Continuous Depth Control of Phase-Only Hologram With Depth Embedding Block

Won Jong Ryu, Jin Su Lee, and Yong Hyub Won

***Abstract*—Digital holography is a promising candidate for advanced display, although several obstacles remain, such as the problem of heavy time consumption in the generation of phase-only holograms. Recently, deep-learning-based methods have achieved the real-time generation of holograms while maintaining high image quality. However, the holograms created with deep neural networks can reproduce images only at a specific distance because their target depth is fixed in the training process. This paper suggested and demonstrated a deep neural network that can continuously control the depth of the phase-only hologram. The network takes a target depth and an input image and generates a phase-only hologram. We added a depth embedding block that moves the hologram latent vector depending on the target depth. Thus, we can change the location of the image plane without retraining. The numerical and optical experiments show that the network understands the relationship between the depth and the appearance of the phase-only hologram. As a result, phase-only holograms generated with the proposed network can reconstruct images with around 25-dB PSNR.**

***Index Terms*—Holography, phase-only hologram, deep neural network, depth embedding block, hologram latent space.**

I. Introduction

DIGITAL holographic display is of great interest for next-generation displays to implement natural 3D images. The digital holographic display can reproduce realistic 3D objects at the desired location by adjusting light's amplitude and phase [1]. However, the current technologies are unsuitable for commercialization because of several problems, such as strong noises [2]–[6], heavy time consumption [7], [8], and a narrow field of view [9], [10]. Furthermore, a spatial light modulator (SLM), the main component in a digital holographic display, can only control either amplitude or phase. When using a phase-only SLM, the amplitude of each pixel is fixed to a uniform constant. Therefore, a complex-amplitude hologram must be converted into phase-only or amplitude-only hologram.

Manuscript received January 25, 2022; revised March 4, 2022; accepted March 17, 2022. Date of publication March 22, 2022; date of current version April 12, 2022. This work was supported in part by the National Research Foundation of Korea grant funded by the Korea government under Grant NRF2020R1F1A107408912 and in part by BK21 FOUR Program. *(Corresponding author: Jin Su Lee.)*

Won Jong Ryu and Yong Hyub Won are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea (e-mail: wonjong@kaist.ac.kr; yhwon@kaist.ac.kr).

Jin Su Lee is with the Spatial Optical Information Research Center, Korea Photonics Technology Institute, Gwangju 61007, Republic of Korea (e-mail: eventually27@kopti.re.kr).

Digital Object Identifier 10.1109/JPHOT.2022.3161225

Iterative phase retrieval algorithms are commonly used to convert a complex-amplitude hologram into a phase-only hologram [11]–[13]. However, because speed and accuracy are trade-offs, it is not easy to generate high-quality holograms in real-time (30 frames per second). Moreover, the computation time increases exponentially as the hologram resolution increases.

Recently, deep learning has been applied to hologram generation to overcome this problem. Deep learning approximates a complex function with a simple but large number of calculations. Several neural networks have been designed to generate holograms as follows. Horisaki *et al.* suggested a deep learning model for generating a phase-only hologram [14]. The proposed neural network is based on a U-net [15] with supervised learning. Eybposh *et al.* proposed a training strategy of the self-supervised method [16]. Aamir *et al.* proposed GAN-holo, constructed of a generator and a discriminator [17]. Peng *et al.* proposed HoloNet, which is composed of two sub neural networks [18]. They are a network that infers a wavefront of the target plane and a network that synthesizes a phase-only hologram from a complex-amplitude hologram.

Because the methods introduced above have a limitation that they target a specific single-depth, methods for targeting multi-depths have been proposed as follows. Eybposh *et al.* developed DeepCGH, a deep learning-based 3D computer-generated holography [19]. Lee *et al.* suggested a multi-depth hologram generation network (MDHGN). This network takes five images as an input and outputs an amplitude-only hologram [20]. Horisaki *et al.* introduced the 3D-CGH method with a priori knowledge of reproduced optical patterns based on deep learning [21]. These networks still treat the target depth as a constant parameter in the training process. In other words, images can be reconstructed at specific locations, and users should retrain the networks if they want to change the target depth.

Shi *et al.* proposed Tensor Holography synthesizing a photorealistic 3D hologram from a single RGB-depth image [22]. Tensor Holography requires a depth estimation map as well as an RGB image and outputs a complex-amplitude hologram. The complex-amplitude hologram is converted into the phase-only hologram, and the captured scene with quasi-continuous depth is well reconstructed with a high peak-to-noise ratio (PSNR). It is possible to change the focal plane, but it is hard to change the location of the object in the image by adjusting the depth map. We suggest a neural network where we can position each object directly.

