



DOI: 10.29026/oea.2023.220135

4K-DMDNet: diffraction model-driven network for 4K computer-generated holography

Kexuan Liu, Jiachen Wu, Zehao He and Liangcai Cao*

Deep learning offers a novel opportunity to achieve both high-quality and high-speed computer-generated holography (CGH). Current data-driven deep learning algorithms face the challenge that the labeled training datasets limit the training performance and generalization. The model-driven deep learning introduces the diffraction model into the neural network. It eliminates the need for the labeled training dataset and has been extensively applied to hologram generation. However, the existing model-driven deep learning algorithms face the problem of insufficient constraints. In this study, we propose a model-driven neural network capable of high-fidelity 4K computer-generated hologram generation, called 4K Diffraction Model-driven Network (4K-DMDNet). The constraint of the reconstructed images in the frequency domain is strengthened. And a network structure that combines the residual method and sub-pixel convolution method is built, which effectively enhances the fitting ability of the network for inverse problems. The generalization of the 4K-DMDNet is demonstrated with binary, grayscale and 3D images. High-quality full-color optical reconstructions of the 4K holograms have been achieved at the wavelengths of 450 nm, 520 nm, and 638 nm.

Keywords: computer-generated holography; deep learning; model-driven neural network; sub-pixel convolution; oversampling

Liu KX, Wu JC, He ZH, Cao LC. 4K-DMDNet: diffraction model-driven network for 4K computer-generated holography. *Opto-Electron Adv* 6, 220135 (2023).

Introduction

Computer-generated holography is a technology that can accurately modulate the light field distribution in three-dimensional (3D) space. It has been widely applied in various fields, including holographic display^{1–4}, meta-surface design^{5–8}, and laser fabrication⁹. Currently, only a few spatial light modulators (SLMs) can modulate both the amplitude and phase of the wavefront simultaneously. The complex amplitude hologram (CAH) needs to be converted to an amplitude hologram or a phase-only hologram (POH) for display. Compared with the amplitude hologram, the POH provides higher diffraction efficiency and avoids the twin image in the recon-

struction. The POH is calculated only with the amplitude constraints on the object plane without any phase constraints. So the solution is not unique. And the amplitude constraints on the hologram plane are equal to 1, so the solution may not exist. The POH generation process is a typical ill-posed inverse problem. The algorithms used to fit this inverse problem are mostly iterative methods, such as Gerchberg–Saxton (GS) algorithm^{10,11}, Wirtinger algorithm¹², and non-convex optimization algorithm¹³. But they are time-consuming and only can find the local optimal solutions with speckle noise.

Since the high-performance network structure ResNet

State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instruments, Tsinghua University, Beijing 100084, China.

*Correspondence: LC Cao, E-mail: clc@tsinghua.edu.cn

Received: 2 August 2022; Accepted: 2 October 2022 ; Published online: 10 January 2023



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023. Published by Institute of Optics and Electronics, Chinese Academy of Sciences.

was proposed in 2016¹⁴, the powerful ability of deep learning on highly ill-posed inverse problems has been gradually demonstrated^{15–17}. Deep learning has been widely used for types of hologram generation since its parallel operational framework and complex structure based on convolution layers. Compared with the traditional POH algorithms^{4,18–21}, the learning-based POH has the great potential to realize real-time and speckle-free holographic display. Nowadays, the learning-based POH can be mainly divided into two types: the data-driven deep learning^{1,22–29} and the model-driven deep learning^{30–38}.

In the data-driven deep learning, the neural networks fit the inverse problem by learning the coding method from the approximate solutions calculated by traditional algorithms. The image datasets and their corresponding approximate POHs compose the labeled training datasets. The network is trained to extract the non-linear mapping between the labeled training dataset, which can be regarded as a black box. The training process is realized by calculating the loss function between the output POHs and the label POHs, and updating the network parameters by gradient descent algorithms. The trained data-driven network can effectively speed up the POH generation processes of images outside the training dataset^{1,22} and has the advantage of simple network structures. However, since the neural network with the data-driven deep learning purely learns the mapping of the image dataset and label POHs, the labeled training dataset quality limits the ceiling of the training performance and POH generalization. And the label POHs need to be generated in advance, which requires large datasets, huge computing resources, and long calculation time. The above two challenges are particularly prominent in the learning-based POH generation problem^{24–26}, which prohibits the practical application of the data-driven method. To avoid using iterative algorithms in the label POH generation process, random phase patterns and their propagating speckle-like intensity patterns are used for training²⁷. This method was also used for binary amplitude holograms²⁸. However, the trained networks work well only for the speckle-like patterns.

To solve the above two challenges, the model-driven deep learning is proposed for the POH generation^{30,31}. In the model-driven deep learning, the corresponding forward process model of the inverse problem is used as the constraints to train the networks. It can directly fit the inverse problem without the limitation of the approxi-

ate solutions. For POH generation, the physical diffraction model is incorporated into the network structure. The loss function can be directly calculated between the image dataset and the output reconstructions, which effectively eliminates the need for label POHs. The network can automatically learn the latent encodings of POHs in an unsupervised way. Current studies have successfully explored and demonstrated the availability of model-driven deep learning for various POH generation tasks. Peng et al. optimized the diffraction model by the camera-in-the-loop (CITL) strategies, which obtained speckle-free holographic images with coherent light sources^{32,33} and partially coherent light sources³⁴. Shimobaba et al. realized zoomable reconstruction larger than the holograms³⁵. Liu et al. proposed the phase dual-resolution network (PDRNet) structure to learn the mapping on the same optical plane rather than crossing optical planes³⁶. Sun et al. solved dual tasks of amplitude reconstruction and phase smoothing jointly by loss function optimization³⁷. In our previous research, the model-driven network Holo-Encoder could generate one single-wavelength 4K POH in 0.15 s³⁸. However, due to the insufficient constraints, the existing model-driven networks face the limited convergence result problem^{39,40}. The transfer learning with the single target image is needed for better reconstruction quality. There is a trade-off between the reconstruction quality and calculation speed, which limits the practical application. The combined-driven method, which combines the advantages of the data-driven and model-driven method, was proposed and achieved high-quality reconstructions of 3D objects⁴¹. But it still faces the time-consuming generation challenge of the labeled training dataset.

In this paper, we systematically investigate the existing learning-based POH research and especially analyze the advantages of the model-driven method over the data-driven method. We propose a high-fidelity 4K POH generation network, called 4K Diffraction Model-driven Network (4K-DMDNet). The constraint of the reconstructed images in the frequency domain is strengthened. The network structure combines the residual method and sub-pixel convolution method, which effectively enhances the fitting ability of the network for inverse problems. The generalization of the 4K-DMDNet is demonstrated with both binary and grayscale images, which can achieve high-fidelity and high-speed POH generation for 4K display.

POH generation network of 4K-DMDNet

Network architecture

Compared with the data-driven deep learning method, the 4K-DMDNet consists of the convolutional neural network (CNN) for POH generation and the diffraction model for loss function calculation, as shown in Fig. 1. The output of the neural network with the data-driven method is the predicted POHs. The loss function calculates the error between the output POHs and the label POHs, as shown in Fig. 1(a). Thus, it requires the preparation of a labeled training dataset consisting of the image dataset and its corresponding label POHs. However, the label POHs need to be generated by iterative algorithms, which is time-consuming and limits the ceiling of the training performance and generalization. In comparison, the proposed 4K-DMDNet works in an unsupervised way by incorporating the diffraction model as a part of the neural network, as is shown in Fig. 1(b). The diffraction model simulates the light field propagation process. Therefore, the loss function can be directly calculated between the image dataset and the output images. The latent POH encodings can be sought without the label POHs by 4K-DMDNet:

$$\begin{aligned} &\text{find } \mathbf{H} \\ &\text{s.t. } |\text{PROP}(\mathbf{H})|^2 = \mathbf{I}, \end{aligned} \quad (1)$$

where \mathbf{H} represents the POH, \mathbf{I} represents the target image, and PROP represents the propagation process from the hologram plane to the object plane. Moreover, 4K-DMDNet directly learns the optimal encoding POHs, avoiding the quality degradation caused by the extra complex-to-phase-only conversion operation. Since no random phase is added in the whole calculation process, the ubiquitous speckle noises are significantly suppressed.

