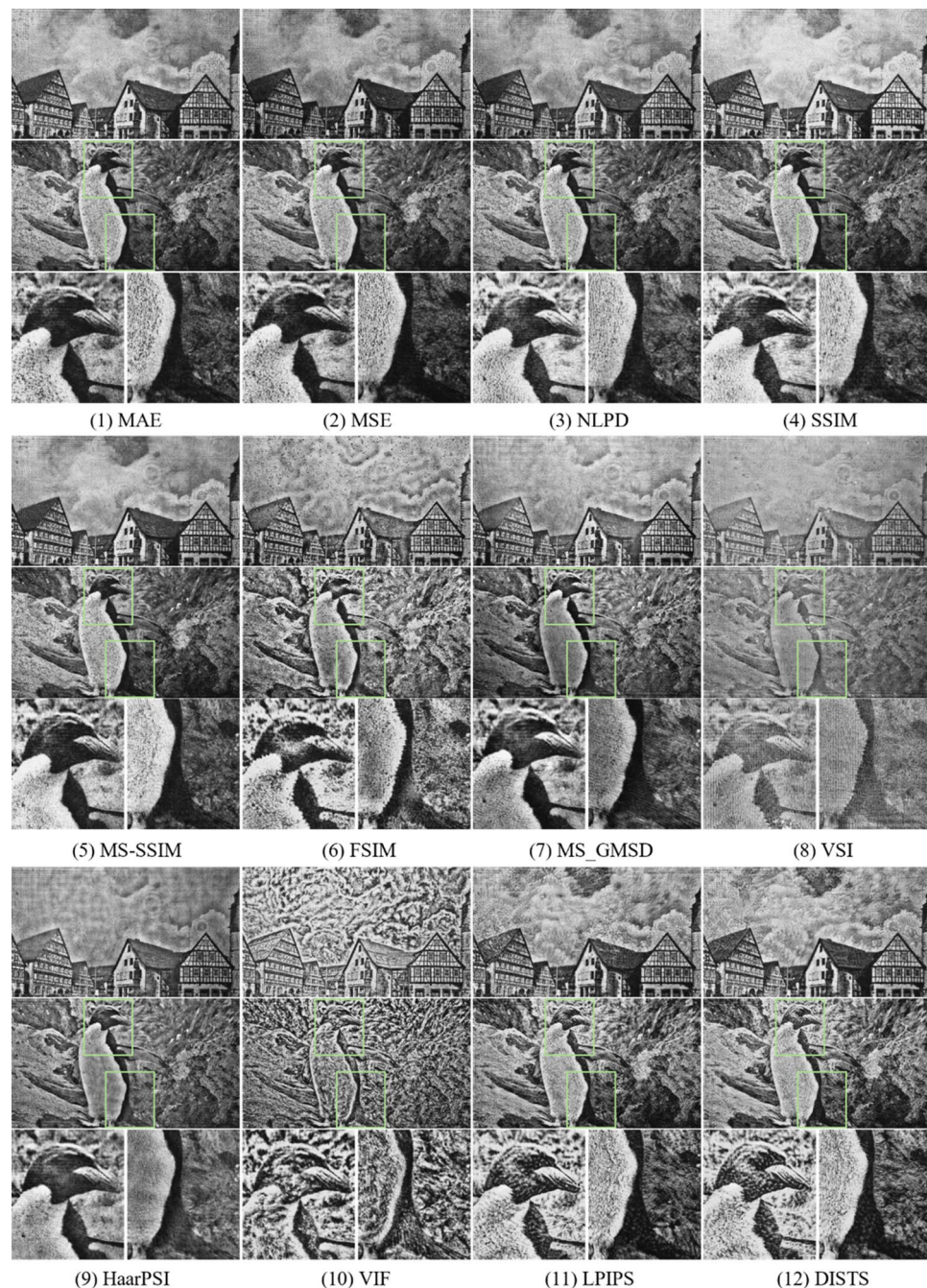


Figure 6. Captured reconstruction results. For target images, we display phase holograms optimized by IQM losses. Reconstructions results of IQM losses are captured with our holographic display prototype for image quality comparison.

metrics derived from Table 2 and normalizing them to zero mean. Scores have been reformulated to ensure a higher score indicates a higher predicted quality.

