RESEARCH ARTICLE

H2HSR: Hologram-to-Hologram Super-Resolution With Deep Neural Network

YOUCHAN NO[1], JAEHONG LEE[1], HANJU YEOM[2],
SUNGMIN KWON[1], (Student Member, IEEE), AND DUKSU KIM[1], (Member, IEEE)
[1]School of Computer Science and Engineering, Korea University of Technology and Education (KOREATECH), Cheonan 31253, South Korea
[2]Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, South Korea

Corresponding author: Duksu Kim (bluekdct@gmail.com)

ABSTRACT In holography, the resolution of the hologram significantly impacts both display size and angle-of-view, yet achieving high-resolution holograms presents formidable challenges, whether in capturing real-world holograms or in the computational demands of Computer-Generated Holography. To overcome this challenge, we introduce an innovative Hologram-to-Hologram Super-Resolution network (H2HSR) powered by deep learning. Our encoder-decoder architecture, featuring a novel up-sampling block in the decoder, is adaptable to diverse backbone networks. Employing two critical loss functions, data fidelity and perceptual loss, we guide H2HSR to attain pixel-wise accuracy and perceptual quality. Rigorous evaluations, using the MIT-CGH-4K dataset, demonstrate H2HSR's consistent superiority over conventional interpolation methods and a prior GAN-based approach. Particularly, in conjunction with the SwinIR encoder, H2HSR achieves a remarkable 8.46% PSNR enhancement and a 9.30% SSIM increase compared to the previous GAN-based method. Also, we found that our H2HSR shows more stable reconstruction quality across varying focal distances. These results demonstrate the robustness and effectiveness of our H2HSR in the context of hologram super-resolution.

INDEX TERMS Holography, deep learning, super-resolution, hologram super-resolution.

I. INTRODUCTION

The ability of holography to directly project three-dimensional images, without the need for additional accessories like 3D glasses, has firmly established it as the pinnacle of display technology. At its core, holography relies on the hologram, serving as the medium for precisely encoding and storing holographic information.

Resolution plays a pivotal role in determining the quality of displayed holographic images. The hologram's size and angle-of-view (AoV) are closely tied to its resolution. However, achieving high-resolution holograms is a challenging task, both in capturing real-world holograms due to their

inherent complexity and in Computer-Generated Holography (CGH) [1] due to the substantial computational demands.

Recent works have explored the use of deep neural networks for hologram generation [2], [3], with notable examples such as Tensor Holography [4], which takes RGB-D images as input and produces holograms. While impressive, Tensor Holography requires RGB-D input and needs to regenerate the hologram when the target resolution changes.

Addressing these challenges, particularly the enhancement of hologram resolution, represents a promising avenue for further research. Recent advancements in deep learning, including Convolutional Neural Networks (CNNs) and Transformer-based models, have demonstrated exceptional results in various image super-resolution tasks [5], [6],

---

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

[7], [8], [9], [10], [11], [12]. However, their application in holography remains largely unexplored.

In response, we present an innovative Hologram-to-Hologram Super-Resolution network (H2HSR). Our H2HSR is designed as an encoder-decoder architecture with a unique up-sampling block in the decoder (Sec. III-A). The encoder component is flexible and can utilize various backbone networks. To guide the H2HSR network in effectively capturing both fidelity to ground truth and the perceptual quality of reconstructed holograms, we employ two specific loss functions: data fidelity loss and perceptual loss (Sec. III-B).

We have developed three versions of H2HSR, each leveraging a different backbone network for the encoder. These versions are thoroughly evaluated using the MIT-CGH-4K dataset [4] (Sec. IV). Our evaluation includes comparisons with conventional interpolation methods and a prior deep learning-based (GAN) approach [13].

The results clearly demonstrate the superior performance of H2HSR across multiple dimensions. H2HSR consistently outperforms alternative methods, regardless of the specific encoder network, in terms of both reconstruction quality (Sec. IV-A) and hologram similarity (Sec. IV-C). Notably, when employing the SwinIR [11] encoder, H2HSR achieves an impressive PSNR of 32.85 and an SSIM of 0.91 in reconstruction quality. This represents a significant improvement of 8.46% and 9.30%, respectively, over the prior GAN-based method. Furthermore, our research reveals that H2HSR produces super-resolution results with a more stable quality across varying focal distances (Sec. IV-B).

These findings underscore the robustness and effectiveness of H2HSR in the context of hologram super-resolution, making it a valuable contribution to the field.

II. RELATED WORK
A. COMPUTER-GENERATED HOLOGRAM
Computer-generated holography (CGH) is capable of generating holograms from a virtual environment that includes three-dimensional information, making it one of the most commonly used methods for hologram generation. The techniques employed in CGH vary depending on the type of object representation, such as point clouds [14], polygon meshes [15], [16], [17], and layer-based methods [18], [19]. Additionally, Shimobaba et al. [20] proposed a hybrid approach known as the wavefront recording plane (WRP) method. This technique records a point wave field onto near planes (i.e., WRPs) and propagates them to the hologram plane, thereby reducing the computational complexity associated with point-based CGH.

The quality of a hologram is generally proportional to its resolution. However, as the resolution increases, so do the computational and spatial overheads. To mitigate these overheads, parallel computing approaches utilizing multi-core CPUs [21] and GPUs [22] have been implemented. Additionally, memory-efficient CGH algorithms have been proposed [23]. Despite these performance improvements,

the CGH process for generating ultra high-resolution holograms remains time-consuming. For example, generating a $100K^2$ hologram using CGH can take several hours [24].

B. HOLOGRAM GENERATION WITH DEEP NEURAL NETWORK
Generating high-quality holograms is a complex and time-consuming process. Recent work has attempted to use deep neural networks (DNNs) to overcome these challenges. Hirosaki et al. [2] demonstrated the use of DNNs in generating a phase-only hologram from an image, which can be an alternative to iterative phase refinement algorithms such as the Gerchberg Saxton (GS) algorithm [25]. Eybposh et al. [3] presented DeepCGH, a CNN-based hologram generation network that creates a 3D volume consisting of planes at specific depths, and then uses a U-Net-based network to perform image plane holography. As a result, DeepCGH achieved a speedup with higher accuracy than traditional approaches such as the GS algorithm.

Learning the wave propagation model using neural networks is a key strategy for incorporating neural networks into CGH [26], [27]. Peng et al. [28] proposed the camera-in-the-loop (CITL) optimization method to learn the wave propagation model of a specific display. Based on CITL, they also presented HoloNet, which generates full-color-high-quality holographic images in real-time.

Shi et al. [4] proposed Tensor Holography, a neural network that takes RGB-D images and outputs a hologram. The network has a simple CNN architecture consisting of thirty convolution layers, and it shows real-time hologram generation performance on a consumer-grade GPU (e.g., 60Hz for $1,920 \times 1,080$ pixel resolution). To train the network, they generated a set of scenes in a virtual environment and made a set of holograms based on CGH. This dataset is called MIT-CGH-4K, and it consists of an RGB image, a depth image, and holograms in two resolutions. Since loss functions like MSE do not work properly for holograms, they also proposed novel loss functions, including ASM (Angular Spectrum Method) loss and total-variation loss, to measure the inference accuracy over the ground-truth hologram. They recently extended Tensor Holography to version 2 with an unsupervised learning approach to decrease noise at the outline in the reconstructed image [29].

The MIT-CGH-4K dataset includes a pair of holograms in two resolutions: 192-by-192 with a 16 µm pixel pitch and 384-by-384 with an 8 µm pixel pitch. We use this dataset to train our super-resolution network. Furthermore, we employ the loss function proposed in Tensor Holography to ensure the accuracy of our super-resolution results.

C. IMAGE SUPER-RESOLUTION
Image super-resolution has been actively studied for a long time [30]. Traditional approaches involve interpolating the values of existing pixels to fill the empty pixels added to increase resolution. Common interpolation methods include

nearest-neighbor, bilinear, and bicubic algorithms [31], [32]. Although simple and generally applicable to various applications, these methods are not suitable for super-resolution (or up-sampling) because they cannot guarantee the quality of the result since they only exploit the values of neighboring pixels.