Once the network training is complete, the network parameters can be solidified into the computer chip to realize the rapid generation of holograms. The generation and reconstruction process of 4K POHs by the 4K-DMDNet is shown in Fig. 2. A set of images or series of video frames outside the image dataset are input into the network in sequence. The trained network is used to predict the POHs of images. The output POHs are uploaded on the 4K SLM. To realize full-color optical reconstruction, the 4K-DMDNet is trained in three different wavelength versions corresponding to the three channels of the RGB image. The POHs corresponding to the three channels are loaded onto the SLM in turn, and three illumination lasers with different wavelengths are switched simultaneously. This time-multiplexing method can present high-quality color images without interference-induced noise. Finally, observers can see the

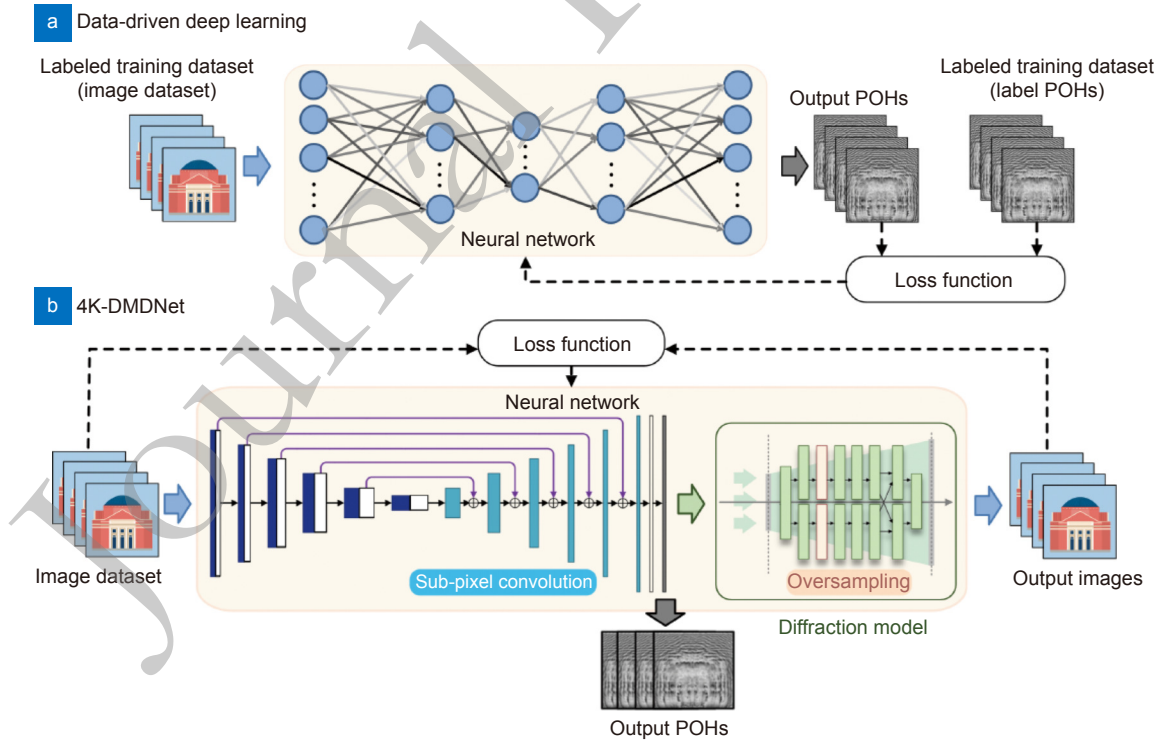


Fig. 1 | Training processes of (a) data-driven deep learning and (b) 4K-DMDNet, respectively.

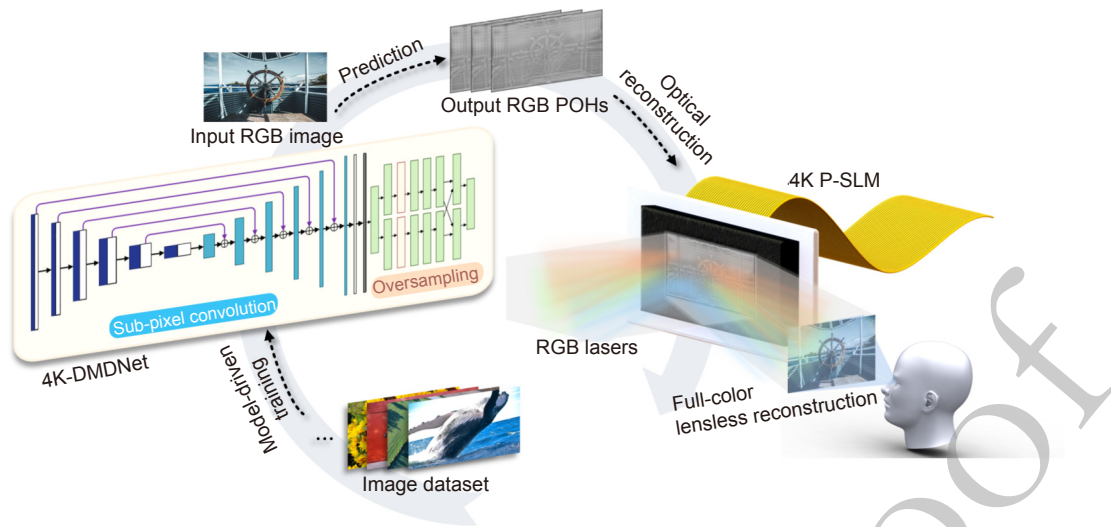


Fig. 2 | Generation and reconstruction process of 4K POHs by the 4K-DMDNet. The sub-pixel convolution method and oversampling method have played decisive roles to achieve it.

optically reconstructed image or video at the set distance.

The U-Net network as the CNN of 4K-DMDNet is shown in Fig. 3(a). It consists of a contracting down-sampling path to capture context and a symmetric expanding upsampling path that enables precise localization, which achieves excellent performance on different image-to-image problems⁴². And the skip connection from the downsampling path to the upsampling path is another feature of the U-Net, which makes the output images include more details. Here the downsampling path is a residual neural network, consisting of down-sampling blocks and corresponding residual blocks. Each block is composed of two sets of batch normalization, nonlinearity (ReLU), and a 3×3 convolutional layer stacked one above the other. And the residual block effectively solves the degradation problem with skip connections¹⁴. The downsampling is realized directly by convolutional layers with a stride of two. The upsampling path consists of upsampling blocks, as shown in Fig. 3(b). In order to achieve 4K hologram generation, the upsampling is realized by the sub-pixel convolution method⁴³. It includes convolutional layers to increase the channel number and a pixel shuffle layer to turn the tensor from $H \times W \times 2C$ to $2H \times 2W \times C/2$. And the residual method is also used. Finally, the output layer of the U-Net is a tanh function that limits the POH value in the range of $[-\pi, \pi]$.