The scatter plot indicates that the MS-SSIM is the top-ranking loss function, as agreed upon by both subjective and objective evaluations. NLPD and SSIM losses are statistically indistinguishable from the MSE loss for subjective performance. The MSE loss unexpectedly achieves higher performance in the subjective test than HaarPSI, and MAE losses, despite performing far worse in objective performance. A similar trend also occurs in VSI and VIF losses versus FSIM loss. This disagreement is due to different objective and subjective weighting strategies on image structure similarity, image smoothness, luminance, and contrast.

We further calculate Spearman's rank-order correlation coefficient (SRCC) between objective and subjective scores, as shown in Table 4. Higher SRCC scores indicate a better correlation of a metric with subjective ratings. Although most modern image quality metrics show superior performance in existing image databases, we

IQM losses	Objective image quality metrics											
	MAE	MSE	NLPD	SSIM	MS-SSIM	FSIM	MS-GMSD	VSI	HaarPSI	VIF	LPIPS	DISTS
MAE	0.104	0.021	0.754	0.382	0.568	0.770	0.270	0.893	0.264	0.118	0.600	0.265
MSE	0.120	0.028	0.862	0.315	0.458	0.720	0.276	0.877	0.234	0.078	0.618	0.273
NLPD	0.118	0.024	0.717	0.365	0.566	0.783	0.258	0.905	0.287	0.117	0.601	0.271
SSIM	0.107	0.021	0.739	0.371	0.563	0.779	0.262	0.904	0.279	0.112	0.604	0.272
MS-SSIM	0.096	0.018	0.696	0.414	0.610	0.795	0.256	0.913	0.296	0.133	0.589	0.253
FSIM	0.185	0.058	1.083	0.219	0.305	0.648	0.294	0.795	0.187	0.067	0.664	0.387
MS-GMSD	0.153	0.040	0.833	0.328	0.451	0.744	0.258	0.879	0.274	0.098	0.608	0.283
VSI	0.158	0.040	0.816	0.299	0.430	0.761	0.256	0.894	0.276	0.079	0.628	0.406
HaarPSI	0.145	0.035	0.748	0.380	0.526	0.783	0.245	0.901	0.313	0.121	0.589	0.272
VIF	0.171	0.051	0.895	0.338	0.413	0.633	0.294	0.790	0.200	0.197	0.580	0.314
LPIPS	0.127	0.029	0.896	0.288	0.430	0.696	0.289	0.852	0.216	0.084	0.635	0.247
DISTS	0.130	0.030	0.911	0.279	0.415	0.690	0.289	0.852	0.212	0.077	0.636	0.246

Table 2. Objective performance of IQM-based model evaluated by IQMs as quality metrics. Significant values are in bold.

IQM losses	Objective image quality metrics			
	SSIM	MS-SSIM	SSIM (piq)	MS-SSIM (piq)
MS-SSIM	0.414	0.610	0.619	0.641
NLPD	0.365	0.566	0.567	0.601
HaarPSI	0.380	0.526	0.550	0.591
MAE	0.382	0.568	0.577	0.602
SSIM	0.371	0.563	0.568	0.596
MS-GMSD	0.328	0.451	0.463	0.505
MSE	0.315	0.458	0.446	0.484

Table 3. Objective performance of IQM-based model evaluated on different libraries.

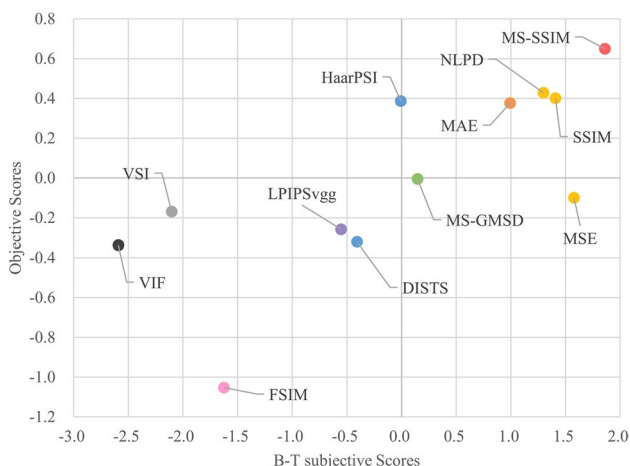


Figure 7. Quantitative comparison of IQM-based CGH optimization. Scatter points represent the losses for CGH optimization. Points with the same color are statistically indistinguishable for subjective results. Vertical and horizontal axes indicate the objective performance and the subjective performance of each loss respectively.