This paper suggested and demonstrated a neural network that generates a phase-only hologram corresponding to the input

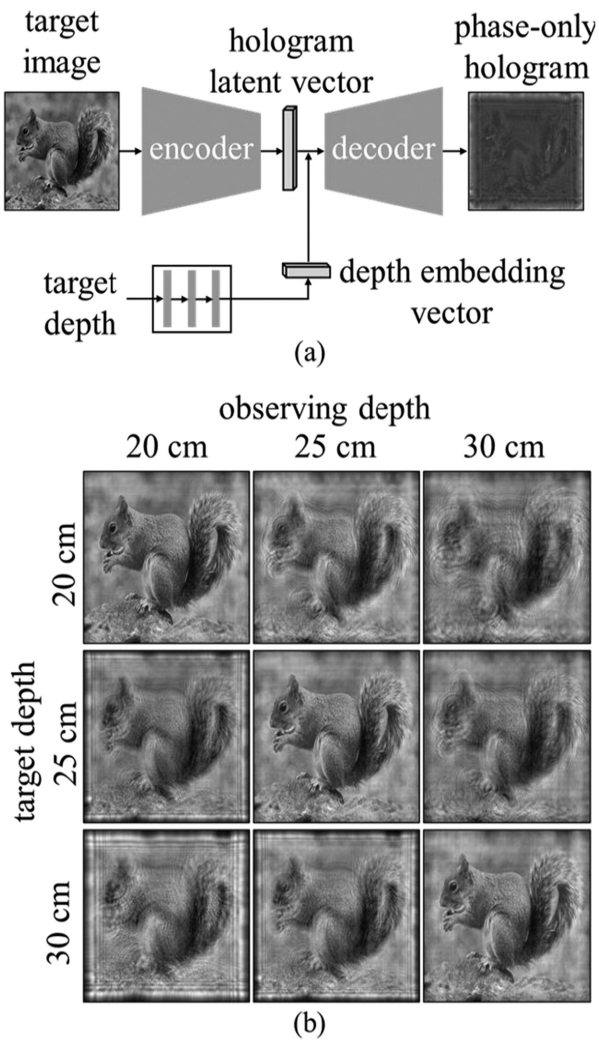

Fig. 1. (a) The proposed network generates a phase-only hologram corresponding to the target depth. (b) Reconstructed images by varying the observing depth. The phase-only holograms are generated from a single neural network without retraining.

target depth. Continuous depth phase-only hologram generation network (CDHGN) takes a target depth as an input and outputs a phase-only hologram according to the depth. The depth enters a depth embedding block that converts a target depth into a depth embedding vector, as shown in Fig. 1(a). The depth embedding vector is added to the hologram latent vector of the target image.

We have brought a simple experimental result to explain the concept of the proposed network, as shown in Fig. 1(b). In this case, we trained the network with the depth ranging from 20 cm to 30 cm, and phase-only holograms were generated from the network without retraining. The phase-only holograms reconstructed images by varying the observing depths. They reconstructed high-quality images if the observing depth is the same as the target depth. For example, the phase-only hologram generated with the target depth of 20 cm well reconstructed the image only at the observing depth of 20 cm.

We compared three types of network structure as depicted in Fig. 2. We chose a network structure that generates high-quality phase-only holograms with a relatively small number of

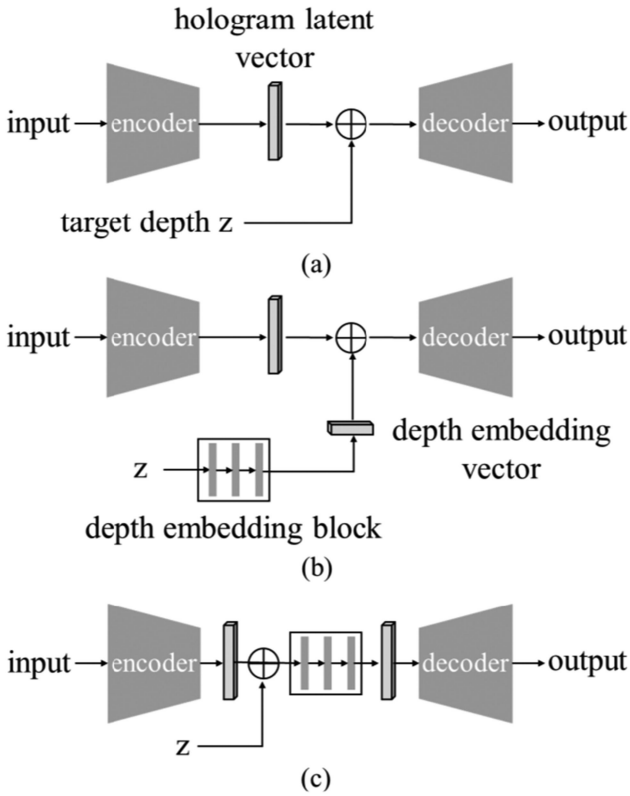

Fig. 2. Different methods to inject the depth information into the network. (a) A target depth is added to the hologram latent vector in net A while (b) Net B and (c) Net C exploit a depth embedding block.

parameters in a short time. We also evaluated the generalization ability for depth and verified the network works fine for depths not used in training. In other words, the network understands the relationship between a depth and a phase-only hologram. However, this work still has a limitation in that the network is only working for a single plane image. Extending a phase-only hologram to a 3D holographic display is still challenging.