Sub-pixel convolution method

The learnable parameter number presents the learning capability of network frameworks. For CNN, it is the

number and size of the convolution kernels. When the pending data exceeds the learning capability, the network will not converge stably. Increasing the number of convolutional layers is usually used for more learnable parameters. But the deeper the network, the more difficult the training. The upsampling method is another important factor which prominently affects the learnable parameter number. Here we use the sub-pixel convolution method to achieve the 4K hologram generation. It can increase the learnable parameters of the upsampling path by four times without changing the network depth. To highlight the strong learning capability of the sub-pixel convolution method, we compare it with the other two common upsampling methods, as shown in Fig. 4.

The task of the upsampling path is to upscale the low-resolution feature map gradually to the target size. The transposed convolution method, also called the deconvolution method, realizes the upsampling by zero-padding around the input layer to double size and then convoluting with a stride of one. It is the original upsampling method used in the U-Net network. However, its reconstructions face the “checkerboard artifacts” problem, caused by the uneven overlap in the convolution process⁴⁴, as shown in Fig. 4(a). The nearest neighbor resize convolution (NN-resize convolution) can effectively solve this problem. It replaces the zero-padding operation with the nearest neighbor interpolation, adding more valid information. The sub-pixel convolution further uses learnable parameters instead of interpolation information to enhance network performance. It includes a convolutional layer with a stride of one and four

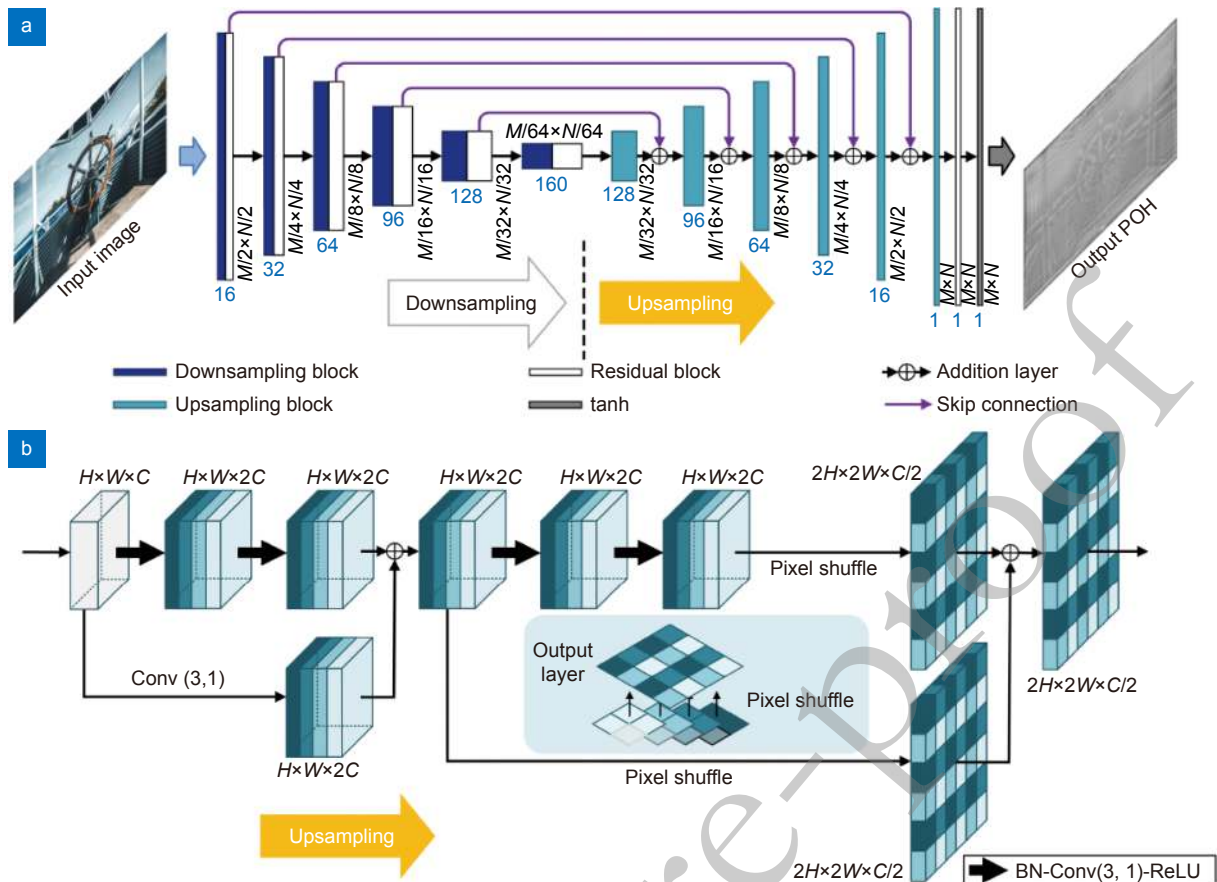


Fig. 3 | (a) U-Net neural network architecture of 4K-DMDNet. (b) Upsampling block architecture. The figures between the brackets present the kernel size and the stride of the convolutional layer, respectively.

times the original channel number, and a pixel shuffle layer to permute data L^i from the channel dimension into blocks of 2-D spatial data,

$$PS(L^i)_{h,w,c} = L^i_{\lfloor h/2 \rfloor, \lfloor w/2 \rfloor, c \cdot 2 \cdot \text{mod}(w,2) + c \cdot \text{mod}(h,2)} \cdot (2) \quad (2)$$

The peak signal-to-noise ratio (PSNR) is employed to quantitatively evaluate the reconstruction quality of the above three upsampling methods. We can see that the sub-pixel convolution method can achieve high-fidelity and artifact-free images compared with the other two methods. The PSNR of the reconstruction with the sub-pixel convolution method is 19.27 dB.

Diffraction model of 4K-DMDNet

Fresnel diffraction model in network layer manner

The Fresnel diffraction model is used as the diffraction model of 4K-DMDNet, which is advantageous for computational speed. It calculates the propagation process from the hologram plane to the object plane, as shown in Fig. 5(a). The intensity distribution on the object plane can be formulated as

$$\begin{aligned} \hat{I}(\mathbf{x}, \mathbf{y}) &= |\hat{C}(\mathbf{x}, \mathbf{y})|^2 = |\mathcal{F}\{\exp[i\Phi(\mathbf{x}_0, \mathbf{y}_0)]\}|^2 \\ &= |\mathcal{F}\{\exp[i\varphi(\mathbf{x}_0, \mathbf{y}_0)] \cdot \exp[i\frac{\pi}{\lambda d}(\mathbf{x}_0^2 + \mathbf{y}_0^2)]\}|^2, \end{aligned} \quad (3)$$

where $\mathbf{x}, \mathbf{y}, \mathbf{x}_0, \mathbf{y}_0$ represent the coordinates on the object plane and the hologram plane, respectively, $\hat{C}(\mathbf{x}, \mathbf{y})$ represents the complex amplitude distribution on the object plane, \mathcal{F} denotes the Fourier transform, $\varphi(\mathbf{x}_0, \mathbf{y}_0)$ is the output POH of the U-Net, λ is the wavelength of the light sources, and d is the distance between the two planes. By changing the parameters λ and d , full-color and multi-plane holographic displays can be obtained by the 4K-DMDNet. The calculation process of the Fresnel diffraction model is realized in the network layer manner, as shown in Fig. 5(b). Because the neural network can only backpropagate with real numbers, the light field distribution on the hologram plane is split into real and imaginary parts at the beginning according to the Euler's formula:

$$\exp[i\Phi(\mathbf{x}_0, \mathbf{y}_0)] = \cos[\Phi(\mathbf{x}_0, \mathbf{y}_0)] + i\sin[\Phi(\mathbf{x}_0, \mathbf{y}_0)]. \quad (4)$$

Since the FFT results of real numbers are complex numbers, they need to be split again according to the