observe that for CGH they have less correlation than pixel-error-based metrics to human judgments. This may be because the most common image databases for benchmarking such as LIVE⁶¹, TID2008⁶² and TID2013⁶³ comprise source images with synthetically distorted images. The synthetic distortion types, including White Gaussian Noise, JPEG2000 compression, and Gaussian Blur with varied distortion levels, attempt to reflect various image impairments found in image processing. Experimental CGH reconstructed images, such as those seen here, can be rather more complex with more types of distortions produced during the optical reconstruction and image

Objective image quality metrics	SRCC
MAE	0.846
MSE	0.825
NLPD	0.657
SSIM	0.587
MS-SSIM	0.839
FSIM	0.692
MS-GMSD	0.434
VSI	0.678
HaarPSI	0.566
VIF	0.189
LPIPS	0.266
DISTS	0.427

Table 4. SRCC between objective scores and subjective scores of IQM-based CGH optimization.

acquisition. Furthermore, CGHs are predominantly tainted by noise, whereas some IQMs were developed for recognizing blurry objects, inferring details in deblurred objects, or super-resolution imaging tasks. Current IQMs are not well specifically benchmarked for those real-world and CGH distortions. For partial coherent light illumination in the holographic optical system that could bring more blurry effect and contrast reduction in the replay field^{57,64}, modern IQMs may take advantage in inferring blurry and contrast-reduced information. Therefore, the use of IQMs may potentially have better performance in partial coherent holographic displays.

Conclusion

In this work, we have conducted a comprehensive study of the real-world performance of using IQM as loss functions in the CGH optimization process. By benchmarking with a standard optical reconstruction dataset, we have collated the results of applying 12 distinct IQMs as loss functions in both objective and subjective ratings. The results from the comparison study show that IQM losses can achieve better image quality than the MSE loss in generating holograms, with the MS-SSIM loss outperforming all the other losses. This extensive comparison provides guidance for finding a specific perceptually-motivated loss function for CGH generation.

Beyond this study, individual IQM losses can be further combined based on their complementarity to incorporate the specific CGH distortions. We recognize that our analysis is limited to 2D hologram reconstruction. For 3D holographic applications, the authors believe that there are several extensions to the work conducted in this study, such as the use of blurring distortion, which could be a significant perceptual factor to be considered in hologram optimization.

Data availability

The datasets generated and/or analysed during the current study are available in the GitHub repository, https://github.com/fy255/perceptual_cgh.

Code availability

The code for hologram generation and evaluation are publicly available in the GitHub repository, https://github.com/fy255/perceptual_cgh. Additional codes are available from the corresponding authors upon reasonable request.

Received: 27 January 2022; Accepted: 14 April 2022

Published online: 11 May 2022

References

- Huang, C. *et al.* Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends. *IEEE Wirel. Commun.* **27**, 118 (2020).
- Li, L. *et al.* Electromagnetic reprogrammable coding-metasurface holograms. *Nat. Commun.* **8**, 1–7 (2017).
- Marquet, P. *et al.* Digital holographic microscopy: A noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Opt. Lett.* **30**(5), 468–470 (2005).
- Maimone, A., Georgiou, A. & Kollin, J. S. Holographic near-eye displays for virtual and augmented reality. *ACM Trans. Graph. (TOG)* **36**, 1–16 (2017).
- Yaraş, F., Kang, H. & Onural, L. State of the art in holographic displays: A survey. *IEEE/OSA J. Display Technol.* **6**, 443 (2010).
- Hsueh, C. K. & Sawchuk, A. A. Computer-generated double-phase holograms. *Appl. Opt.* **17**, 3874 (1978).
- Tsang, P. W. M., Cheung, W. K., Poon, T.-C. & Zhou, C. Novel method for converting digital Fresnel hologram to phase-only hologram based on bidirectional error diffusion. *Opt. Exp.* **21**(20), 23680–23686 (2013).
- Eschbach, R. Comparison of error diffusion methods for computer-generated holograms. *Appl. Opt.* **30**, 3702–3710 (1991).
- Pang, H. *et al.* Non-iterative phase-only Fourier hologram generation with high image quality. *Opt. Exp.* **25**, 14323–14333 (2017).
- Buckley, E. Invited paper: Holographic laser projection technology. *SID Sympos. Digest. Tech. Pap.* **39**, 1074–1079 (2008).
- Cable, A. J. *et al.* Real-time binary hologram generation for high-quality video projection applications. *SID Sympos. Digest Tech. Pap.* **35**, 1431–1433 (2004).
- Shi, L., Li, B., Kim, C., Kellnhofer, P. & Matusik, W. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).