The evolution of deep learning has provided a solution to the problem of interpolation-based approaches [30]. SRCNN [5] showed that a convolutional neural network (CNN) could generate a more natural high-resolution image from a low-resolution input image than the bicubic interpolation method. Kim et al. [6] achieved higher accuracy than SRCNN by using a very deep CNN. Further works improved the quality of the super-resolution result using more complex or advanced CNN architectures [7], [8], [9], [33]. The network depth grew using a residual block [34] or a memory block [35]. Shi et al. [36] introduced an innovative decoder component known as the pixel shuffle layer, which demonstrates superior performance compared to the transposed convolution layers used in previous studies. This efficiency is achieved by rearranging the pixels of a feature map in a channel-wise order. While it effectively reduces the number of channels when upscaling the resolution of features, it accomplishes this without the need to train any additional parameters. While the deeper network generated better results, Zhang et al. [10] found that prior work did not fully utilize the hierarchical features in each convolutional layer. Zhang et al. proposed the residual dense network (RDN) to address this issue. RDN has dense residual connections, making it possible to fully use the hierarchical features from a low-resolution input image. RDN is one of the state-of-the-art CNN models for super-resolution, and we employ it as one of the backbones for our hologram super-resolution network.

Dosovitskiy et al. [37] introduced the vision Transformer (ViT), which adapts Transformer [38] to image tasks. Swin Transformer [39] extends ViT with a hierarchical feature map to a general-purpose backbone for computer vision. Liang et al. [11] proposed SwinIR, an image super-resolution model based on Swin Transformer. Also, Chen et al. [12] presented HAT (Hybrid Attention Transformer) based on Swin Transformer, which combines color channel-wise attention mechanisms. These Transformer-based methods for image super-resolution are currently state-of-the-art models. Therefore, we employ SwinIR and HAT as candidates for the backbone of our network.

D. HOLOGRAM SUPER-RESOLUTION

In the field of microscopy, various optical-based super-resolution methods have been studied, including those that utilize holography to enhance resolution at nanometer scales [40], [41], [42]. For coherent imaging systems, various super-resolution methods based on optical and numerical techniques have been proposed to overcome the limitations imposed by pixel size and diffraction [43], [44], [45]. Also, deep learning-based super-resolution frameworks were proposed [46], [47], [48], [49], like Liu et al. [50] used a generative adversarial network (GAN) to enhance the resolution of both pixel size-limited and diffraction-limited coherent imaging systems. Coherent imaging systems are commonly used in microscopy, and the goal of those work is to retrieve missing information (e.g., phase) due to sensor chip limitations. Unlike those works, our method targets the super-resolution of a full-color hologram generated by computer-generated holography (CGH).

Several works have attempted to use deep learning-based super-resolution for holograms generated from CGH [13], [51], [52]. Liu et al. [52] proposed a new method that exploits a Variational Autoencoder (VAE) [53] to develop real-time hologram generation. They introduced a new hologram format, SSHM (Spatial Spectrum of Hologram Modulator), to better represent complex-valued frequency data to train the network. They used three hologram animations for each 3D object as the dataset and achieved impressive results. However, it may be challenging to apply their network to holograms with different contents from the original data used in training, as VAE is generally better suited for generating data similar to the original data used in training. Additionally, they did not present any quantitative evaluations, such as PSNR (Peak Signal-to-Noise Ratio) or SSIM (Structural Similarity Index Map), to support their results. Jee et al. [13] proposed using dual GAN for hologram generation, with the input and output being amplitude and phase. They used the MSE as the loss function, and the MIT-CGH-4K dataset [4] to train their network, achieving better results than the interpolation method. However, we found that evaluating the quality of the reconstructed image is crucial for displaying holograms. Like Jee et al. [13], we used the MIT-CGH-4K dataset to train our network for holograms generated from CGH. However, we further analyzed the results by measuring the reconstruction quality.

III. HOLOGRAM SUPER-RESOLUTION WITH DEEP NEURAL NETWORK

In this work, we aim to design an end-to-end super-resolution network that takes a low-resolution hologram as input and generates a high-resolution hologram as output. The resolution of a hologram affects the reconstructed image in two ways: the reconstructed image's size and the hologram's AoV. Consequently, there are two types of hologram super-resolution. The first method involves increasing the number of pixels while decreasing the pixel pitch, which leads to a wider AoV while maintaining the reconstructed image size. The second method involves up-scaling the volume of the hologram while maintaining the pixel pitch, similar to 2D image super-resolution methods. Our approach targets the first case, which involves increasing the resolution of the hologram by decreasing the pixel pitch.

Dataset: To train our super-resolution network, we used the MIT-CGH-4K dataset [4]. This dataset includes four thousand samples, each containing a pair of low- and high-resolution holograms generated from the same 3D

scene. The low-resolution hologram has a resolution of 192×192 and a pixel pitch of 16 μm, while the high-resolution hologram has a 384×384 and a pixel pitch of 8 μm. Both resolutions' physical width and height are 3.072mm, and the depth volume is 6mm. It is a full-color hologram consisting of three color channels whose wavelengths, ordered by red, green, and blue, are 638nm, 520nm, and 450nm, respectively. This dataset satisfies the requirement of having a pair of low- and high-resolution holograms generated from the same content, which is essential for training our super-resolution network. Also, since the physical volume of the two resolutions is the same, we can properly compare the reconstructed images with the same focal length. The MIT-CGH-4K dataset is partitioned into 3,800 training images, 100 validation images, and 100 test images, with all evaluations performed using the test set.

A. H2HSR NETWORK

Our hologram super-resolution network (i.e., H2HSR) has an encoder-decoder model, and the input and output are a set of amplitudes and phases. Since each color channel has amplitude and phase and the color hologram consists of RGB (red, green, and blue) colors, the input and output of H2HSR are six channels.

1) INPUT AND OUTPUT

H2HSR takes the amplitude and phase of a hologram as its input, and the output is also a pair of amplitude and phase with a resolution four times larger than the input (e.g., $192^2 \rightarrow 384^2$). Before feeding the input into the H2HSR network, we normalize the input amplitude and phase values, ranging from 0 to 1. However, we need to set the maximum value as a hyper-parameter for the amplitude input.

Regarding the MIT-CGH-4K dataset, the phase values are already normalized within the range of 0 to 1, so we directly use these normalized phase values. However, for amplitude normalization, we calculate the maximum value among all the samples (holograms) in the dataset and use it as the maximum value for the normalization process.

The output of H2HSR, including both the amplitude and phase, also falls within the range of 0 to 1. To obtain the final output amplitude, we apply inverse normalization by multiplying the amplitude value with the hyper-parameter (i.e., the maximum value). This process ensures that the output amplitude is restored to its original scale and can be properly interpreted. The phase output remains within the range of 0 to 1 as it does not require further normalization adjustments.

2) ENCODER

The encoder module plays a crucial role in extracting informative features from the input hologram. In our approach, we leverage well-established models that have shown exceptional performance in image super-resolution tasks. Specifically, we evaluated three state-of-the-art image

super-resolution models: RDN [10], SwinIR [11], and HAT [12].

To incorporate these image super-resolution models into our hologram super-resolution network, we made modifications to the first layer of the backbone models. By extending the first layer, we adapted the models to handle holograms with six channels, corresponding to the amplitude and phase components. Consequently, the encoder generates a feature map with dimensions $[6, H, W]$, where H and W denote the height and width of the input hologram, respectively. This feature map serves as a representation of the input hologram and is subsequently utilized for further processing in the network.

The output dimension of the encoder is $[C, H, W]$, where C denotes the number of feature map channels and is determined by the choice of the backbone model. In our method, the value of C is 64 for the RDN-based encoder, while it is 180 for the SwinIR and HAT-based encoders.

3) DECODER

The decoder network plays a crucial role in generating the super-resolved hologram from the feature maps obtained from the encoder. In the field of image super-resolution, several methods have been explored, including convolution with nearest-neighbor interpolation [5], [54], transposed-convolution [55], and pixel-shuffle [36]. Among these methods, the pixel-shuffle technique has demonstrated superior performance in capturing high-frequency details, which are particularly important for holograms due to their sensitivity to high-frequency information. Our experimentation revealed that the H2HSR network attains the best outcomes when the decoder layer employs the pixel-shuffle method. Therefore, we have integrated the pixel-shuffle technique into our decoder network to ensure the highest fidelity in preserving high-frequency details within the super-resolved hologram.

Based on the pixel-shuffle method, we have devised an up-sampling block that consists of two convolution layers along with a pixel-shuffle layer (Fig. 1). For SwinIR and HAT encoder, we added a convolution layer before the up-sampling block to reduce the feature map channels from 180 to 64 due to the GPU memory limit.