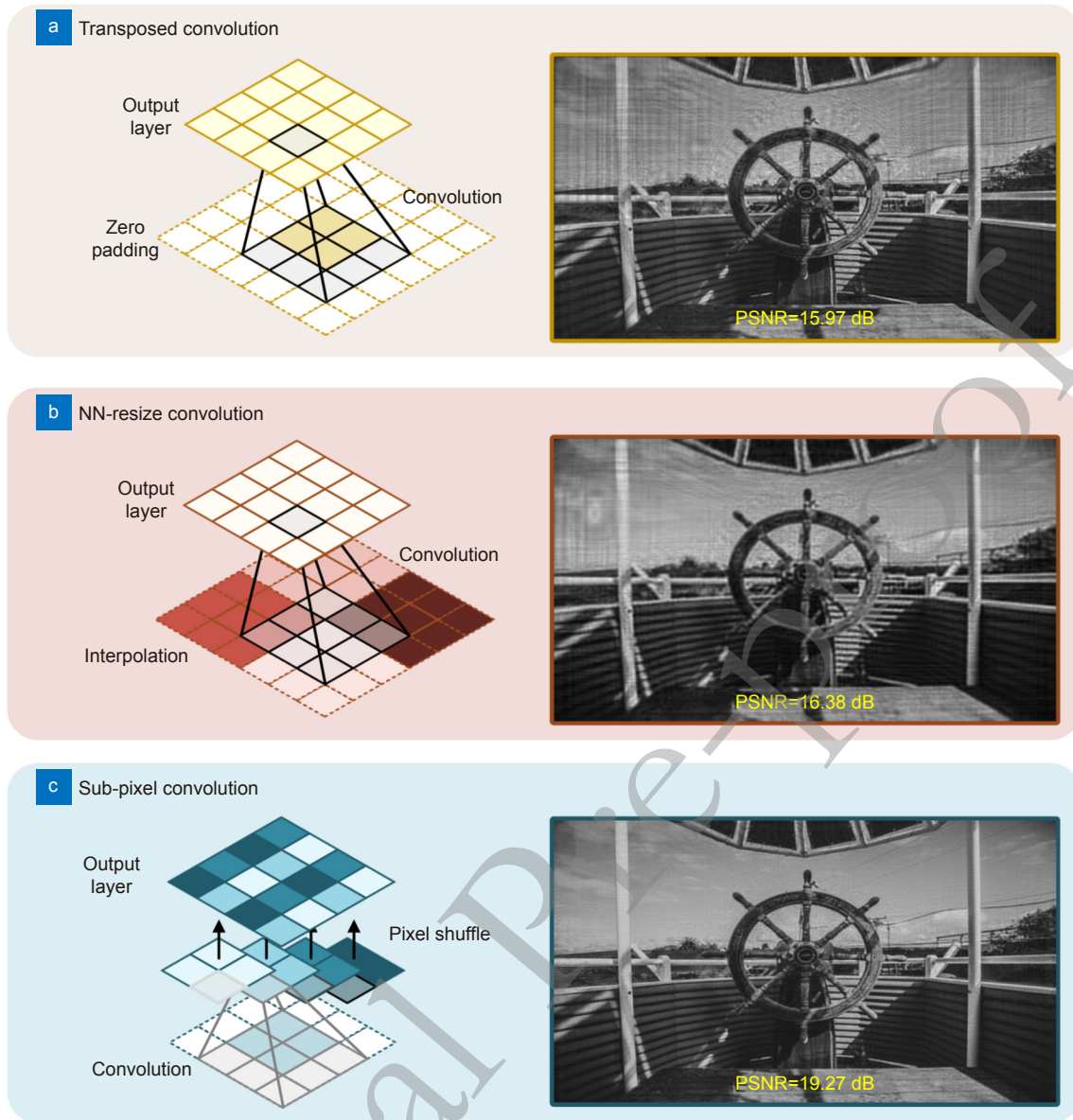


Fig. 4 | (a–c) Schematic diagram of the transposed convolution, NN-resize convolution, and sub-pixel convolution with their corresponding numerical simulations.

following formulas:

$$\begin{aligned} \text{Re}(\hat{C}_{\text{shift}}) &= R_c + i \cdot i \cdot I_s = R_c - I_s, \\ \text{Im}(\hat{C}_{\text{shift}}) &= I_c + R_s, \end{aligned} \quad (5)$$

where \hat{C}_{shift} is the complex amplitude distribution on the object plane before the second fftshift operation, R_c and I_c are the real and imaginary parts of FFT result in the cos path, R_s and I_s are the real and imaginary parts of FFT result in the sin path. The intensity distribution on the object plane is calculated by $\hat{I} = |\hat{C}|^2 = \text{Re}(\hat{C})^2 + \text{Im}(\hat{C})^2$.

Oversampling method

Although the model-driven deep learning is an effective tool for high-quality POH generation, the insufficient constraints cause the artifacts on the reconstructions. Here we propose to strengthen the constraints in the frequency domain for solving this problem. The spectrum of the light field is zero padded to double the size in the calculation process. The spectrum of the object plane S can be calculated as

$$\begin{aligned} S &= \mathcal{F}\{\hat{C}(x, y)\} = \mathcal{F}\{\mathcal{F}\{\exp[i\Phi(x_0, y_0)]\}\} \\ &= \exp[i\Phi(-x_0, -y_0)]. \end{aligned} \quad (6)$$