The proposed network, including the architecture, the dataset, and training details, will be described in Section II. The result of numerical and optical experiments will be reported in Section III. In detail, we first figured out how to inject depth information into the network. Next, we tested the generalization ability for depth. We experimentally determined the ratios of error metrics and compared the quality of the reconstructed image with other methods. Finally, a conclusion will follow in the final section.

II. Overview of CDHGN

A. Network Architecture

The proposed network consists of an encoder, a decoder, and a depth embedding block. The structures of the encoder and the decoder refer to HoloNet [18], a state-of-the-art deep learning model generating a phase-only hologram. The HoloNet consists of two subnetworks. They both have the same structure but different roles. The first subnetwork creates an initial phase corresponding to a target image. Then the phase is combined

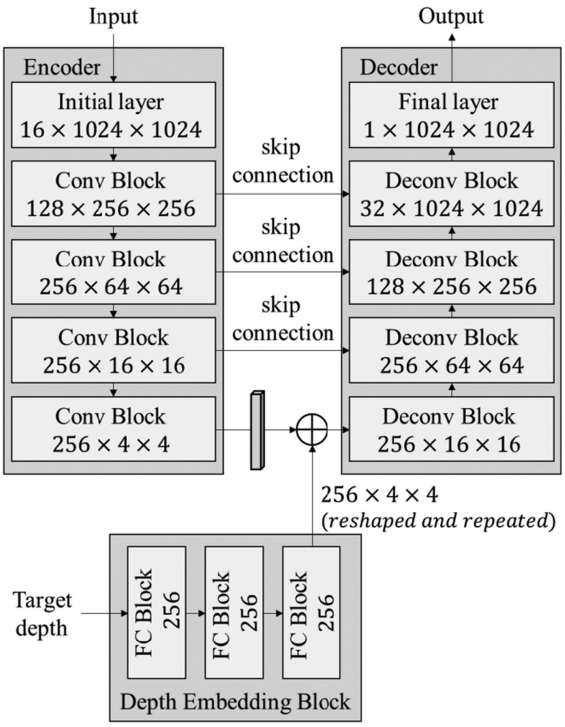

Fig. 3. Detail structures of net B. An output shape of each layer or block is written in the box.

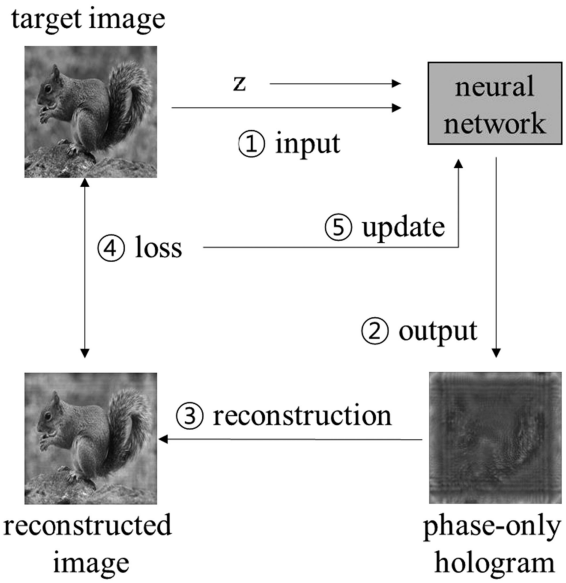

Fig. 4. The training process of the proposed neural network repeats this sequence by changing target images and target depths.

with the target image and numerically calculated as if the light propagates to the SLM plane. The complex-amplitude hologram in the SLM plane enters the second subnetwork. The second subnetwork outputs a final phase which can be reconstructed to the target image.

As shown in Fig. 2, we modified the network architecture to reflect the depth information. We experimented with how the performance varies depending on adding the target depth to the network. As shown in Fig. 2(a), target depth z is added to the hologram latent vector without any processing. Otherwise, as shown in Fig. 2(b), the target depth passes through the depth embedding block. The depth embedding block takes the target depth and returns a depth embedding vector. As shown in Fig. 3, the block is a sequence of a fully connected layer and a rectified linear unit. Each fully connected layer has 256 nodes, and the size of the depth embedding vector is repeated to $4 \times 4 \times 256$ before it is added to the hologram latent vector whose size is $4 \times 4 \times 256$. In the case of net C, as shown in Fig. 2(c), the hologram latent vector enters the depth embedding block after the target depth is added to the hologram latent vector. Fully connected layers in the depth embedding block of net C have 4096 nodes. Note that this network architecture change is applied to the two subnetworks.

As mentioned above, the target depth is added to the hologram latent vector without any preprocessing in net A. Therefore, the scale of the parameters in the hologram latent vector may not match the target depth. Moreover, the depth embedding block, in net B and net C, can map the target depth to the probability distribution, which has an advantage in the training.