The first convolution layer of the up-sampling block increases the number of feature map channels by a factor of four, preparing the feature map for the subsequent pixel-shuffle layer that enlarges the feature map size by consolidating four channels into one. This progression entails that the first convolution layer takes in a $[C, H, W]$ feature map and generates an output of $[4C, H, W]$. Following the pixel-shuffle layer, the output is further transformed into $[C, 2H, 2W]$. Finally, the final convolution layer shapes the up-sampling block's output dimensions to $[6, 2H, 2W]$, with each channel corresponding to the amplitude and phase components for each color channel, mirroring the arrangement of the input hologram.
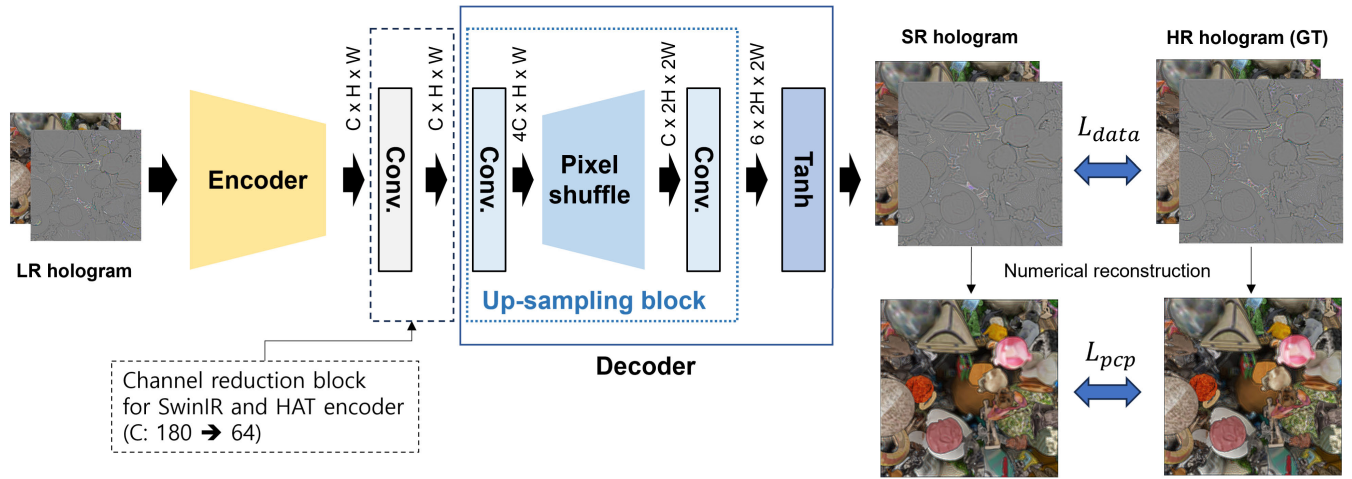
FIGURE 1. Overview of H2HSR architecture.

At the termination of the decoder, a hyperbolic tangent layer is employed to normalize the output values, confining them within a range of -1 to 1. This range mirrors the value range of the input amplitude and phase components, ensuring consistency in the representation of the hologram data.

B. LOSS FUNCTION

In the context of image super-resolution, assessing the similarity between images by comparing the pixel values and using metrics such as Mean Square Error (MSE) to measure the performance is common. This approach works well for images as they primarily consist of color values for each pixel, and the quality of the super-resolution result can be evaluated by comparing it to the ground-truth high-resolution image.

However, when working with holograms, measuring quality using MSE presents challenges for two reasons. Firstly, holograms consist of amplitude and phase components, unlike images with only color values. Directly comparing pixel values is not sufficient for evaluating holograms. Additionally, the periodicity of the phase makes MSE unsuitable for measuring hologram quality accurately. Secondly, in the hologram context, the quality of the reconstructed results is more important than the individual values in the amplitude and phase components.

To address these challenges, we utilize two specific loss functions proposed by Shi et al. [4]: the data fidelity loss and the perceptual loss. These loss functions are specifically designed to handle holograms and effectively capture both the fidelity to the ground truth and the perceptual quality of the reconstructed hologram.

The loss function we use to train the H2HSR network is represented by Eq. 1. In this equation, α represents a weight parameter that balances the contributions of the data fidelity loss, denoted as L_{data} (Sec. III-B1), and the perceptual loss, denoted as L_{pcp} (Sec. III-B2). By combining these two loss functions, we can effectively guide the training process of the H2HSR network to ensure both fidelity to the ground truth

and perceptual quality in the super-resolved hologram.

$$L = \alpha L_{data} + (1 - \alpha) L_{pcp} \qquad (1)$$

1) DATA FIDELITY LOSS

The data fidelity loss measures the pixel-wise accuracy between the super-resolution result and the ground-truth high-resolution hologram, considering both the amplitude and phase components. Eq. 2 represents the data fidelity loss used in our H2HSR network.

$$L_{data} = \|\hat{A} - A e^{\delta(\hat{\theta}, \theta) - \overline{\delta}(\hat{\theta}, \theta)}\|_2, \qquad (2)$$

where the estimated amplitude and phase are denoted as $\hat{A}$ and $\hat{\theta}$, respectively, while the ground-truth amplitude and phase are represented as A and θ. The term $\overline{\delta}$ corresponds to the average phase difference across the RGB channels, which helps reduce disparities. The function captures the phase difference between the estimated and ground-truth phases $\delta(\hat{\theta}, \theta)$, as defined in Eq. 3.

$$\delta(\hat{\theta}, \theta) = \text{atan2}(\sin(\hat{\theta} - \theta), \cos(\hat{\theta} - \theta)) \qquad (3)$$

The phase is a periodic value ranging from $-\pi$ to π, and it is treated within an absolute coordinate system to ensure accurate comparison. Minimizing the phase difference in the δ term, the sine term approaches zero, and the cosine term approaches one, resulting in the exponential term being close to one in Eq. 2. This implies that the absolute angle tends to approach zero, leading to a better alignment between the estimated and ground-truth amplitudes.

2) PERCEPTUAL LOSS

The perceptual loss measures the similarity between images reconstructed from the super-resolved hologram and the ground-truth high-resolution hologram. It incorporates two components: the ASM (Angular Spectrum Method) loss and the total-variation loss introduced by Shi et al. [4].

The ASM loss is specifically designed for near hologram reconstruction, such as those found in the MIT-CGH-4K dataset. Using the ASM, we can numerically reconstruct the display result on a hologram display device, such as a Spatial Light Modulator (SLM), at a specific distance. This reconstructed result is represented as an image. The ASM loss measures the difference between the reconstructed image using the super-resolved hologram ($\hat{H}$) and the ground-truth high-resolution hologram (H). It is defined as the average pixel-wise absolute difference between the two images, as shown in Eq. 4:

$$L_{ASM}(d) = \frac{1}{N} \sum_{i=1}^{N} W(d, i) \cdot |ASM(\hat{H}, d)_i - ASM(H, d)_i| \tag{4}$$

In this equation, $ASM(X, d)$ represents the reconstructed image using a hologram X at distance d, and $ASM(X, d)_i$ denotes the pixel value at the i-th pixel in $ASM(X, d)$. N is the total number of pixels in the image.

The weight function $W(d, i)$ is introduced to give pixel-wise weights based on the proximity of the i-th pixel's depth to the desired focal distance d, as shown in Eq. 5:

$$W(d, i) = \frac{\exp(\beta \cdot (Z_{max} - |Z_i - d|))}{Z_{max}} \tag{5}$$

In this equation, Z is the depth map associated with the ground-truth hologram H, and Z_{max} and Z_i are the maximum depth value in Z and the depth value of the i-th pixel, respectively. The parameter β controls the rate of weight decay. The weight function $W(d, i)$ plays a crucial role in the perceptual loss by assigning higher weights to pixels that are closer to the desired reconstruction distance. This enables the network to focus more on the regions of the hologram that have a perceptually stronger impact on the reconstructed image quality.

The total-variation loss captures the variation among pixels in the reconstructed image, promoting a more natural image reconstruction. It measures the difference between the total variations of $ASM(\hat{H}, d)$ and $ASM(H, d)$, as shown in Eq. 6, where $\nabla ASM(X, d)_i$ represents the gradient of $ASM(X, d)_i$.

$$L_{TV}(d) = \frac{1}{N} \sum_{i=1}^{N} W(d, i) \cdot |\nabla ASM(\hat{H}, d)_i - \nabla ASM(H, d)_i| \tag{6}$$

The perceptual loss (L_{pcp}) is computed as the average of the ASM loss (L_{ASM}) and the total-variation loss (L_{TV}) over all the reconstruction distances d in the set D, as shown in Eq. 7:

$$L_{pcp} = \frac{1}{|D|} \sum_{d \in D} (L_{ASM}(d) + L_{TV}(d)) \tag{7}$$

By utilizing the perceptual loss, we ensure that the network is trained to optimize both the similarity to the ground-truth hologram at various distances and the naturalness of the reconstructed images. This comprehensive loss function encourages the network to generate super-resolved holograms that exhibit high fidelity and perceptual quality.

