# Opto-Electronic Advances

CN 51–1781/TN ISSN 2096–4579 (Print) ISSN 2097–3993 (Online)


## IncepHoloRGB: multi-wavelength network model for full-color 3D computer-generated holography


Xuan Yu, Zhilin Teng, Xuhao Fan, Tianchi Liu, Wenbin Chen, Xinger Wang, Zhe Zhao, Wei Xiong and Hui Gao



**Citation:** Yu X, Teng ZL, Fan XH, et al. IncepHoloRGB: multi-wavelength network model for full-color 3D computer-generated holography. *Opto-Electron Adv* **8**, 250130(2025).

https://doi.org/10.29026/oea.2025.250130

Received: 30 May 2025; Accepted: 27 August 2025; Published online: 23 October 2025


## Related articles

Opto-Electronic Advances

Article

October 2025, Vol. 8, No. 10

DOI: 10.29026/oea.2025.250130 CSTR: 32247.14.oea.2025.250130

# IncepHoloRGB: multi-wavelength network model for full-color 3D computer-generated holography

Xuan Yu[1], Zhilin Teng[1], Xuhao Fan[1], Tianchi Liu[2], Wenbin Chen[1], Xinger Wang[1], Zhe Zhao[1], Wei Xiong[1,3*] and Hui Gao[1,3*]

The popularity of deep learning has boosted computer-generated holography (CGH) as a vibrant research field, particularly physics-driven unsupervised learning. Nevertheless, present unsupervised CGH models have not yet explored the potential of generating full-color 3D holograms through a unified framework. In this study, we propose a lightweight multi-wavelength network model capable of high-fidelity and efficient full-color hologram generation in both 2D and 3D display, called IncepHoloRGB. The high-speed simultaneous generation of RGB holograms at 191 frames per second (FPS) is based on Inception sampling blocks and multi-wavelength propagation module integrated with depth-traced superimposition, achieving an average structural similarity (SSIM) of 0.88 and peak signal-to-noise ratio (PSNR) of 29.00 on the DIV2K test set in reconstruction. Full-color reconstruction of numerical simulations and optical experiments shows that IncepHoloRGB is versatile to diverse scenarios and can obtain authentic full-color holographic 3D display within a unified network model, paving the way for applications towards real-time dynamic naked-eye 3D display, virtual and augmented reality (VR/AR) systems.

**Keywords:** computer-generated holography; deep learning; physics-driven model; full-color holographic 3D display

Yu X, Teng ZL, Fan XH et al. IncepHoloRGB: multi-wavelength network model for full-color 3D computer-generated holography. *Opto-Electron Adv* **8**, 250130 (2025).

## Introduction

Computer-generated holography (CGH), as one of the most attractive next-generation three-dimensional (3D) display technology, possesses the capacity to provide authentic depth cues of 3D scenes via faithfully recording the optical field with computational simulations[1–6] and has been widely applied in many fields such as dynamic holographic display[7–10], additive manufacturing[11] and information encryption[12,13]. Among the various modulation principles of CGH, phase-only hologram (POH) is preferred owing to its higher diffraction efficiency, decent compatibility with commercial spatial light modulators (SLMs), and absence of conjugate images during reconstruction. However, the process of POH generation is predominately an ill-posed inverse problem, since only the intensity or amplitude information of the target scene is known while the original wavefront phase remains inaccessible in the initial calculation of

[1]Wuhan National Laboratory for Optoelectronics and School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China; [2]National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China; [3]Optics Valley Laboratory, Wuhan 430074, China.

*Correspondence: W Xiong, E-mail: weixiong@hust.edu.cn; H Gao, E-mail: gaohui_wnlo@hust.edu.cn

Received: 30 May 2025; Accepted: 27 August 2025; Published online: 23 October 2025

holograms[14]. Conventional CGH methods employ iterative approaches like Gerchberg-Saxton (GS) algorithm[15], Wirtinger holography[16], stochastic gradient descent (SGD) algorithm[17] and their variants, or non-iterative techniques such as double phase-amplitude coding (DPAC)[18], all facing a fundamental trade-off between reconstruction fidelity and computational efficiency. Iterative algorithms take substantial computational time and resources to improve the image quality as iteration count increases, while speckle noise still exits due to the initial random phase and other mismatched errors[19,20]; non-iterative algorithms such as DPAC run fast but the reconstructed images are deteriorated by severe artifacts near high-frequency boundaries[21]. These limitations hinder current methods from meeting the escalating demands for high-quality, full-color, naked-eye 3D display in modern information applications.