Fig. 3 shows the detailed structure of net B. An output shape of each layer or block is written in the box. Conv Block consists of a convolutional layer, a rectified linear unit, and batch normalization. Deconv Blocks consist of a transposed convolutional network, a rectified linear unit, and batch normalization. Deconv Blocks take input from two ways by concatenating them along the channel axis. FC Block consists of a fully connected layer and a rectified linear unit. In net B, the output shape of the FC block is 256, and it is reshaped and repeated to $4 \times 4 \times 256$. In net C, the depth embedding block is adopted after adding the target to the hologram latent vector. The number of nodes is 4096 and it is reshaped to $4 \times 4 \times 256$ to match the size.

B. Training and Test

Neural networks require a training process to understand the input and output relationship. When the training is just beginning, the output of the network is far from a phase-only hologram. As the training progresses, the network learns how to extract features from the input and how to make the phase-only hologram from the features. Although the training takes a long time, the trained network can generate a phase-only hologram in a short time.

The process of training is presented in Fig. 4. First, a target image and a target depth enter the neural network. The target depth is randomly sampled in the predefined depth range. The neural network outputs a phase-only hologram, and the hologram is reconstructed to the image as if the light propagates to the image plane. After that, the error between the reconstructed image and the target image is computed with some error metrics such as a mean square error (MSE) and a perceptual loss function. Finally, the network is modified by analyzing the error. This sequence is repeated by changing the target images.

The test process is similar to the training process but does not require updates to the network. The reconstruction process

is performed by either numerical or optical propagation. In the optical experiment, no further calculations are required at this stage during application to the holographic display system.

At the reconstruction step, we use the Fresnel diffraction model to propagate from an origin plane to a destination plane:

$$U_2 = F^{-1}\left\{F\left\{U_1\left(x,y\right)\right\}H\left(f_x,f_y\right)\right\} \tag{1}$$

Here, U_1 and U_2 are an origin plane and a destination plane, respectively; $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and the inverse Fourier transform, respectively; and H is the transfer function given as:

$$H\left(f_x,\ f_y\right) = \exp\left(-j\pi\lambda z\left(f_x^2+f_y^2\right)\right) \tag{2}$$

where k is the wavenumber, z is the depth, λ is the wavelength of light, and f_X, f_Y are the coordinates of the frequency domain.

C. Training Details

The training dataset is a significant element in determining the generalization ability and the performance of the network. The dataset must be large enough and must include various images for the network to be applied to all images in general. Overfitting occurs if the dataset is small or skewed. The network then loses its generality and cannot process unseen images. In previous research, Lee *et al.* used images of randomly combined dots and circles [20], Eybposh *et al.* used datasets of 2D images with a random number of disks, squares, or lines [19]. Peng *et al.* used the DIV2K dataset [18], a widely-used open dataset consisting of high-resolution images [23]. We exploited the DiV2K dataset containing 1000 images. Among these, 800 images are used for training, and the remaining 200 images are used for testing.

The depth embedding block consists of fully connected layers and rectified linear units. The encoder and the decoder refer to HoloNet [18], but we set strides to 4, initially 2. We modified the loss ratio to 0.25, which is experimentally decided. Moreover, we did not use hardware optimization. We used the NVIDIA GeForece RTX Titan for training the network used in Figs. 6 and 8. NVIDIA GeForece RTX 3090 was used to train the networks used in other figures. While the maximum step is set to 300 steps, and one step is recorded for every 100 images processed. The training was ended within 6 hours.

III. Experimental Results

We verified the performance of the CDHGN with several tests. We evaluated the quality of reconstructed images with the peak signal-to-noise ratio (PSNR). We used 200 test data of the DiV2K dataset to evaluate the network performance, which is never used in the training process. Because we explained the image quality and the functionality of the network in Fig. 1(b), we show the change of the average PSNR according to the training steps or the target depth.

A. Network Architecture Test

The architecture of the network determines the complexity and capacity of the network. We exploited the HoloNet and modified it to inject the target depth into the network. As described above, three types of injecting methods are compared. In Fig. 2(a), a target depth is directly added to the hologram latent vector. In the case of net B and net C, the depth embedding block is used to preprocess the depth information. Net B takes the depth embedding vector, while net C takes the target depth and passes the hologram latent vector to the depth embedding block.

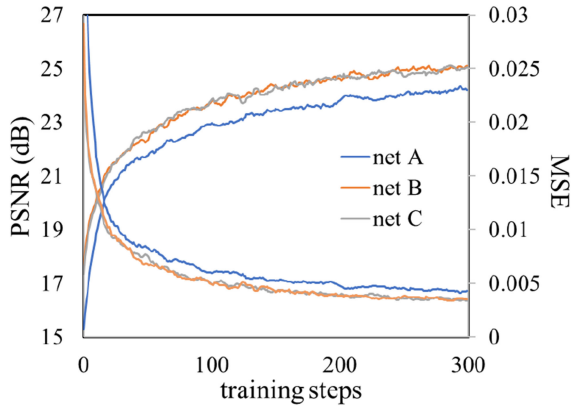

Fig. 5. The PSNR and MSE of the reconstructed images during the training. Net B and net C have a depth embedding block and higher performance than net A.