13. Peng, Y., Choi, S., Padmanaban, N. & Wetzstein, G. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* **39**, 1–14 (2020).
14. Eybposh, M. H., Caira, N. W., Atisa, M., Chakravarthula, P. & Pégard, N. C. DeepCGH: 3D computer-generated holography using deep learning. *Opt. Exp.* **28**, 26636–26650 (2020).
15. Seldowitz, M. A., Allebach, J. P. & Sweeney, D. W. Synthesis of digital holograms by direct binary search. *Appl. Opt.* **26**, 2788–2798 (1987).
16. Kirk, A. G. & Hall, T. J. Design of binary computer generated holograms by simulated annealing: Coding density and reconstruction error. *Opt. Commun.* **94**, 491–496 (1992).
17. Gerchberg, R. W. & Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik (Stuttg)* **35**, 237–246 (1971).
18. Fienup, J. R. Phase retrieval algorithms: A comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
19. Bauschke, H. H., Combettes, P. L. & Luke, D. R. Hybrid projection–reflection method for phase retrieval. *JOSA A* **20**, 1025–1110 (2003).
20. Liu, S. & Takaki, Y. Optimization of phase-only computer-generated holograms based on the gradient descent method. *Appl. Sci.* **10**, 4283 (2020).
21. Chakravarthula, P., Peng, Y., Kollin, J., Fuchs, H. & Heide, F. Wirtinger holography for near-eye displays. *ACM Trans. Graph. (TOG)* **38**, 1–13 (2019).
22. Chakravarthula, P., Tseng, E., Srivastava, T., Fuchs, H. & Heide, F. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Trans. Graph. (TOG)* **39**, 1–18 (2020).
23. Chen, C. *et al.* Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function. *Opt. Exp.* **29**, 15089–15103 (2021).
24. Zhang, J., Pégard, N., Zhong, J., Adesnik, H. & Waller, L. 3D computer-generated holography by non-convex optimization. *Optica* **4**, 1306–1313 (2017).
25. Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 586–595. <https://doi.org/10.1109/CVPR.2018.00068> (2018).
26. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
27. Wang, Z. & Bovik, A. C. Mean squared error: Lot it or leave it? A new look at signal fidelity measures. *IEEE Signal Process. Mag.* **26**, 98–117 (2009).
28. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 9906. LNCS. 694–711. (2016).
29. Rivenson, Y., Zhang, Y., Günaydin, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141–17141 (2017).
30. Ahar, A. *et al.* Comprehensive performance analysis of objective quality metrics for digital holography. *Signal Process. Image Commun.* **97**, 116361 (2021).
31. Ahar, A. *et al.* Subjective quality assessment of numerically reconstructed compressed holograms. in *Proceedings of the SPIE 9599, Applications of Digital Image Processing XXXVIII, 95990K (22 September 2015)*. Vol. 9599. 188–202. (2015).
32. Blinder, D. *et al.* Signal processing challenges for digital holographic video display systems. *Signal Process. Image Commun.* **70**, 114–130 (2019).
33. Laparra, V. *et al.* Perceptual image quality assessment using a normalized Laplacian pyramid. in *Proceedings of the IS&T International Symposium on Electronic Imaging: Human Vision and Electronic Imaging, 2016*. Vol. 28. 1–6. (2016).
34. Chakravarthula, P. *et al.* Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays. *IEEE Trans. Vis. Comput. Graph.* **27**, 4194–4203 (2021).
35. Walton, D. R. *et al.* Metameric Varifocal Holography. Preprint at <https://arxiv.org/abs/2110.01981> (2021).
36. Jo, Y., Yang, S. & Kim, S. J. Investigating loss functions for extreme super-resolution. in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1705–1712. (2020). <https://doi.org/10.1109/CVPRW50498.2020.00220>.
37. Ding, K., Ma, K., Wang, S. & Simoncelli, E. P. Comparison of full-reference image quality models for optimization of image processing systems. *Int. J. Comput. Vis.* **129**, 1258–1281 (2021).
38. Mustafa, A., Mikhailiuk, A., Iliescu, D. A., Babbar, V. & Mantiuk, R. K. Training a task-specific image reconstruction loss. in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 21–30. (2022). <https://doi.org/10.1109/WACV51458.2022.00010>.
39. Goodman, J. W. *Introduction to Fourier Optics* 3rd edn. (Roberts and Company Publishers, 2005).
40. Matsushima, K. & Shimobaba, T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Opt. Exp.* **17**, 19662–19673 (2009).
41. Yeh, L.-H. *et al.* Experimental robustness of Fourier ptychography phase retrieval algorithms. *Opt. Exp.* **23**, 33214–33240 (2015).
42. Cao, L. & Gao, Y. Generalized optimization framework for pixel super-resolution imaging in digital holography. *Opt. Exp.* **29**, 28805–28823 (2021).
43. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems* (eds. Wallach, H. *et al.*). Vol. 32. 8024–8035. (Curran Associates, Inc., 2019).
44. Candès, E. J., Li, X. & Soltanolkotabi, M. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory* **61**, 1985–2007 (2015).
45. Duchi, J. & Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization * Elad Hazan. *J. Mach. Learn. Res.* **12**, 2121–2159 (2011).
46. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. in *3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings* (2015).
47. Kastyulin, S., Zakirov, D. & Prokopenko, D. *PyTorch Image Quality: Metrics and Measure for Image Quality Assessment*. (2019).
48. Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. *Thirty-Seventh Asilomar Conf. Signals Syst. Comput.* **2**, 1398–1402 (2003).
49. Zhang, L., Zhang, L., Mou, X. & Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **20**, 2378–2386 (2011).
50. Zhang, B., Sander, P. v. & Bermak, A. Gradient magnitude similarity deviation on multiple scales for color image quality assessment. in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1253–1257. (2017). <https://doi.org/10.1109/ICASSP.2017.7952357>.
51. Zhang, L., Shen, Y. & Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **23**, 4270–4281 (2014).
52. Reisenhofer, R., Bosse, S., Kutyniok, G. & Wiegand, T. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Process. Image Commun.* **61**, 33–43 (2018).
53. Sheikh, H. R. & Bovik, A. C. Image information and visual quality. *IEEE Trans. Image Process.* **15**, 430–444 (2006).