Distance selection: Holograms generate a continuous volume, making it infeasible to evaluate perceptual loss for every possible focal distance. To overcome this challenge, we employ a distance selection strategy that consists of key distances and random distances. To identify the key distances, we construct a histogram using the depth values from the associated depth map. From this histogram, we select the top k depths as the key distances for the hologram. This approach allows us to focus on specific depth characteristics that are deemed important for hologram reconstruction. In addition to the key distances, we include random distances in our set. These random distances are uniformly sampled within the hologram volume (e.g., -3mm to 3mm). By incorporating both key distances and random distances, we consider both localized depth features and the overall depth distribution, enabling a comprehensive evaluation of the reconstructed holograms.

IV. EXPERIMENTS

We developed three distinct variants of the H2HSR network, each featuring a unique encoder network: RDN [10], SwinIR [11], and HAT [12].

- $H2HSR_{RDN}$ is a H2HSR network that incorporates the RDN as its encoder. For RDN, we opted for a configuration featuring 16 RDBs (Residual Dense Blocks), each comprising 8 convolution layers with 64 channels.
- $H2HSR_{SwinIR}$ uses the SwinIR as the encoder of the H2HSR network. The SwinIR is based on a medium-size Transformer model with an embedding space of 180 dimensions.
- $H2HSR_{HAT}$ employs HAT as its encoder. Similar to $H2HSR_{SwinIR}$, HAT is founded on a medium-sized Transformer model with an embedding space of 180 dimensions.

The $H2HSR_{RDN}$ was trained with a batch size of 16, while $H2HSR_{SwinIR}$ and $H2HSR_{HAT}$ used a batch size of 8. We initialized the learning rate at 0.001 and implemented a step-scheduler strategy, reducing the learning rate by half every 50 epochs to ensure stable convergence. For all models, we used a fixed training epoch count of 200 for consistency. As an optimizer, we employed the Adam optimizer [56], configuring alpha to 0.99 and beta to 0.999. For the perceptual loss (Sec. III-B2), we set the weight (β) to 0.35 following the principles of the Tensor Holography approach. For the loss function in Eq. 1, we have empirically determined that setting α to 0.5 yields the best performance in our experiments.

During our experiments, we used two distinct system configurations. For training, we employed four NVIDIA RTX A6000 GPUs, each equipped with 48GB of VRAM. For testing and validation, we relied on a single NVIDIA RTX 3090 GPU with 24GB of VRAM. SwinIR and HAT

TABLE 1. Super-resolution performance comparison of eight methods.

	$nearest$	$bilinear$	$bicubic$	$DGAN_{G1}$	$DGAN_{G2}$	$H2HSR_{RDN}$	$H2HSR_{SwinIR}$	$H2HSR_{HAT}$
PSNR	26.8861	28.3991	27.9790	30.3045	30.2564	31.4138	**32.8486**	32.6513
SSIM	0.6697	0.7298	0.7345	0.8357	0.8354	0.8750	**0.9134**	0.9086

typically required approximately 50 hours, while the RDN model demanded nearly two full days of training.

Baselines: In our evaluation, we benchmark H2HSR against existing methods, encompassing both conventional interpolation techniques and a previously introduced hologram super-resolution approach. Firstly, we explore conventional interpolation methods to understand their effectiveness. Specifically, we independently upscale the amplitude and phase components of the holograms. We assess the performance of three interpolation methods: nearest-neighbor (*nearest*), *bilinear*, and *bicubic*.

Our second baseline is a hologram super-resolution method utilizing GANs, as proposed by Jee et al. [13]. Since the implementation code for this method is not publicly available, we implemented it ourselves while acquiring implementation details from the author. Following the paper's specifications, we constructed a generator based on the Cascading Residual Network (CARN) [57] and a discriminator adapted from EfficientNet (size B2) [58]. Similar to their dual-generator GAN model, we employed and trained two generators. Consequently, we obtained two distinct generators, which we label as $DGAN_{G1}$ and $DGAN_{G2}$.

Evaluation metrics: As previously discussed in the context of loss design (Sec. III-B), assessing the quality of the reconstructed results holds paramount importance in holography. Therefore, our primary focus in evaluating super-resolution performance lies in the assessment of these reconstructed results. Specifically, we employ two widely-used metrics in the super-resolution field: PSNR and SSIM. To conduct this evaluation, we generate seven numerically reconstructed images corresponding to focal distances ranging from -3mm to 3mm at 1mm intervals. Subsequently, we calculate PSNR and SSIM values by comparing these reconstructed images with the high-resolution hologram (i.e., the ground truth), for each of these distances. The final metrics are then derived by averaging these values.

Numerical reconstruction method: For the numerical reconstruction process, we utilized the ASM since the hologram plane is close to the scene in the MIT-CGH-4K dataset. The equation for ASM is detailed in Eq.8:

$$u_2(x_2, y_2) = \mathcal{F}^{-1}\left[U(f_x, f_y) \exp\left(i2\pi z \sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2} \right) \right]$$
(8)

In this equation, $u_2(x_2, y_2)$ represents the reconstructed two-dimensional scene at a specific distance z. The term $U(f_x, f_y)$, which denotes the Fourier space of hologram

plane, is multiplied by an exponential factor incorporating λ (the wavelength) and the spatial frequencies f_x and f_y. To transform the Fourier domain into the spatial domain, we employ the inverse Fourier transform, denoted as F^{-1}. In our method, the ASM incorporates zero-padding to mitigate the effects of circular convolution during the Fourier transform. To reduce aliasing and unnecessary high frequencies in the hologram, our implementation of ASM is based on the band-limited ASM approach described by Matsushima et al. [59].

A. RESULTS

Table 1 provides a quantitative comparison of the super-resolution performance among eight different methods. The first three columns on the left display the results achieved by interpolation-based approaches. While bilinear and bicubic interpolation methods demonstrate higher performance than nearest interpolation, they fall significantly short when compared to neural network-based methods, as indicated by both the PSNR and SSIM metrics.

Two GAN-based methods, denoted as $DGAN_{G1}$ and $DGAN_{G2}$, exhibit similar performance to each other. They achieve considerably higher performance, approximately 2 in terms of PSNR and 0.1 for SSIM, when compared to the bicubic interpolation method.

However, our H2HSR consistently outperforms the *DGAN* models, achieving up to 2.5 in PSNR and 0.078 in SSIM improvement. The key distinction between our H2HSR and the *DGAN* models is that the latter focus on restoring the amplitude and phase of the high-resolution hologram without considering the reconstructed results. As a result, our H2HSR attains the highest performance among all methods evaluated. Among the three encoders we employed, the transformer-family models, specifically SwinIR and HAT, demonstrate superior performance compared to RDN in our H2HSR network. Notably, an intriguing observation emerges: while HAT exhibits superior performance to SwinIR in the image super-resolution field, SwinIR surpasses HAT in the context of H2HSR for hologram super-resolution. Ultimately, $H2HSR_{SwinIR}$ emerges as the top-performing model, boasting impressive scores of 32.85 in PSNR and 0.91 in SSIM.

1) QUALITATIVE COMPARISON

In Fig. 2, we present the numerical reconstruction results achieved through the ASM for holograms super-resolved by three distinct methods and the ground truth high-resolution hologram. Upon closer examination of the zoomed images

FIGURE 2. Numerical reconstruction results obtained from hologram super-resolution using three different methods. The focused objects are highlighted within red boxes, and the second row provides magnified views of these regions. More comparison results are available in the supplementary report.

in the focused region, marked by the red box, notable differences become evident. In the results obtained from the interpolation method (*bicubic*), aliasing artifacts are pronounced, especially around the edges. Additionally, these results exhibit an overall smoothing effect when compared to GT. Also, although the *DGAN* manages to avoid aliasing artifacts around the edges, it still introduces a smoothing effect across the entire image, including the edges.