So the spectrum S is the inverted image of

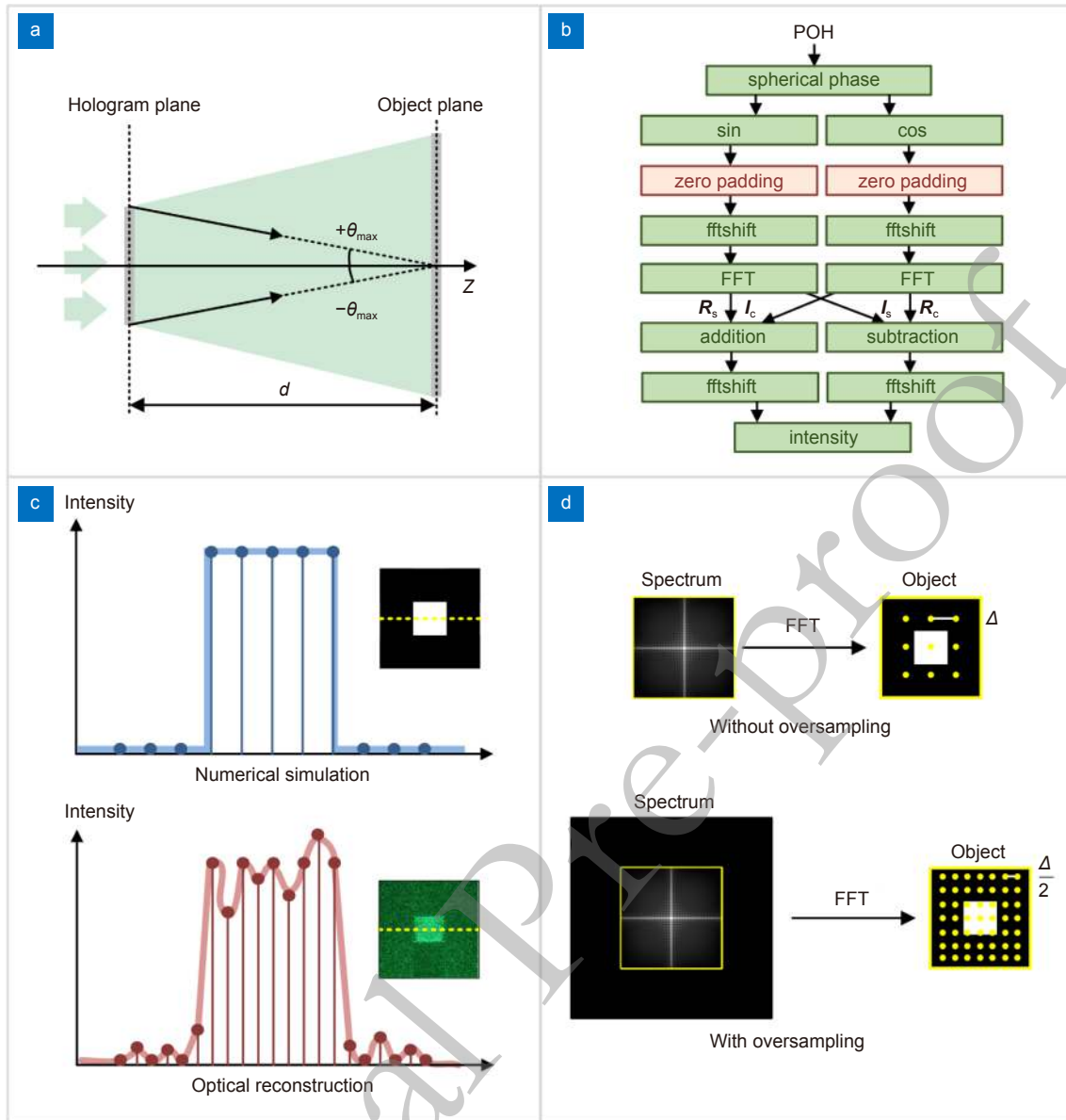


Fig. 5 | (a) Schematic diagram of the Fresnel diffraction model. (b) Fresnel diffraction model with oversampling method realized in the neural network layer manner. (c) Comparison between the numerical simulation and optical reconstruction with the undersampling problem. (d) Schematic diagram of the oversampling method.

$\exp[i\Phi(x_0, y_0)]$. The zero padding operations are directly added after the sin and cos layer, as shown in Fig. 5(b).

In addition, this method also plays the role of oversampling. Here we discuss the cause and practical elimination method of the undersampling problem in the Fresnel diffraction calculation. The frequency analysis below is one-dimensional for simplicity. According to the Nyquist-Shannon sampling theorem, the maximum sampling interval on the object plane is determined by the maximum spatial frequency of the light field. The light ray formed by connecting the edge point of the hologram and the center point of the object represents

the maximum spatial frequency that the hologram needs to recover, as shown in Fig. 5(a). It can be formulated as

$$f_{\max} = \frac{2}{\lambda} \sin \theta_{\max} = \frac{np}{\lambda d}, \quad (7)$$

where n and p are the pixel number and pixel pitch of the hologram, respectively. The maximum sampling interval satisfying the Nyquist-Shannon sampling theorem is

$$\Delta_{\max} = \frac{1}{2f_{\max}} = \frac{\lambda d}{2np}. \quad (8)$$

However, according to Eq. (3), the complex amplitude field on the object plane is obtained by multiplying the spherical phase and performing a single FFT.

According to the corresponding relationship in the frequency domain, the sampling interval of the object plane in the Fresnel diffraction calculation process is

$$\Delta = \frac{\lambda d}{np} = 2\Delta_{\max}. \quad (9)$$

Therefore, the Fresnel diffraction model generally faces the undersampling problem. The numerical simulations can't accurately represent the practical reconstructions, as shown in Fig. 5(c). The speckle noise often exists in experiments, which is a common problem of POH that need to be addressed⁴⁵. According to the reciprocal relationship between the frequency domain range and the spatial sampling interval, the zero-padding to the spectrum is Fourier transferred to the double pixel number and half pixel pitch on the object plane, as shown in Fig. 5(d). Therefore, the Nyquist-Shannon sampling theorem is satisfied without any additional information.

Loss function

The network training process is realized by calculating the loss function between the image dataset and the corresponding output images, and updating the convolution kernel by the Adam optimization algorithm according to the loss⁴⁶. The negative Pearson correlation coefficient (NPCC) is chosen as the loss function for the 4K-DMDNet. It guarantees linear amplification and bias-free reconstruction, which increases the convergence probability. The NPCC between the input image \mathbf{I} and output image $\hat{\mathbf{I}}$ can be formulated as

$$\mathcal{L}_{\text{NPCC}}(\hat{\mathbf{I}}, \mathbf{I}) = - \frac{\sum_i^n (\hat{\mathbf{I}}_i - \bar{\hat{\mathbf{I}}})(\mathbf{I}_i - \bar{\mathbf{I}})}{\left\{ \sum_i^n (\hat{\mathbf{I}}_i - \bar{\hat{\mathbf{I}}})^2 \sum_i^n (\mathbf{I}_i - \bar{\mathbf{I}})^2 \right\}^{1/2}}. \quad (10)$$

Display demonstration of 4K-DMDNet

We verified the feasibility of the proposed 4K-DMDNet by both numerical simulations and optical reconstructions. The training epoch is set as 40 and the distance between the object plane and the hologram plane is 0.3 m. The network is trained and tested with public image datasets, DIV2K_train_HR and DIV2K_valid_HR, respectively. And we use the Matlab Deep Learning Toolbox to realize the network building and training. The trained network model and training code are shown in ref.⁴⁷. All the algorithms were run on the same workstation with an Intel Xeon Gold 6248R CPU and an

NVIDIA Quadro GV100 GPU. Note that the transfer training was employed in our previous Holo-Encoder work for better display effects. In order to more intuitively compare the performance of different algorithms, this method was not used for all the following results.

We first compared the full-color simulations of the POHs generated by the traditional GS algorithm, Holo-Encoder, and the proposed 4K-DMDNet, as shown in Fig. 6(a-c). The test image was selected from the DIV2K_valid_HR dataset, which wasn't seen by the network before. From the detail views, we can see the GS algorithm faces the speckle noise problem caused by the initial random phase and the amplitude information loss. The contrast of the simulation was low. The Holo-Encoder faces the quality reduction caused by the limited learning capability. The 4K-DMDNet effectively suppressed the above problems and obtained the natural-looking reconstruction. The blurs in Fig. 6(c) are mainly caused by the detailed information loss in the down-sampling path. And it can be effectively improved by using other advanced network structures for POH generation. For example, the HRNet maintains high resolution through the whole process, while the U-Net recovers high resolution from low resolution⁴⁸. Fig. 6(d) shows the PSNR values under different runtimes of the above three algorithms. The GS algorithm achieved a better quality with more iterations and converged to an average quality of 16 dB for above 100 s. However, the 4K-DMDNet broke the trade-off between computation time and reconstruction quality. It can generate the POH in just 0.26 s, with the PSNR of 20.49 dB.

The experimental setup is shown in Fig. 7(a). A coherent beam was attenuated, expanded, and polarized before illuminating the SLM. A Holoeye GAEA-2 phase-only SLM with the resolution of 3840×2160 pixels was employed. The pixel pitch of the SLM was $3.74 \mu\text{m}$. The POHs of the target object were uploaded on the SLM. The reconstructed pattern was photographed at the distance of 0.3 m. Color holographic display was realized by time-multiplexing with the 638 nm red, 520 nm green, and 450 nm blue laser sources, as shown in Fig. 7(b). During the time period T , the POHs corresponding to the three channels were loaded onto the SLM in turn. A programmable light switch synchronously controls one of the red, green, and blue lasers that passes through and illuminates the SLM. When the period T is less than the human eye response time, the reconstructed images become a color image. It was high-fidelity and without