Fig. 5 shows the effect of the depth embedding block. The PSNR of the net A is much lower than the others. Net B and net C, the networks including the depth embedding block, show a similar result. However, net B is more efficient in terms of parameters. There are 256 nodes in the fully connected layer of net B, while there are 4096 nodes in net C. Therefore, net C uses more parameters than net B. Without considering the parameters of the encoder and the decoder, the depth embedding block of net B has 131328 parameters while net C has 68 billion parameters.

Furthermore, we test the computation time. It is a more direct way to compare the performance of net B and net C rather than the parameters needed for each structure. We measured the time to process 200 images using the GPU. Net A takes 31 ms per image, net B takes 35 ms per image, and net C takes 42 ms per image. On the other hand, a traditional non-network-based method such as Gerchberg–Saxton (GS) algorithm [11] and double phase-amplitude coding (DPAC) [24] takes 232 ms and 1315 ms, respectively. Net A, without the depth embedding block, satisfies the real-time condition, and the result of net B is also close to it. However, net C takes a relatively longer time than net A and net B. As a result, net B can reduce processing time and memory usage while maintaining high image quality. Note that the latter results are based on net B.

B. Depth Range Test

The generalization ability refers to how accurately the network can process unseen data, and we also evaluated the generalization ability for depth. We tried to input depths out of the depth range used for the training. Suppose a network trained in the range of 20 cm to 25 cm can make a high-quality phase-only hologram even when inputting 40 cm. It can be said that the network understands the relationship between a depth and a

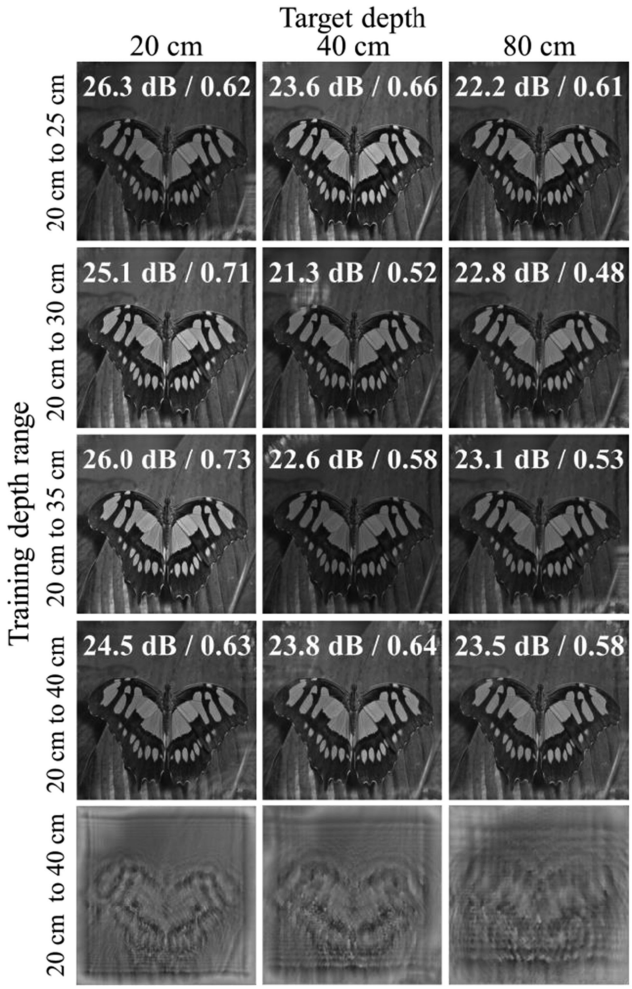

Fig. 6. The PSNR and the SSIM of the reconstructed images according to the target depth. The last row shows the phase-only hologram when the training depth range is 20 cm to 40 cm.

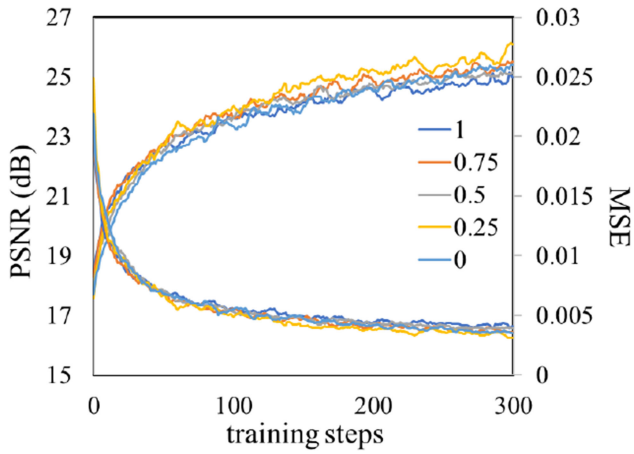

Fig. 7. The PSNR of the reconstructed images according to the ratio of the error metrics.