In contrast to the two aforementioned methods, our $H2HSR_{SwinIR}$ achieves results that closely match the GT. Notably, in the first scene, where the green object and the white cube overlapped, we observed a remarkable preservation of edges, mirroring the quality seen in the GT results. This also demonstrates that our method successfully captures the volume information embedded in the hologram to authentically represent the scene. A similar result can
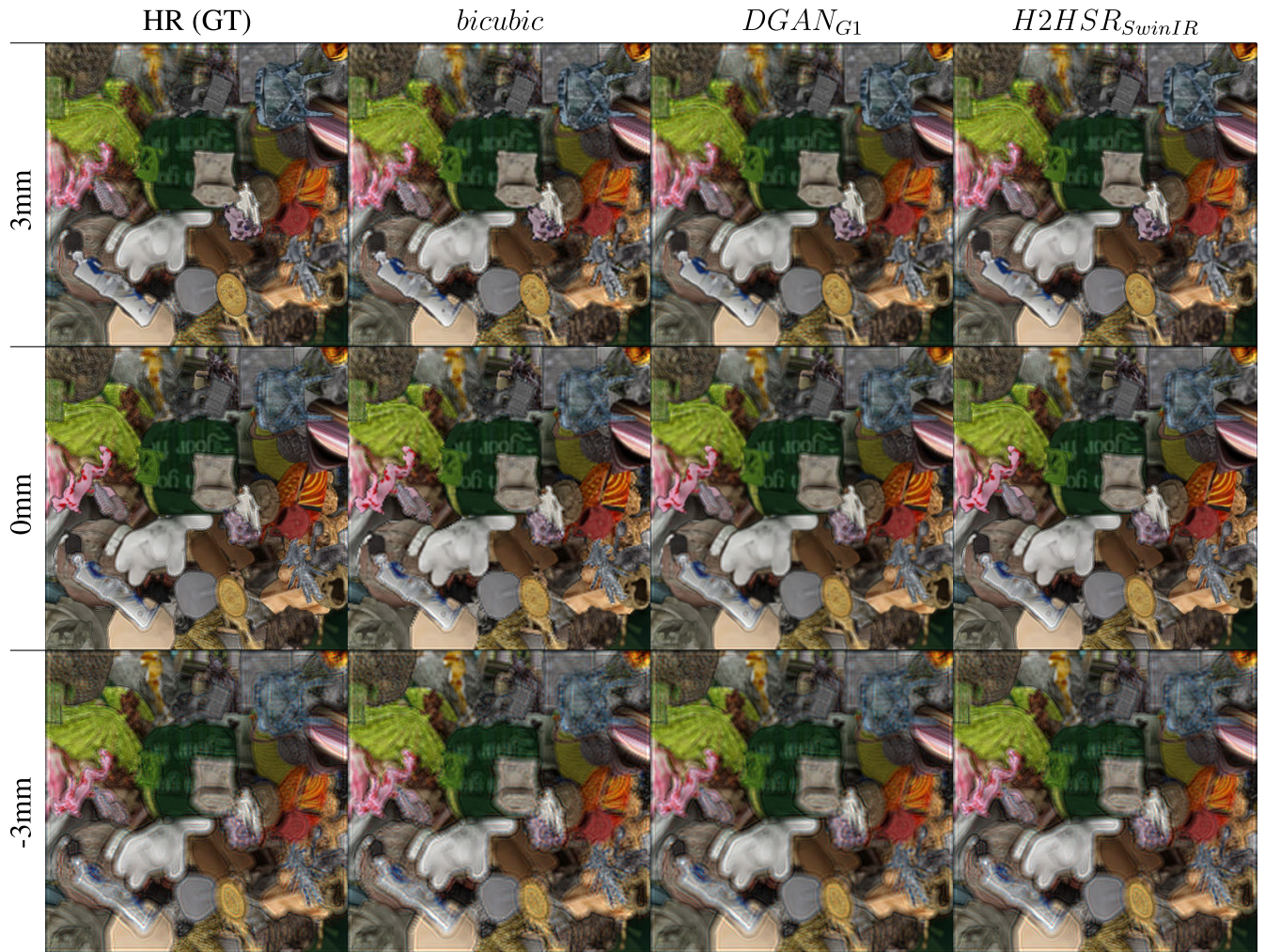
HR (GT)	*bicubic*	$DGAN_{G1}$	$H2HSR_{SwinIR}$

FIGURE 3. Numerical reconstruction results for three focal distances.

be found in the second scene, featuring a yellow circular object. Additionally, $H2HSR_{SwinIR}$ exhibits minimal smoothing effects when compared to the other methods, all while successfully preserving the texture in a manner similar to the GT.

B. RECONSTRUCTION QUALITY ACROSS THE FOCAL DISTANCE

Fig. 3 illustrates the reconstruction results obtained at three different focal distances: $3mm$, $0mm$, and $-3mm$. The qualitative trends observed in the reconstructed results align with those found in Fig. 2. To assess the influence of focal distance on reconstruction quality, we conducted separate evaluations of PSNR and SSIM for specific distances.

We found that the PSNR values remained relatively stable across varying focal distances. However, it is important to emphasize that the SSIM values for the *bicubic* method displayed a notable decline as the distance from the center (i.e., $0mm$) increased (Table 2). For instance, the SSIM values exhibited a difference of 0.044 between $0mm$ and $-3mm$, with a standard deviation of 0.0159. This decline in

SSIM aligns with prior research findings [60], suggesting that hologram super-resolution may encounter challenges related to preserving volumetric properties within the reconstructed scene.

Differing from the interpolation method, which solely considers pixel values, both the *DGAN* and *H2HSR* approaches leverage deep learning to capture hologram characteristics in the super-resolution context. This may contribute to their lower standard deviation, indicating higher robustness. Furthermore, *H2HSR* differs from *DGAN* in its training approach. While *DGAN* primarily focuses on the hologram's amplitude and phase, *H2HSR* takes into account reconstruction quality through perceptual loss (see Sec. III-B2). This comprehensive approach results in *H2HSR* exhibiting the most robust performance across varying focal distances.

C. HOLOGRAM SIMILARITY AND RECONSTRUCTION QUALITY

Given that holograms are mediums for recording three-dimensional images, the quality of a hologram is intrinsically

TABLE 2. This table presents SSIM values across varying focal distances, along with the corresponding standard deviation (SD).

focal distance	-3mm	-2mm	-1mm	0mm	1mm	2mm	3mm	SD
$bicubic$	0.7445	0.7563	0.7710	0.7885	0.7899	0.7848	0.7691	0.0159
$DGAN_{G1}$	0.8350	0.8468	0.8484	0.8447	0.8350	0.8246	0.8156	0.0113
$H2HSR_{SwinIR}$	0.9051	0.9088	0.9130	0.9161	0.9189	0.9186	0.9135	**0.0047**

TABLE 3. Hologram similarity (amplitude and phase) to the ground truth hologram as assessed by PSNR and SSIM. The values within brackets in the $DGAN_{G1}$ column represent reported values from the paper [13], while the remaining values in the column are measured through our experiments.

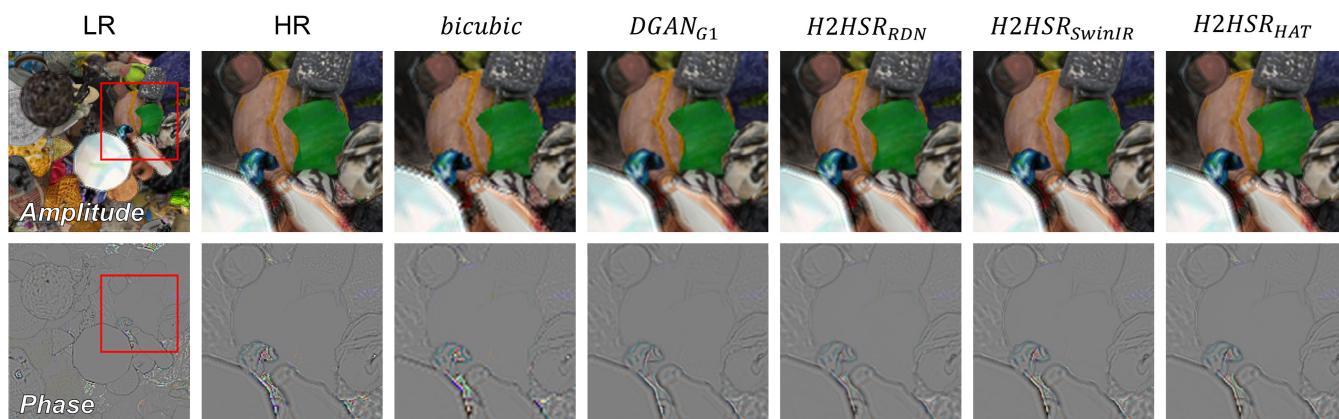|		$bicubic$	$DGAN_{G1}$	$H2HSR_{RDN}$	$H2HSR_{SwinIR}$	$H2HSR_{HAT}$
Amplitude	PSNR	29.6625	31.9558 (29.3761)	33.0012	**34.4330**	34.2299
	SSIM	0.7541	0.8158 (0.8933)	0.8588	**0.9017**	0.8960
Phase	PSNR	24.5054	**26.3356 (28.2232)**	22.7960	22.1054	22.7648
	$PSNR_{Phase}$	25.6759	26.5909	26.6793	**28.2339**	27.7295
	SSIM	0.7300	0.7911 (0.8776)	0.8209	**0.8756**	0.8665

FIGURE 4. Comparison of the amplitude and phase components in holograms generated by five super-resolution methods.

linked to the visual quality of the reconstructed images. One might intuitively assume that higher hologram similarity should correspond to visually similar reconstruction outcomes. However, encapsulating this intricate relationship within a single, comprehensive similarity metric remains a formidable challenge.