Deep learning-based CGH models have emerged as a transformative solution to the longstanding quality-speed trade-off in hologram generation with the development of artificial intelligence (AI) algorithms and high-performance computing hardware like graphics processing units (GPUs)[22]. Common supervised approaches require massive dataset comprised of target scenes and corresponding pre-computed holograms for fitting the mapping relationship[23–26], which is a process fundamentally constrained by the inherent paradox that perfect hologram pre-calculation through conventional methods remains unattainable due to the ill-posed nature of inverse problems. To address this limitation, physics-driven unsupervised models have leveraged the learning abilities of deep neural networks (DNNs) informed by wave optics principles instead of treating networks as black-box approximators[17,27–32], enabling high-quality POH generation without label holograms. Nevertheless, present unsupervised models predominantly focus on one certain aspect in CGH. Some models generate single-wavelength holograms limited to monochromatic 2D display[27,31,33–38], or obtain RGB holograms through three separate single-wavelength network models that triple computational resource consumption[17,29,30,32], which is difficult to satisfy the escalating global demand for efficient computing solutions[39]. Although notable advances have been made in 3D holographic reconstruction[28,37,40], there is still a lack of comprehensive support in full high definition (FHD) resolution (1920×1080), full-color display and multi-depth-varying perception using a single high-performance model.

In this paper, we propose a lightweight unsupervised CGH model named IncepHoloRGB, which generates RGB FHD POHs simultaneously through a unified framework fulfilled in both 2D and 3D display mode. The network structure combines the Inception-inspired sampling blocks and multi-wavelength propagation module, which prominently enhances the computational efficiency with high reconstruction fidelity. The depth-traced superimposition and 3D enhanced training strategy guarantee that the model can learn correct depth cues. The scalability of IncepHoloRGB is demonstrated with full-color scenario reconstruction in both simulations and experiments, which can achieve high-performance POH generation for full-color holographic 3D display, promising unprecedented potential for real-time dynamic 3D display systems such as virtual and augmented reality (VR/AR).

## End-to-end unsupervised learning of full-color 3D CGH

### Workflow of full-color hologram generation

Directly generating RGB POHs is not only suitable for practical applications such as full-color display but also beneficial for computational efficiency, due to the fact that the calculation process of every hologram at different wavelength is similar with only minor variations in physical settings, permitting a shared computational workflow rather than requiring separate and repetitive computations. Therefore, we propose an end-to-end deep learning-based full-color CGH framework to generate RGB POHs simultaneously. As illustrated in Fig. 1, a target RGB image is processed through depth estimation method to generate corresponding depth map, forming the RGBD (RGB-Depth) image as an input of the network model. IncepHoloRGB model is composed of two similar subnetworks. The 1st subnetwork RGB(D)2CNet (RGB(D)-image-to-complex-field network) predicts the amplitude $A_{\mathrm{tar}}$ and phase $\varphi_{\mathrm{tar}}$ of the target RGB(D) image at red, green and blue channels to derive the complex field, which are then propagated to midpoint plane via the multi-wavelength propagation module. The 2nd subnetwork C2PNet (complex-field-to-phase network) transforms and decomposes the multi-wavelength complex optical field of the midpoint plane $A_{\mathrm{mid}}\exp(\mathrm{i}\varphi_{\mathrm{mid}})$ into two pure phase values $\Phi 1$ and $\Phi 2$, finally interlaced into RGB POHs through checkerboard sampling (specific analysis and discussion are detailed in Supplementary information Section 1).