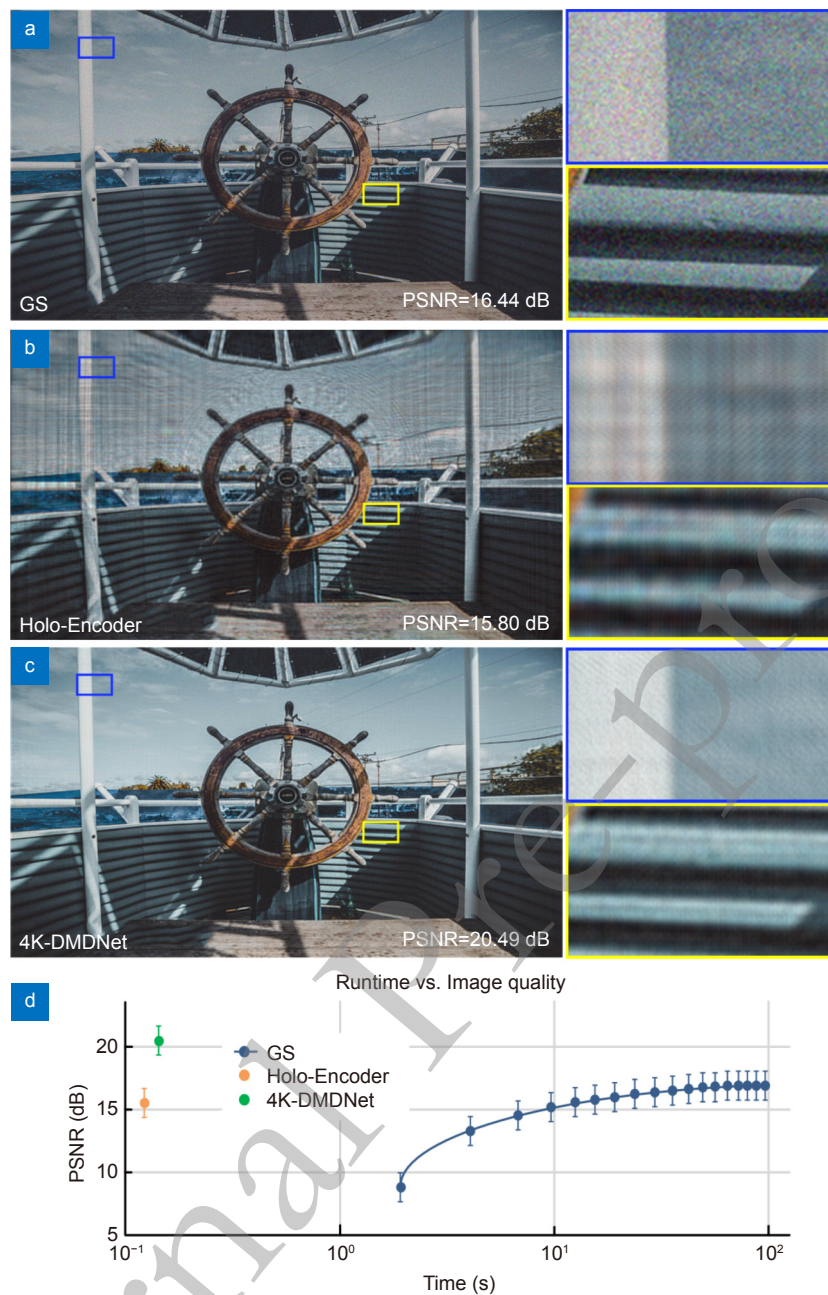


Fig. 6 | Contrast between the numerical simulations of POHs by (a) the GS algorithm, (b) Holo-Encoder, and (c) 4K-DMDNet. (d) Evaluation of algorithm runtime and image quality. The length of the bar represents the standard deviation of 100 samples (DIV2K_valid_HR).

interference-induced noise, as shown in Fig. 7(c). As to the GS algorithm, since the practical light propagation is not ideal, the speckle noise problem is always magnified on the optical reconstructions. An incoherent Light Emitting Diode (LED) or a partially coherent Self-scanning Light Emitting Device (SLED) source could be employed to reduce the speckle noise. However, the low-coherence light sources also create blurred details and reduced image sharpness. By contrast, in this work, the POHs predicted by the 4K-DMDNet are too smooth to generate the vortex phase. Therefore, compared with the

simulations, the optical reconstructions with the laser illumination have no quality degradation and the speckle noise is mostly suppressed.

The 4K-DMDNet learns the latent encodings of POHs in an unsupervised way, which enables a better generalization compared with the data-driven deep learning networks. In the field of two-photon microscopic imaging, optical micromanipulation, and laser nuclear fusion, the patterns with simple shapes and high contrast are widely used. A binary object was experimentally reconstructed to demonstrate the high generalization of the 4K-DMDNet,

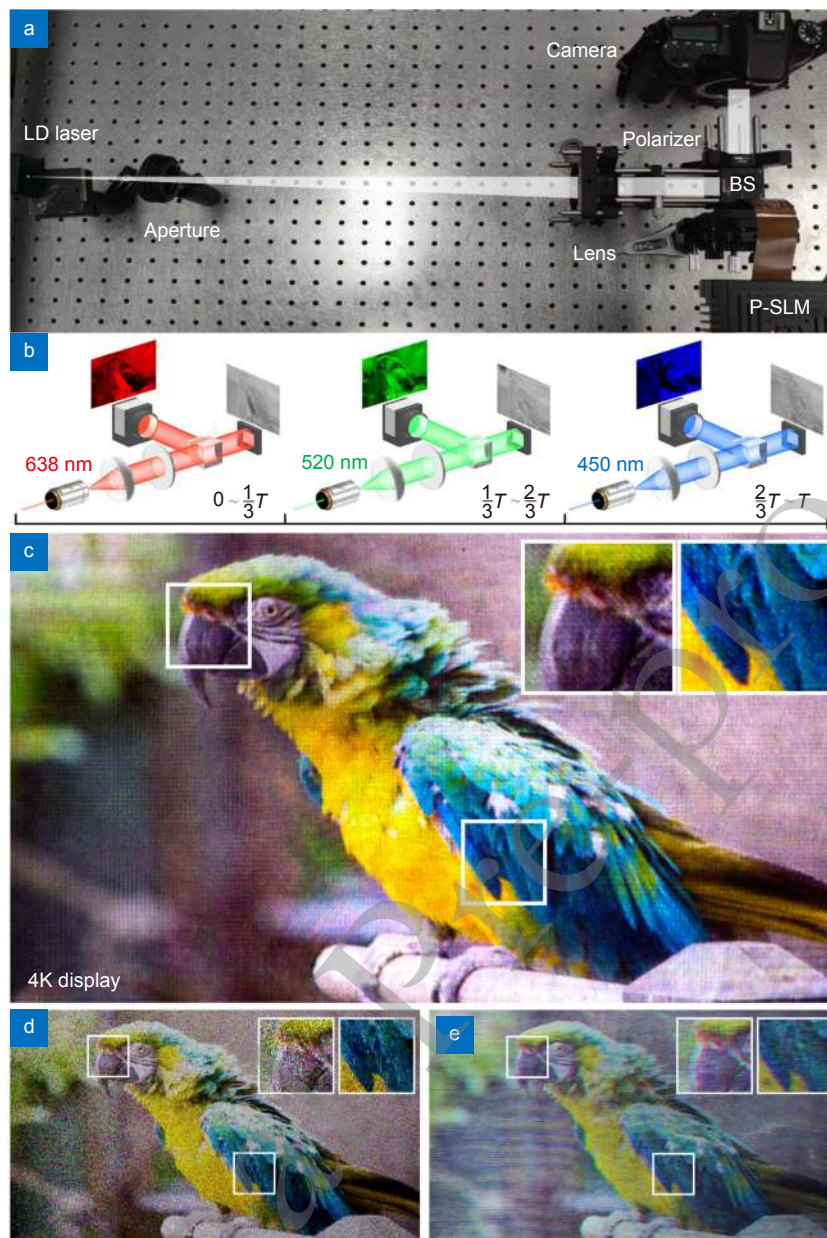


Fig. 7 | (a) Photograph of the experimental setup. (b) Schematic diagram of the time multiplexing method for full-color display. (c) Full-color 4K optical reconstruction by 4K-DMDNet and its detail views. (d) Optical reconstruction by GS algorithm. (e) Optical reconstruction by Holo-Encoder.

as shown in Fig. 8. The intensity of the signal part was uniform and the background showed no bottom noise. It is pretty applicable to the head-up display (HUD) and diffractive optical elements (DOEs) design.