Jee et al. [13] sought to evaluate the performance of their hologram super-resolution method by quantitatively comparing the amplitude and phase of the super-resolved hologram with those of the ground-truth hologram. To assess the similarity, they conducted measurements of PSNR and SSIM independently for both the amplitude and phase components. However, unfortunately, they did not provide information regarding the reconstruction quality of the super-resolved hologram produced by their network.

To investigate the relationship between hologram similarity and reconstruction quality, we evaluated the PSNR and SSIM for both the amplitude and phase of holograms super-resolved by different methods (Table 3). We observed a strong correlation between hologram similarity, as assessed

by PSNR and SSIM, and reconstruction quality, particularly for the amplitude component. This correlation aligns with the quantitative evaluation results presented in Table 1, where $DGAN_{G1}$ demonstrated higher PSNR and SSIM values compared to $bicubic$. Furthermore, the H2HSR methods consistently achieved higher values in both metrics than the other alternatives.

For phase, SSIM appears to have a correlation with reconstruction quality. However, PSNR does not align with the reconstruction quality. Although $DGAN_{G1}$ achieves a higher PSNR than $bicubic$, which may appear to align with reconstruction quality, $H2HSR_{SwinIR}$, which exhibits the highest reconstruction quality, receives a significantly lower PNSR score compared to other methods.

Fig. 4 compares the amplitude and phase of holograms super-resolved from a low-resolution hologram using different super-resolution methods. We observed that the phase obtained from $DGAN_{G1}$ exhibits more blurriness compared to the ground truth (HR), while the H2HSR method achieves a phase more similar to HR than $DGAN_{G1}$. It means that
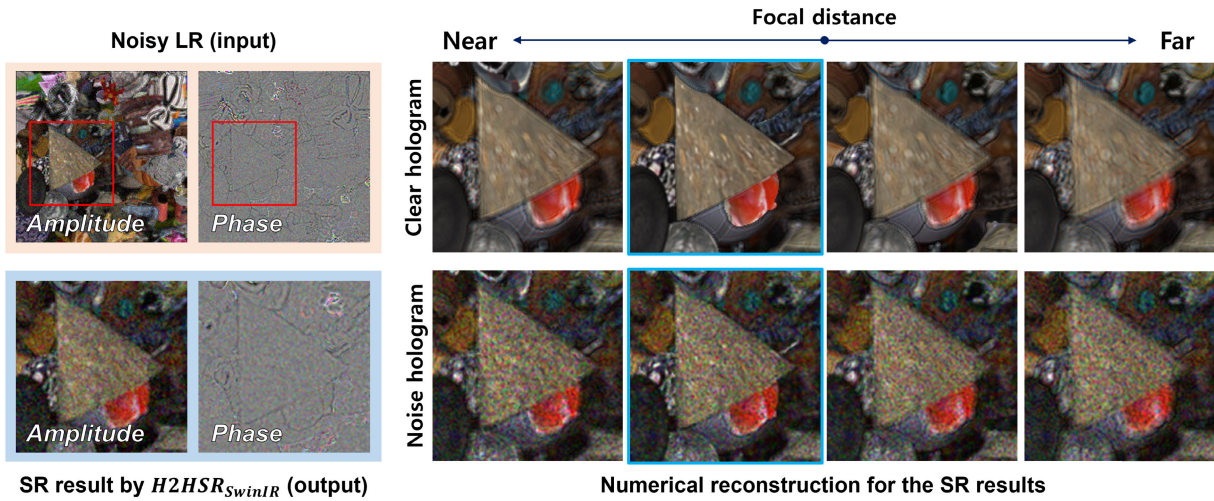
FIGURE 5. The LR hologram inputs with noise and the corresponding super-resolution results achieved by H2HSR.

H2HSR methods restore the high-frequency information. Among H2HSR methods, $H2HSR_{RDN}$ and $H2HSR_{HAT}$ gets a little more blurred results in phase than $H2HSR_{SwinIR}$ while the $H2HSR_{SwinIR}$ achieve lower $PSNR_{phase}$ than others. In terms of images, applying blur can reduce artifacts and enhance visual quality. However, smoothing the high-frequency information in phase can lead to loss of details in reconstruction results. Consequently, we can conclude that it is hard to measure the quality of hologram by PSNR.

It is important to note that standard PSNR is not suitable for directly assessing phase information, given its angular nature expressed in radians. Therefore, we utilized $PSNR_{phase}$ as introduced by Oh et al. [61] to address this limitation. As a result, we figure out that the $PSNR_{phase}$ values alleged with the reconstruction quality while $H2HSR_{SwinIR}$ gets the highest value.

We incorporated perceptual loss into our training process, a departure from the approach employed by *DGAN*, which relies solely on MSE loss for amplitude and phase. This design choice enables our H2HSR method to adjust phase in tandem with amplitude values, resembling optimization techniques like Stochastic Gradient Descent (SGD) for phase refinement [62]. As a result, H2HSR excels in accurately restoring the phase within the context of reconstructed images when compared to *DGAN*. Although individual pixel values may be less similar to the HR, these results underscore the effectiveness of our approach for the task of hologram super-resolution.

D. ROBUSTNESS TO NOISE HOLOGRAM
Unlike holograms generated by CGH, those captured by holographic cameras or microscopy often contain noise. Our method directly processes holograms, enabling its application to real-world captures. To evaluate our system's noise robustness, we conducted experiments using H2HSR
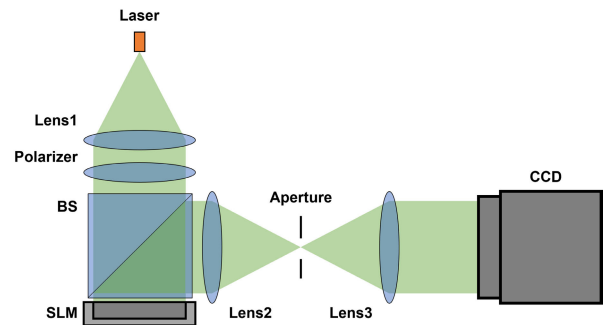

FIGURE 6. Overview of the optical-based experiment configuration.

to super-resolve noisy holograms. We introduced Gaussian noise into a LR hologram in both amplitude and phase components, characterized by zero mean and a standard deviation of 0.05. Fig. 5 displays the noisy LR hologram and the super-resolved (SR) result achieved through H2HSR. Although the results exhibit noise, originating from the noisy input LR hologram, the hologram retains robustness in terms of focal distance. In this work, we focused on super-resolution rather than denoising while training the networks with clear hologram pairs. Nonetheless, these results indicate that our method could work with damaged holograms too.