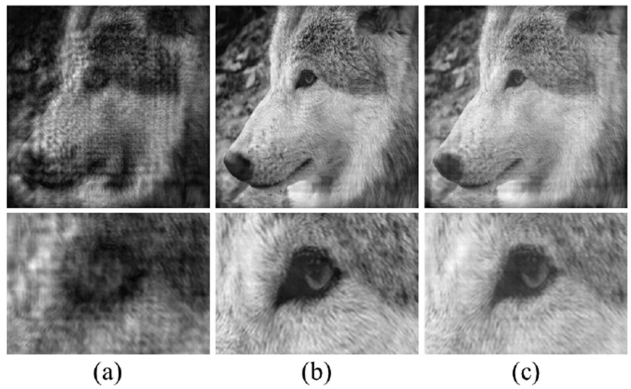

Fig. 8. Images in the first row are reconstructed from the phase-only hologram generated from three networks. Images in the second row are the enlarged image of them. They were trained with different error ratios. (a) Alpha is 0, (b) Alpha is 0.25, (c) Alpha is 1.

phase-only hologram. It means the deep neural network has the ability to generate phase-only holograms with continuous depth control.

The result of the depth range test is shown in Fig. 6. We evaluate the reconstructed image with the peak-to-noise ratio and the structural similarity (SSIM). The reconstructed images have PSNR over 20 dB even at depths outside the depth range used for training. The brightness of the image may have changed while saving the image to fit the 0–255 range. The last row shows the phase-only hologram is varying according to the target depth. Generating phase-only holograms for unlearned depths means the neural network works with depth as an understandable parameter.

C. Error Metric Test

Error metrics measure the difference between the target and the reconstructed images. The PSNR is a well-known error metric for estimating image quality. It is directly related to the mean square error (MSE), and the MSE is a convenient method to calculate errors. However, they have a limitation that their ability to capture texture detail is relatively low because they are defined based on a pixel-level image difference. Therefore, we use both mean square error and VGG-loss [25] to evaluate the reconstructed images. The MSE is used to compare pixel values, and the VGG-loss estimates feature-level errors. The ratio of the MSE and VGG-loss was experimentally decided as follows.

$$total\ loss = MSE\_loss + \alpha \times VGG\_loss \tag{3}$$

As shown in Fig. 7, the PSNR is the highest when alpha is 0.25. Even if alpha is changed, PSNR is similar, but the image quality differs. Fig. 8. shows the results of the networks trained with different alpha. The image quality is significantly low when only the MSE is used. This is why the perceptual loss, the VGG-loss, is necessary. Also, the pixel level loss, the MSE-loss, could harm the generalization ability of the network.

D. Comparing With Fixed Target Depth Network

We tested whether there is a disadvantage in adding the depth embedding block. Therefore, we removed the depth embedding block from our model and compared it with the original model. The network with a fixed distance was trained separately for

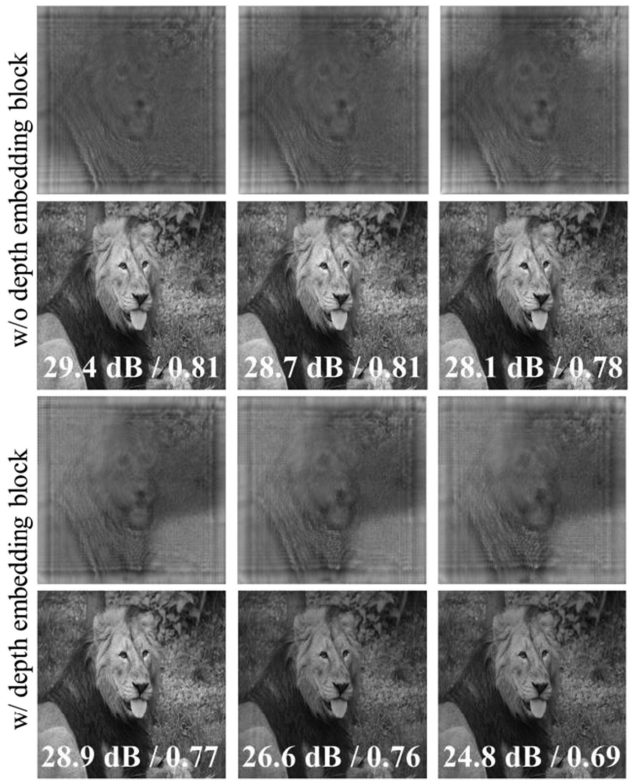

Fig. 9. Comparing the proposed network with a hologram generation network without the depth embedding block.