The ability to reconstruct 3D scenes of the 4K-DMDNet is presented in Fig. 9. The objects in the scene were at different depths which is indicated by the grayvalue in Fig. 9(b). The 4K-DMDNet could be applied to generate POHs for the layer-oriented objects with the value of the depth. The reconstruction distances were set as 0.28 m, 0.3 m and 0.32 m, respectively. The obvious focusing and defocusing effects can be observed by using a camera.

Conclusions

In summary, we propose the 4K-DMDNet model-driven neural network capable of generating high-fidelity 4K computer-generated holograms. The constraint of the reconstructions in the frequency domain is strengthened, which ensures the high-precision optical reconstructions. The sub-pixel convolution method solves the limited learning capability problem which typically appears in the existing hologram generation networks. Compared with the transposed convolution method and NN-resize convolution method, the image quality can be improved to 19.27 dB. Full-color and binary optical reconstructions

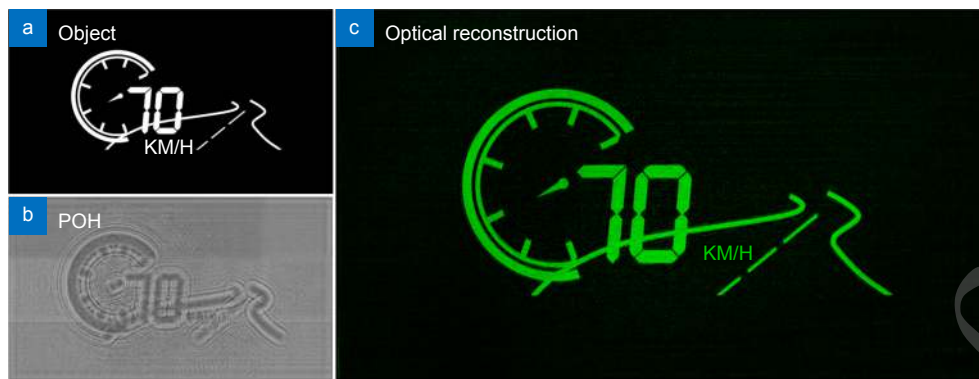


Fig. 8 | (a) Object, (b) POH, and (c) optical reconstruction of the binary target.

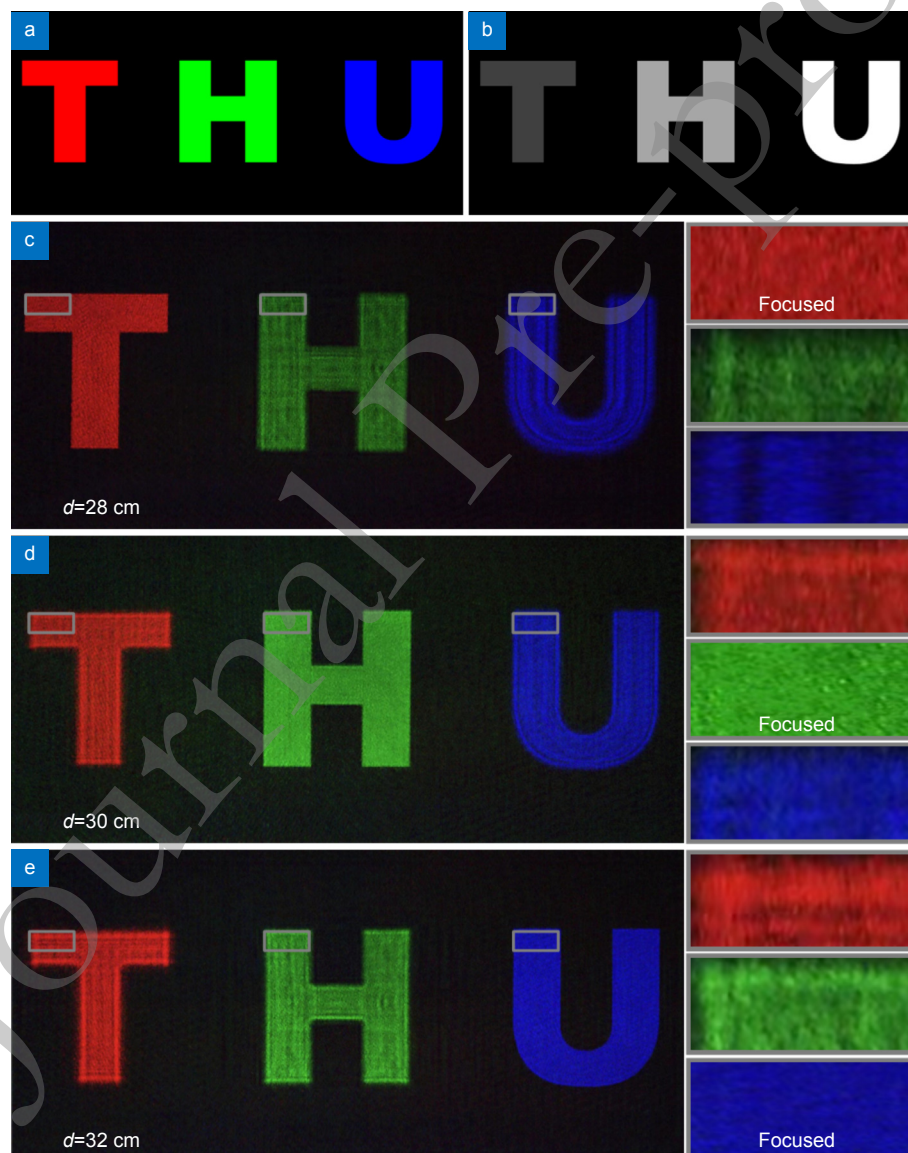


Fig. 9 | (a) All-in-focus image of the 3D scene. (b) Depth map of the 3D scene. (c) (d) and (e) Optical reconstructions of the 4K-DMDNet for 3D scene at the 28 cm, 30 cm and 32 cm, respectively. The enlarged views are presented at the right side.

tions have been obtained. The display quality outperforms the traditional iterative algorithms and data-driven deep learning algorithms. We believe that our approach further makes the computer-generated holographic display theory to be a viable technology for productive practice.

The current network architecture is based on the universal U-Net. It is suggested that the accurate physical models and smart mapping relations can also be applied to other advanced network architectures, such as generative adversarial network and graph neural network. More efforts will be needed to accelerate the calculation speed in the future. The proposed 4K-DMDNet can also be integrated for laboratory studies such as metasurface design and additive manufacturing. It should be a very powerful algorithm for portable virtual and augmented reality with the rapid development of ASICs. And it provides a versatile CNN framework for the solutions of various ill-posed inverse problems with mass data.