E. OPTICAL RECONSTRUCTION RESULTS
We conducted a qualitative evaluation of optical reconstruction using our Spatial Light Modulator (SLM), HOLOEYE LETO 3 Phase only LCoS. The SLM has a resolution of 1920×1080 and a pixel pitch of 6.4 μm. Fig. 6 shows the configuration for the optical experiment. To facilitate the use of a phase-only SLM, we employ the double phase amplitude coding (DPAC) [63] to extract the phase information from the complex-valued hologram. The light source used is a laser
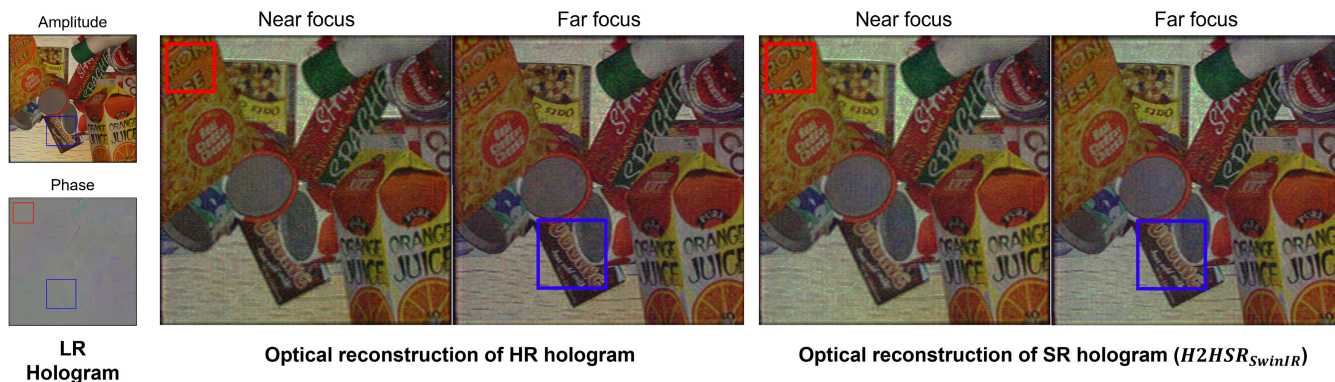
FIGURE 7. The left-most column shows the LR hologram used for the optical experiment, while the second and third columns display the optical reconstruction results of the HR hologram and the SR hologram from H2HSR, respectively. The color boxes indicate the zoomed regions shown in Fig. 8.
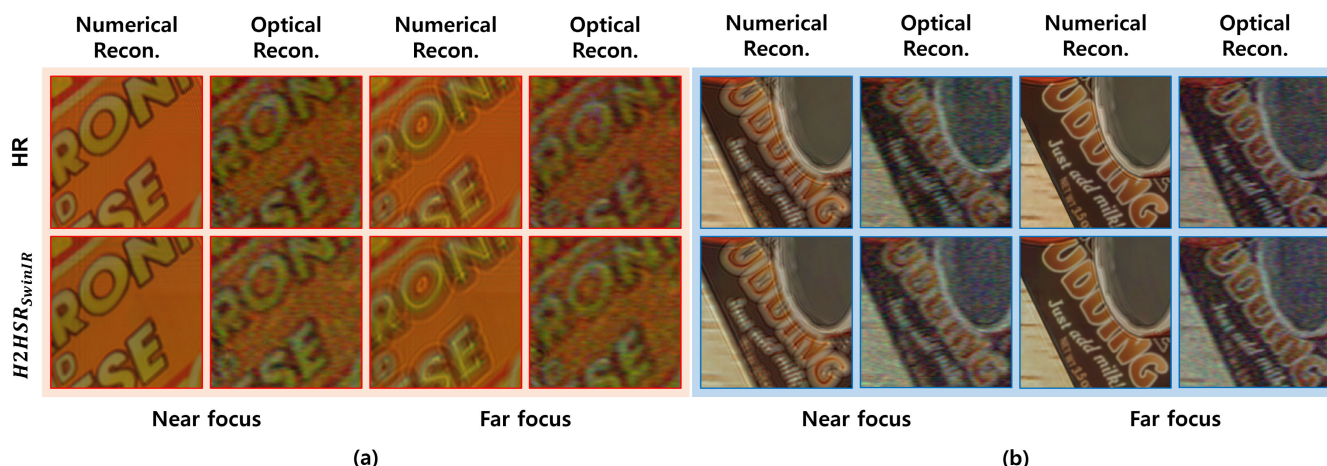

FIGURE 8. Comparison between numerical and optical reconstruction results. The first row shows the reconstruction result of the HR hologram, while the second row displays the result of the super-resolved hologram from our $H2HSR_{SwinIR}$. (a) and (b) are the zoomed regions of the red boxes in Fig. 7. The near focus refers to the focal distance being near the red box, while the far focus refers to the focal distance being near the blue box.

(FISBA RGBeam) with a maximum output power of 70 mW and the wavelengths for red, green, and blue are 638 nm, 520 nm, and 450 nm, respectively. The coherent light from the laser is converted into a plane wave by a lens (Lens1), and it reaches the SLM by passing through a polarizer. The modulated beam from the SLM is directed into the 4F system, which consists of two lenses (Lens2 and Lens3), by the beam splitter (BS). In the frequency plane of the 4F system, the aperture filters out high-frequency components to meet the requirements of DPAC and ensure high-quality optical reconstruction. Finally, the reconstructed scene is captured by a camera (FLIR Grasshopper3 USB Color, CCD) through the 4F system. Full-color optical reconstruction is achieved by merging the results of the red, green, and blue channels, which are captured separately.

For the optical experiment, we generated a hologram as an input to our network since the pixel pitch of the SLM (6.4 μm) differs from that of the MIT-CGH-4K dataset, which has 16 μm for low-resolution and 8 μm for high-resolution holograms. For hologram generation, we employed the point-based method and the wavefront recording plane (mid-point hologram), using objects from the HOPE dataset [64]. The left-most column in Fig. 7 shows the generated LR hologram. For the LR hologram, we set the resolution and pixel pitch to 540 × 540 and 12.8 μm, respectively, to match the SLM's specifications for the super-resolution results. To validate the optical reconstruction result, we also generated a HR hologram for the same scene as the LR hologram. To fit the resolution to the SLM, both the HR and SR holograms were revised by adding zero-padding, increasing the resolution from 1080 × 1080 to 1920 × 1080.

Fig. 7 shows the optical reconstruction results, captured by a camera, focusing around the red (near focus) and blue (far focus) boxes. Fig. 8 displays the zoomed images of the red and blue boxes and compares them with the numerical reconstruction results. Similar to the high similarity observed between HR and SR in numerical reconstruction, the optical reconstruction results of HR and SR also demonstrate a high degree of similarity. We found that the SR hologram accurately displays details across varying focal lengths. For

example, letters in the focusing region are clear, while those at different depths appear out-of-focus, as seen in Fig. 8.

It should be noted that our H2HSR network successfully inferred the SR hologram without depth distortion, even though the pixel pitch of the holograms used in the optical experiment (12.8 μm and 6.4 μm for LR and HR, respectively) is different from the pixel pitch of the training data (16 μm and 8 μm for LR and HR, respectively). These results demonstrate the potential of our H2HSR method to be applied to holograms with non-trained pixel pitches.

V. CONCLUSION AND FUTURE WORK
In this study, we introduced H2HSR, a novel hologram super-resolution method harnessing the power of deep learning to enhance the quality of reconstructed holographic images. Our H2HSR framework is built upon an encoder-decoder architecture, with a focus on designing a simple yet efficient up-sampling block in the decoder. The encoder can seamlessly integrate with various backbone models, offering flexibility and adaptability to different hologram super-resolution tasks. A fundamental aspect of our approach lies in the incorporation of perceptual loss, which complements the conventional data fidelity loss. Unlike methods relying solely on MSE loss, H2HSR's adaptive phase adjustment, guided by perceptual loss, closely aligns with the characteristics of real-world holograms. This holistic approach contributes to the overall enhancement of reconstructed image quality.

Our comprehensive evaluation reveals the superior performance of H2HSR when compared to conventional interpolation techniques and a prior GAN-based approach. Particularly noteworthy is the exceptional performance of H2HSR with the SwinIR encoder, achieving an impressive PSNR of 32.85 and SSIM of 0.91 in terms of reconstruction quality. This represents a significant improvement of 8.46% and 9.30%, respectively, over the prior GAN-based method.

These findings underscore the substantial contributions of H2HSR to the field of hologram super-resolution. Beyond superior quantitative metrics, our approach offers a promising pathway to efficiently adapt given holograms to suit specific hologram display requirements. As holographic imaging continues to advance and find applications across various domains, we anticipate that H2HSR will play a pivotal role in shaping the future of high-quality holographic displays and immersive experiences.

A. FUTURE WORK
While our super-resolution method effectively reduces pixel pitch, the impact of increasing the Angle of View (AoV) remains challenging to quantify despite our diligent efforts. The MIT-CGH-4K dataset, designed primarily for RGB-D to hologram conversion, provides holograms with limited resolutions (e.g., 192-by-192) and depth volumes, making it inadequate for thoroughly evaluating the effects of pixel pitch alterations. To address this limitation, our future work will concentrate on the generation of a comprehensive full-parallax hologram dataset. This dataset will encompass

holograms with varying pixel pitches and higher resolutions, facilitating the exploration of pixel pitch adjustments' true implications. Additionally, with the introduction of this new dataset, we aim to design a hologram resolution adaptation network. This network will not only support super-resolution but also down-resolution tasks, broadening its applicability to diverse scenarios.

H2HSR uses inputs of amplitude and phase derived from a complex-valued matrix according to Euler's formula. The network is inclined to treat the hologram as a real-valued image, common in image processing tasks. To effectively train using holograms, converting the hologram plane to a mid-point hologram is crucial, enabling the amplitude to resemble a clearer image. However, if the hologram plane is distant from the three-dimensional scene, the network's performance diminishes, unlike current methods that utilize mid-point holograms for training. To overcome this, the network should be designed to directly train complex-valued holograms without converting them into amplitude and phase or diffracting into a mid-point. Developing a specialized network for complex holograms represents our future direction, aiming to deepen our understanding and enhance our manipulation of holograms.

DISCLOSURES
The authors declare no conflicts of interest.

DATA AVAILABILITY
Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

SUPPLEMENTAL DOCUMENT
See Supplement 1 for supporting content.

REFERENCES
[1] C. Slinger, C. Cameron, and M. Stanley, "Computer-generated holography as a generic display technology," *Computer*, vol. 38, no. 8, pp. 46–53, Aug. 2005.
[2] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Appl. Opt.*, vol. 57, no. 14, p. 3859, 2018.
[3] M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, "DeepCGH: 3D computer-generated holography using deep learning," *Opt. Exp.*, vol. 28, no. 18, p. 26636, 2020.
[4] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature*, vol. 591, no. 7849, pp. 234–239, Mar. 2021.
[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[7] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.
[8] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.
[9] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[10] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[11] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[12] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," 2022, *arXiv:2205.04437*.