TABLE I
AVERAGE VALUES OF THE PSNR AND THE SSIM OF TEST SET

Target depth	20 cm	25 cm	30 cm
w/o depth embedding block	28.6 dB 0.79	28.2 dB 0.78	27.6dB 0.71
w/ depth embedding block	25.7 dB 0.73	25.4 dB 0.72	24.7dB 0.66

20 cm, 25 cm, and 30 cm as their fixed target depths. After generating a phase-only hologram, the image reconstructed at a fixed distance is shown in Fig. 9.

The proposed method can control the location of the image plane by inputting the target depth with the image. Thus, memory and training time can be saved, and the distance can be continuously adjusted without retraining. However, as shown in Fig. 9 and Table I, the PSNR and SSIM are lower than when the target distance is fixed.

Without regard to the presence or absence of the depth embedding block, the performance tends to decrease slightly as the distance increases. It seems that the diffraction pattern becomes slightly more complex as the distance increases.

E. Optical Experiment

Simple holographic display optics are set up on the optical table to verify the proposed idea practically, as shown in Fig. 10(a). A laser diode, which has a 520 nm wavelength, is combined with an x10 beam expander. The phase-only SLM is put in front of a

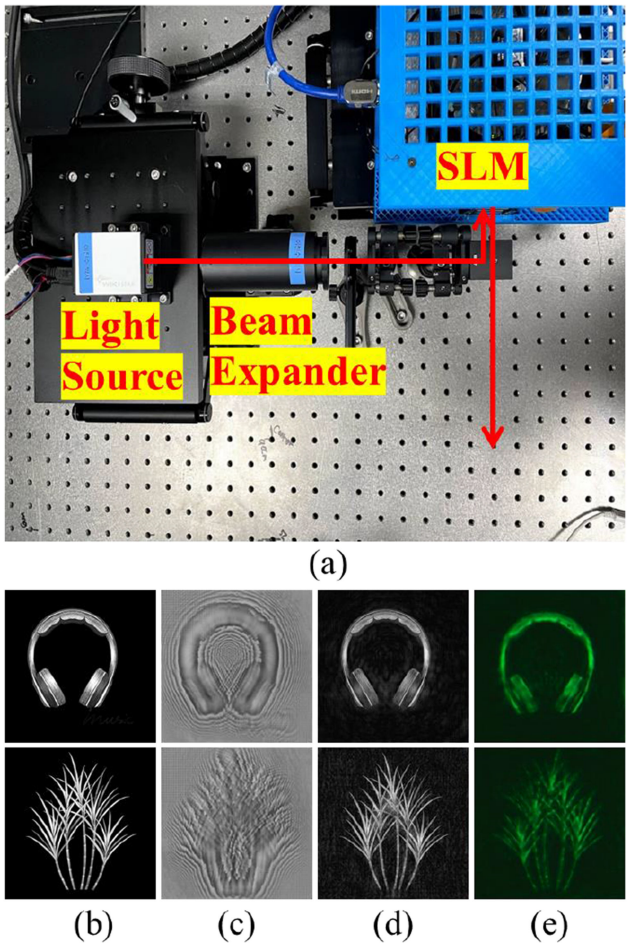

Fig. 10. (a) Holographic display setup, (b) Target images, (c) generated phase-only holograms, (d) Reconstructed images from the phase-only holograms, (e) Captured hologram.

cube beam splitter. It has 3.6 μm pixel pitch and 4K resolution and the number of phase levels of the spatial light modulator is 256. To figure out the strict performance of the proposed idea, additional optical components such as 4F optics, spatial filter, and eyepiece are not used in the experiment. Capturing device is Nikon A5000 which has 20M pixels.

Before the optical experiment, we generated phase-only holograms using the proposed network whit a target depth of 25 cm. At this time, the network is trained with a depth range from 20 cm to 30 cm. Target images and numerically reconstructed images are shown in Fig. 10(b) and (c). The images reconstructed through the optical system are shown in Fig. 10(e). Although the entire optical system was calibrated precisely, inevitable noises such as speckles, dc noise, and aberrations were added into the captured hologram. Nevertheless, captured holograms show sufficiently similar to simulated ones.

IV. CONCLUSION

This paper proposed a method for the continuous depth control of a phase-only hologram. We control the depth of the phase-only hologram by adding the depth embedding vector

to the hologram latent vector. As a result, a single neural network can generate phase-only holograms with continuous target depth. We demonstrated that target images from phase-only holograms are well reconstructed at the given target depth. Moreover, numerical experiments show the effect of the depth embedding bock and the generalization ability of this method. As a result, we can reproduce an image at the desired location in end-to-end processing. However, this work still has a limitation in that the network is only working for a single plane. Our future work is a one-shot generation of a phase hologram to reconstruct multi-plane images.

REFERENCES

[1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, pp. 777–778, May 1948, doi: 10.1038/161777a0.

[2] Y. Takaki and M. Yokouchi, "Speckle-free and grayscale hologram reconstruction using time-multiplexing technique," *Opt. Exp.*, vol. 19, no. 8, pp. 7567–7579, Apr. 2011, doi: 10.1364/OE.19.007567.