References

- Shi L, Li BC, Kim C, Kellnhofer P, Matusik W. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).
- Zhang CL, Zhang DF, Bian ZP. Dynamic full-color digital holographic 3D display on single DMD. *Opto-Electron Adv* **4**, 200049 (2021).
- He ZH, Sui XM, Jin GF, Cao LC. Progress in virtual reality and augmented reality based on holographic display. *Appl Opt* **58**, A74–A81 (2019).
- Zhao Y, Cao LC, Zhang H, Kong DZ, Jin GF. Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method. *Opt Express* **23**, 25440–25449 (2015).
- Jiang Q, Jin GF, Cao LC. When metasurface meets hologram: principle and advances. *Adv Opt Photonics* **11**, 518–576 (2019).
- Huang LL, Chen XZ, Mühlenbernd H, Zhang H, Chen SM et al. Three-dimensional optical holography using a plasmonic metasurface. *Nat Commun* **4**, 2808 (2013).
- Guo JY, Wang T, Quan BG, Zhao H, Gu CZ et al. Polarization multiplexing for double images display. *Opto-Electron Adv* **2**, 180029 (2019).
- Gao H, Fan XH, Xiong W, Hong MH. Recent advances in optical dynamic meta-holography. *Opto-Electron Adv* **4**, 210030 (2021).
- Saha SK, Wang D, Nguyen VH, Chang Y, Oakdale JS et al. Scalable submicrometer additive manufacturing. *Science* **366**, 105–109 (2019).
- Gerchberg RW, Saxton WOA. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1972).
- Tian SZ, Chen LZ and Zhang H. Optimized Fresnel phase hologram for ringing artifacts removal in lensless holographic projection. *Appl Opt* **61**, B17–B24 (2022).
- Chakravarthula P, Peng YF, Kollin J, Fuchs H, Heide F. Wirtinger holography for near-eye displays. *ACM Trans Graph* **38**, 213 (2019).
- Zhang JZ, Pégard N, Zhong JS, Adesnik H, Waller L. 3D computer-generated holography by non-convex optimization. *Optica* **4**, 1306–1313 (2017).
- He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016); <http://doi.org/10.1109/CVPR.2016.90>.
- Liao MH, Zheng SS, Pan SX, Lu DJ, He WQ et al. Deep-learning-based ciphertext-only attack on optical double random phase encryption. *Opto-Electron Adv* **4**, 200016 (2021).
- Li YX, Qian JM, Feng SJ, Chen Q, Zuo C. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement. *Opto-Electron Adv* **5**, 210021 (2022).
- Blinder D, Birnbaum T, Ito T, Shimobaba T. The state-of-the-art in computer generated holography for 3D display. *Light Adv Manuf* **3**, 35 (2022).
- He ZH, Sui XM, Jin GF, Chu DP, Cao LC. Optimal quantization for amplitude and phase in computer-generated holography. *Opt Express* **29**, 119–133 (2021).
- Sui XM, He ZH, Jin GF, Chu DP, Cao LC. Band-limited double-phase method for enhancing image sharpness in complex modulated computer-generated holograms. *Opt Express* **29**, 2597–2612 (2021).
- Liu KX, He ZH, Cao LC. Double amplitude freedom Gerchberg–Saxton algorithm for generation of phase-only hologram with speckle suppression. *Appl Phys Lett* **120**, 061103 (2022).
- Liu KX, He ZH, Cao LC. Pattern-adaptive error diffusion algorithm for improved phase-only hologram generation. *Chin Opt Lett* **19**, 050501 (2021).
- Kang JW, Park BS, Kim JK, Kim DW, Seo YH. Deep-learning-based hologram generation using a generative model. *Appl Opt* **60**, 7391–7399 (2021).
- Lee J, Jeong J, Cho J, Yoo D, Lee B et al. Deep neural network for multi-depth hologram generation and its training strategy. *Opt Express* **28**, 27137–27154 (2020).
- Zheng HD, Hu JB, Zhou CJ, Wang XX. Computing 3D phase-type holograms based on deep learning method. *Photonics* **8**, 280 (2021).
- Liu SC, Chu DP. Deep learning for hologram generation. *Opt Express* **29**, 27373–27395 (2021).
- Khan A, Zhang ZJ, Yu YJ, Khan MA, Yan KT et al. GAN-Holo: generative adversarial networks-based generated holography using deep learning. *Complexity* **2021**, 6662161 (2021).
- Horisaki R, Takagi R, Tanida J. Deep-learning-generated holography. *Appl Opt* **57**, 3859–3863 (2018).
- Goi H, Komuro K, Nomura T. Deep-learning-based binary hologram. *Appl Opt* **59**, 7103–7108 (2020).
- Chang CL, Wang D, Zhu DC, Li JM, Xia J et al. Deep-learning-based computer-generated hologram from a stereo image pair. *Opt Lett* **47**, 1482–1485 (2022).
- Hossein Eybposh M, Caira NW, Atisa M, Chakravarthula P, Pégard NC. DeepCGH: 3D computer-generated holography using deep learning. *Opt Express* **28**, 26636–26650 (2020).
- Horisaki R, Nishizaki Y, Kitaguchi K, Saito M, Tanida J. Three-dimensional deeply generated holography [Invited]. *Appl Opt* **60**, A323–A328 (2021).

32. Peng YF, Choi S, Padmanaban N, Wetzstein G. Neural holography with camera-in-the-loop training. *ACM Trans Graph* **39**, 185 (2020).
33. Gopakumar M, Kim J, Choi S, Peng YF, Wetzstein G. Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays. *Opt Lett* **46**, 5822–5825 (2021).
34. Peng YF, Choi S, Kim J, Wetzstein G. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Sci Adv* **7**, eabg5040 (2021).
35. Ishii Y, Shimobaba T, Blinder D, Birnbaum T, Schelkens P et al. Optimization of phase-only holograms calculated with scaled diffraction calculation through deep neural networks. *Appl Phys B* **128**, 22 (2022).
36. Yu T, Zhang SJ, Chen W, Liu J, Zhang XY et al. Phase dual-resolution networks for a computer-generated hologram. *Opt Express* **30**, 2378–2389 (2022).
37. Sun XH, Mu XY, Xu C, Pang H, Deng QL et al. Dual-task convolutional neural network based on the combination of the U-Net and a diffraction propagation model for phase hologram design with suppressed speckle noise. *Opt Express* **30**, 2646–2658 (2022).
38. Wu JC, Liu KX, Sui XM, Cao LC. High-speed computer-generated holography using an autoencoder-based deep neural network. *Opt Lett* **46**, 2908–2911 (2021).
39. Situ GH. Deep holography. *Light Adv Manuf* **3**, 13 (2022).
40. Zuo C, Qian JM, Feng SJ, Yin W, Li YX et al. Deep learning in optical metrology: a review. *Light Sci Appl* **11**, 39 (2022).
41. Shi L, Li BC, Matusik W. End-to-end learning of 3D phase-only holograms for holographic display. *Light Sci Appl* **11**, 247 (2022).
42. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015); http://doi.org/10.1007/978-3-319-24574-4_28.
43. Shi WZ, Caballero J, Huszár F, Totz J, Aitken AP et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* 1874–1883 (IEEE, 2016); <http://doi.org/10.1109/CVPR.2016.207>.
44. Dumoulin V, Shlens J, Kudlur M. A learned representation for artistic style. In *Proceedings of the 5th International Conference on Learning Representations* (IEEE, 2016). <https://arxiv.org/abs/1610.07629>
45. Shimobaba T, Blinder D, Birnbaum T, Hoshi I, Shiomi H et al. Deep-learning computational holography: a review. *Front Photonics* **3**, 854391 (2022).
46. Kingma DP, Ba J. Adam: a method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations* (2014). <https://arxiv.org/abs/1412.6980>
47. Source code: <https://github.com/THUHoloLab/4K-DMDNet>
48. Wang JD, Sun K, Cheng TH, Jiang BR, Deng CR et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell* **43**, 3349–3364 (2021).

Acknowledgements

We are grateful for financial supports from National Natural Science Foundation of China (62035003, 61775117), China Postdoctoral Science Foundation (BX2021140), and Tsinghua University Initiative Scientific Research Program (20193080075).

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at <https://doi.org/10.29026/oea.2023.220135>