54. Ding, K., Ma, K., Wang, S. & Simoncelli, E. P. Image quality assessment: Unifying structure and texture similarity. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1–1. (2020). <https://doi.org/10.1109/TPAMI.2020.3045810>.
55. Timofte, R., Gu, S., van Gool, L., Zhang, L. & Yang, M. H. NTIRE 2018 challenge on single image super-resolution: Methods and results. in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2018–June. 965–976. (2018).
56. Agustsson, E. & Timofte, R. NTIRE 2017 Challenge on single image super-resolution: Dataset and study. in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2017–July. pp. 1122–1131 (2017).
57. Lee, B., Kim, D., Lee, S., Chen, C. & Lee, B. High-contrast, speckle-free, true 3D holography via binary CGH optimization. *Sci. Rep.* **12**, 1–12 (2022).
58. ITU-R. *Methodologies for the Subjective Assessment of the Quality of Television Images*. <https://www.itu.int/rec/R-REC-BT.500> (2002).
59. Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324 (1952).
60. Turner, H. & Firth, D. Bradley-Terry models in R: The BradleyTerry2 Package. *J. Stat. Softw.* **48**, 1–21 (2012).
61. Sheikh, H. R., Sabir, M. F. & Bovik, A. C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **15**, 3440–3451 (2006).
62. Ponomarenko, N. *et al.* TID2008-A database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radio-electron.* **10**, 30–45 (2009).
63. Ponomarenko, N. *et al.* Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **30**, 57–77 (2015).
64. Peng, Y., Choi, S., Kim, J. & Wetzstein, G. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Sci. Adv.* **7**, 5040 (2021).

Acknowledgements

F. Y. would like to thank VividQ for support during the period of this research.

Author contributions

F.Y. conceived the original idea, derived the mathematical model, performed experiments, and wrote the manuscript. A.Kad. and R.F. contributed to the optical setup and provided manuscript feedback. B.W. contributed to the subjective experiment design and statistical analysis. The work was initiated and supervised by A.Kac and T.D.W. All authors have given approval to the final version of the manuscript.

Funding

F. Y. would like to acknowledge funding from The Cambridge Trust as well as The China Scholarship Council. A. Kad. would like to acknowledge funding from the Engineering and Physical Sciences Research Council. R. M. would like to thank the Engineering and Physical Sciences Research Council (EP/P030181/1) for financial support during the period of this research. Additionally, B.W. would like to acknowledge funding from the Department of Engineering, University of Cambridge (Richard Norman Scholarship), as well as The Cambridge Trust.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11373-8>.

Correspondence and requests for materials should be addressed to T.D.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022