[13] M. Jee, H. Kim, M. Yoon, and C. Kim, "Hologram super-resolution using dual-generator GAN," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2596–2600.

[14] K. Matsushima, *Introduction to Computer Holography: Creating Computer-Generated Holograms as the Ultimate 3D Image*. Cham, Switzerland: Springer, 2020.

[15] S.-B. Ko and J.-H. Park, "Speckle reduction using angular spectrum interleaving for triangular mesh based computer generated hologram," *Opt. Exp.*, vol. 25, no. 24, pp. 29788–29797, 2017.

[16] M. Askari, S.-B. Kim, K.-S. Shin, S.-B. Ko, S.-H. Kim, D.-Y. Park, Y.-G. Ju, and J.-H. Park, "Occlusion handling using angular spectrum convolution in fully analytical mesh based computer generated hologram," *Opt. Exp.*, vol. 25, no. 21, p. 25867, 2017.

[17] H. Yeom, S. Cheon, K. Choi, and J. Park, "Efficient mesh-based realistic computer-generated hologram synthesis with polygon resolution adjustment," *ETRI J.*, vol. 44, no. 1, pp. 85–93, Feb. 2022.

[18] Y. Zhao, L. Cao, H. Zhang, D. Kong, and G. Jin, "Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method," *Opt. Exp.*, vol. 23, no. 20, p. 25440, 2015.

[19] H. G. Kim and Y. Man Ro, "Ultrafast layer based computer-generated hologram calculation with sparse template holographic fringe pattern for 3-D object," *Opt. Exp.*, vol. 25, no. 24, p. 30418, 2017.

[20] T. Shimobaba, N. Masuda, and T. Ito, "Simple and fast calculation algorithm for computer-generated hologram with wavefront recording plane," *Opt. Lett.*, vol. 34, no. 20, p. 3133, 2009.

[21] X. Jin, J. Gui, Z. Jiang, G. Wang, and Y. Lou, "Fast calculation of computer generated hologram using multi-core cpus and gpu system," *Proc. SPIE*, vol. 10818, pp. 310–316, Nov. 2018.

[22] B. J. Jackin, S. Watanabe, K. Ootsu, T. Ohkawa, T. Yokota, Y. Hayasaki, T. Yatagai, and T. Baba, "Decomposition method for fast computation of gigapixel-sized Fresnel holograms on a graphics processing unit cluster," *Appl. Opt.*, vol. 57, no. 12, p. 3134, 2018.

[23] H. Kang, J. Lee, and D. Kim, "HI-FFT: Heterogeneous parallel in-place algorithm for large-scale 2D-FFT," *IEEE Access*, vol. 9, pp. 120261–120273, 2021.

[24] J. Lee, H. Kang, H.-J. Yeom, S. Cheon, J. Park, and D. Kim, "Out-of-core GPU 2D-shift-FFT algorithm for ultra-high-resolution hologram generation," *Opt. Exp.*, vol. 29, no. 12, p. 19094, 2021.

[25] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of plane from image and diffraction pictures," *Optik*, vol. 35, no. 2, pp. 237–246, Sep. 1972.

[26] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.

[27] Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–8, Sep. 2019.

[28] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–14, Dec. 2020.

[29] L. Shi, B. Li, and W. Matusik, "End-to-end learning of 3D phase-only holograms for holographic display," *Light, Sci. Appl.*, vol. 11, no. 1, pp. 1–18, Aug. 2022.

[30] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[31] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

[32] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[33] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 814–81409.

[34] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 257–265.

[35] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[36] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[39] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[40] N. Siegel and G. Brooker, "Single shot holographic super-resolution microscopy," *Opt. Exp.*, vol. 29, no. 11, p. 15953, 2021.

[41] J. Chen, Z. Fu, B. Chen, and S.-C. Chen, "Fast 3D super-resolution imaging using a digital micromirror device and binary holography," *J. Biomed. Opt.*, vol. 26, no. 11, Nov. 2021, Art. no. 116502.

[42] K. Samanta, A. Tiwari, P. Samsheerali, and J. Joseph, "Improving resolution in two orthogonal orientations from a single-shot digital holographic microscopy," *Results Opt.*, vol. 14, Feb. 2024, Art. no. 100586.

[43] W. Bishara, T.-W. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution," *Opt. Exp.*, vol. 18, no. 11, p. 11181, 2010.

[44] J. Zhang, J. Sun, Q. Chen, J. Li, and C. Zuo, "Adaptive pixel-super-resolved lensfree in-line digital holography for wide-field on-chip microscopy," *Sci. Rep.*, vol. 7, no. 1, pp. 1–15, Sep. 2017.

[45] C. Fournier, F. Jolivet, L. Denis, N. Verrier, E. Thiebaut, C. Allier, and T. Fournel, "Pixel super-resolution in digital holography by regularized reconstruction," *Appl. Opt.*, vol. 56, no. 1, p. 69, 2017.

[46] Y. Rivenson, H. C. Koydemir, H. Wang, Z. Wei, Z. Ren, H. Günaydın, Y. Zhang, Z. Göröcs, K. Liang, D. Tseng, and A. Ozcan, "Deep learning enhanced mobile-phone microscopy," *ACS Photon.*, vol. 5, no. 6, pp. 2354–2364, Jun. 2018.

[47] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 460–468, May 2018.

[48] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, p. 458, 2018.

[49] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, Jan. 2019.

[50] T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Mar. 2019.

[51] W. Kim, B.-S. Park, J.-K. Kim, K.-J. Oh, J.-W. Kim, D.-W. Kim, and Y.-H. Seo, "Deep learning-based super resolution for phase-only holograms," *J. Broadcast Eng.*, vol. 25, no. 6, pp. 935–943, 2020.

[52] S.-C. Liu and D. Chu, "Deep learning for hologram generation," *Opt. Exp.*, vol. 29, no. 17, p. 27373, 2021.

[53] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[54] A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, "Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize," 2017, *arXiv:1707.02937*.

[55] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.

[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[57] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 252–268.

[58] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[59] K. Matsushima and T. Shimobaba, "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," *Opt. Exp.*, vol. 17, no. 22, p. 19662, 2009.

[60] D.-Y. Park and J.-H. Park, "Generation of distortion-free scaled holograms using light field data conversion," *Opt. Exp.*, vol. 29, no. 1, p. 487, 2021.

[61] K. Oh, J. Kim, and H. Y. Kim, "A new objective quality metric for phase hologram processing," *ETRI J.*, vol. 44, no. 1, pp. 94–104, Feb. 2022.

[62] C. Chen, B. Lee, N.-N. Li, M. Chae, D. Wang, Q.-H. Wang, and B. Lee, "Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function," *Opt. Exp.*, vol. 29, no. 10, p. 15089, May 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-10-15089

[63] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–16, Aug. 2017.

[64] Y. Lin, J. Tremblay, S. Tyree, P. A. Vela, and S. Birchfield, "Multi-view fusion for multi-level robotic scene understanding," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 6817–6824.

HANJU YEOM received the B.S. and M.S. degrees in information and communication engineering from Inha University, Incheon, Republic of Korea, in 2015 and 2017, respectively. From 2017 to 2019, he was a Researcher with Samsung Display. Since 2019, he has been a Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea. His research interests include digital holography and computer-generated holography.

YOUCHAN NO received the B.S. and M.S. degrees from Korea University of Technology and Education (KOREATECH), in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree with the Department of Computer Engineering. His research interests include deep learning for vision tasks and computer-generated holography.

SUNGMIN KWON (Student Member, IEEE) is currently pursuing the B.S. degree with the Department of Computer Engineering, Korea University of Technology and Education (KOREATECH). His research interests include vision tasks and high-performance computing.

JAEHONG LEE received the B.S. and M.S. degrees from Korea University of Technology and Education (KOREATECH), in 2020 and 2022, respectively, where he is currently pursuing the Ph.D. degree with the Department of Computer Engineering. His research interests include computer-generated holography and high-performance computing.

DUKSU KIM (Member, IEEE) received the B.S. degree from Sungkyunkwan University, in 2008, and the Ph.D. degree in computer science from Korea Advanced Institute of Science and Technology (KAIST), in 2014. He is currently an Associate Professor with the Department of Computer Engineering, Korea University of Technology and Education (KOREATECH). He spent several years as a Senior Researcher with the KISTI National Supercomputing Center. His research interests include high-performance computing, computer-generated holography, and deep learning. He is a professional member of ACM.

● ● ●