[3] Y. Mori, T. Fukuoka, and T. Nomura, "Speckle reduction in holographic projection by random pixel separation with time multiplexing," *Appl. Opt.*, vol. 53, no. 35, pp. 8182–8188, Dec. 2014, doi: 10.1364/AO.53.008182.

[4] L. Golan and S. Shoham, "Speckle elimination using shift-averaging in high-rate holographic projection," *Opt. Exp.*, vol. 17, no. 3, pp. 1330–1339, Feb. 2009, doi: 10.1364/OE.17.001330.

[5] A. Jesacher, S. Bernet, and M. Ritsch-Marte, "Broadband suppression of the zero-diffraction order of an SLM using its extended phase modulation range," *Opt. Exp.*, vol. 22, no. 14, pp. 17590–17599, Jul. 2014, doi: 10.1364/OE.22.017590.

[6] J. Liang, S. Y. Wu, F. K. Fatemi, and M. F. Becker, "Suppression of the zero-order diffracted beam from a pixelated spatial light modulator by phase compression," *Appl. Opt.*, vol. 51, no. 16, pp. 3294–3304, Jun. 2012, doi: 10.1364/AO.51.003294.

[7] T. Shimobaba, H. Nakayama, N. Masuda, and T. Ito, "Rapid calculation algorithm of fresnel computer-generated-hologram using look-up table and wavefront-recording plane methods for three-dimensional display," *Opt. Exp.*, vol. 18, no. 19, pp. 19504–19509, Sep. 2010, doi: 10.1364/OE.18.019504.

[8] D. Arai*et al.*, "Acceleration of computer-generated holograms using tilted wavefront recording plane method," *Opt. Exp.*, vol. 23, no. 2, pp. 1740–1747, Jan. 2015, doi: 10.1364/OE.23.001740.

[9] Z. Zeng, H. Zheng, Y. Yu, A. K. Asundi, and S. Valyukh, "Full-color holographic display with increased-viewing-angle," *Appl. Opt.*, vol. 56, no. 13, pp. F112–F120, May 2017, doi: 10.1364/AO.56.00F112.

[10] Y. Z. Liu, X. N. Pang, S. Jiang, and J. W. Dong, "Viewing-angle enlargement in holographic augmented reality using time division and spatial tiling," *Opt. Exp.*, vol. 21, no. 10, pp. 12068–12076, May 2013, doi: 10.1364/OE.21.012068.

[11] R. W. Gerchberg and W. O. Saxton, "Practical algorithm for determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[12] N. Yoshikawa and T. Yatagai, "Phase optimization of a kinoform by simulated annealing," *Appl. Opt.*, vol. 33, no. 5, pp. 863–868, Feb. 1994, doi: 10.1364/AO.33.000863.

[13] M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," *Appl. Opt.*, vol. 26, no. 14, pp. 2788–2798, Jul. 1987, doi: 10.1364/AO.26.002788.

[14] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Appl. Opt.*, vol. 57, no. 14, pp. 3859–3863, May 2018, doi: 10.1364/AO.57.003859.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV USA, 2016, pp. 770–778.

[16] M. H. Eybposh, N. W. Caira, P. Chakravarthula, M. Atisa, and N. C. Pégard, "High-speed computer-generated holography using convolutional neural networks," in *Proc. Biophotonics Congr.: Biomed. Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN)*, 20202, Paper BTu2C.2, doi: 10.1364/BRAIN.2020.BTu2C.2.

[17] K. Aamir, Z. Zhang, Y. Ying-jie, K. Muhammad, Y. Ketao, and A. Khizar, "GAN-holo: Generative adversarial networks-based generated holography using deep learning," *Complexity*, vol. 2021, Jan. 2021, Art. no. 6662161, doi: 10.1155/2021/6662161.

[18] Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, "Neural holography with camera-in-the-loop training," *ACM TOG*, vol. 39, no. 6, pp. 1–14, Nov. 2020, doi: 10.1145/3414685.3417802.

[19] 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, pp. 26636–26650, Aug. 2020, doi: 10.1364/OE.399624.

[20] J. Lee, J. Jeong, J. Cho, D. Yoo, B. Lee, and B. Lee, "Deep neural network for multi-depth hologram generation and its training strategy," *Opt. Exp.*, vol. 28, no. 18, pp. 27137–27154, Aug. 2020, doi: 10.1364/OE.402317.

[21] R. Horisaki, Y. Nishizaki, K. Kitaguchi, M. Saito, and J. Tanida, "Three-dimensional deeply generated holography," *Appl. Opt.,* vol. 60, no. 4, pp. A323–A328, Feb. 2021, doi: 10.1364/AO.404151.

[22] 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, doi: 10.1038/s41586-020-03152-0.

[23] R. Timofte *et al.*, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proc. CVPRW*, Honolulu, HI, USA, 2017, pp. 1110–1121.

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

[25] C. Ledig*et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.