To obtain vivid 3D visual cues, a depth-traced

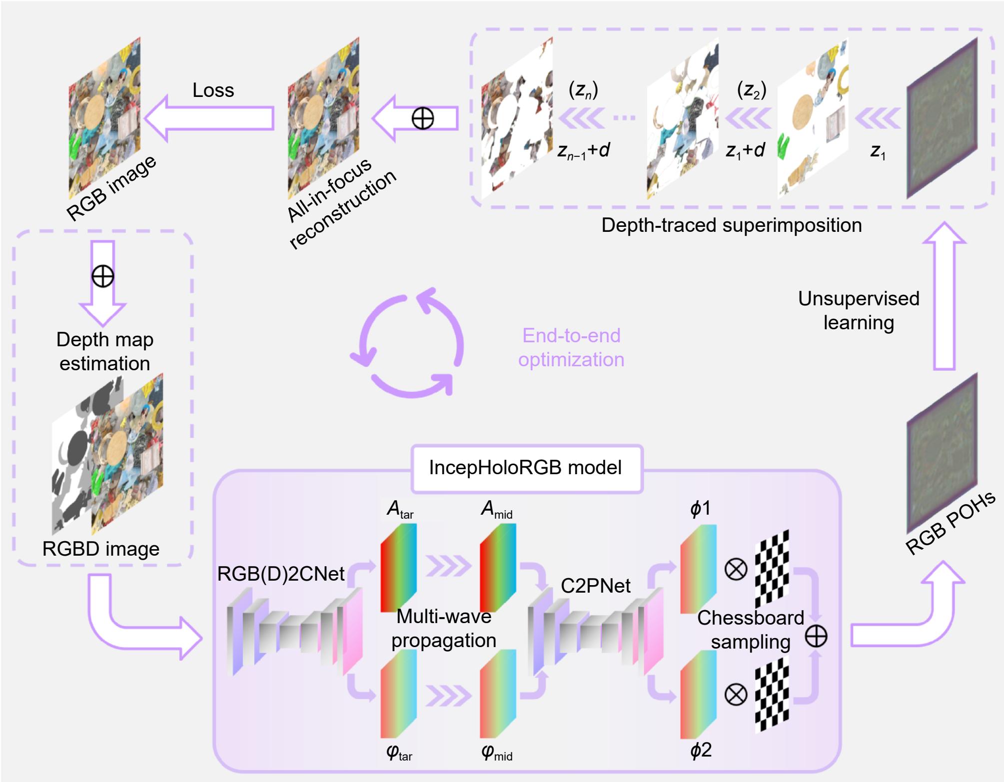


**Fig. 1 |** Full-color hologram generation workflow of IncepHoloRGB.

superimposition method is proposed to ensure the consistency of front-rear spatial relationships. As shown in the upper part of Fig. 1 specifically, 3D POH is firstly propagated from the hologram plane to $z_1$ plane, where the focused image is the square of optical field's amplitude in $z_1$; next, the images of subsequent layers are no longer calculated from the hologram plane but directly from the amplitude and phase of the previous layer's optical field with propagation distance $d$, which are then masked with depth map and superimposed to form an all-in-focus image. This process inherently involves the near-far nature of different depth layers while also completing the numerical reconstruction of holograms, enabling the generation of POHs through pure unsupervised learning without any pre-computed label holograms. The loss function is directly calculated between the reconstructed all-in-focus image and the initial target image to update the network parameters, which constitutes a comprehensive end-to-end optimization, formulizing as:

$$\begin{aligned}\hat{I} &= f_{\mathrm{prop+}}\left(\Phi_{\mathrm{holo}}\right) = f_{\mathrm{prop+}}\left(f_{\mathrm{model}}\left(I\right)\right)\\ &= f_{\mathrm{prop+}}\left(f_{\mathrm{DNN2}}\left(f_{\mathrm{prop\text{-}}}\left(f_{\mathrm{DNN1}}\left(I\right)\right)\right)\right),\\ \Phi_{\mathrm{opt}} &= \underset{\Phi}{\mathrm{minimize}}\,\mathcal{L}_{\mathrm{tar}}\left(\hat{I}, I\right),\end{aligned} \tag{1}$$

where $f_{\mathrm{prop+}}$ and $f_{\mathrm{prop-}}$ represent the forward and backward diffraction of the proposed multi-wavelength propagation module, $f_{\mathrm{DNN1}}$ is RGB(D)2CNet, $f_{\mathrm{DNN2}}$ is C2PNet, $I$ is the target RGB or RGBD image, $\hat{I}$ is the reconstructed image intensity predicted by the network model, $\mathcal{L}_{\mathrm{tar}}$ is the target loss function.

It should be noted that IncepHoloRGB demonstrates flexibility by supporting the generation of both 2D and 3D POHs. To achieve 3D POH generation, the modules delineated within the dashed boxes are required; however, they can be omitted for 2D display. Besides, the whole workflow is processed at full-color channels, enabling rapid generation of high-fidelity POHs for red, green, and blue wavelengths simultaneously within a unified framework.

## Network architecture

The representation learning capability of designed neural networks plays a key role in full-color and high-resolution POH generation, as computational complexity scales with increasing number of pixels. On the other hand, excessive number of network parameters consumes a plethora of computational resources, resulting in a slower running speed. To reconcile feature extraction efficacy with computational economy, a multi-path sampling DNN model is proposed inspired by Inception[41,42] and U-Net[43] and combines the merits of both feature extraction in convolutional operation and improved computational efficiency in architecture design (depicted in Fig. 2), attaining fast and accurate calculation of optical field information including amplitude and phase. Note that RGB(D)2CNet and C2PNet share the analogous architecture differentiated solely by the number of input channels (3 channels for RGB, 4 for RGBD, 6 for RGB amplitude and RGB phase) and the output mapping function (amplitude mapping or phase mapping), so it can be regarded as a universal network architecture.

Considering the tensor format $[B, C, N, M]$, where $B$ is the batch size of the images (usually 1), $C$ represents the number of channels, $N$ denotes the vertical size of the image, and $M$ denotes the horizontal size of the image. Firstly, as shown in Fig. 2(a), the input image undergoes compression and feature extraction through four downsampling blocks and each downsampling reduces the size of the image (including vertical and horizontal dimensions) by half and increases the number of channels. After three upsampling blocks for expansion and feature recovery, the image size doubles and the number of channels decreases gradually. Subsequent feature recovery is performed for different types of information. During the 4$^{th}$ upsampling, the amplitude or phase is passed through the separate upsampling block and residual block, finally transformed into corresponding optical field amplitude or phase information by the mapping layer. Hence the overall network consists of 4 downsampling blocks, 5 upsampling blocks and 2 residual blocks. To match the range of physical values in optics, we choose the sigmoid function as amplitude mapping and a variant of atan2 function[32] as phase mapping. Skip connections are set between every two sampling layers of the same size to preserve high-frequency details of the image. It should be emphasized that traditional U-Net generally follows the rule of doubling the number of channels per downsampling and halving the number of channels per upsampling, but here, in order to pursue higher compression ratios and less computational load, only a small number of channels are expanded, namely 16, 24, 32 and 48, which significantly reduces the number of

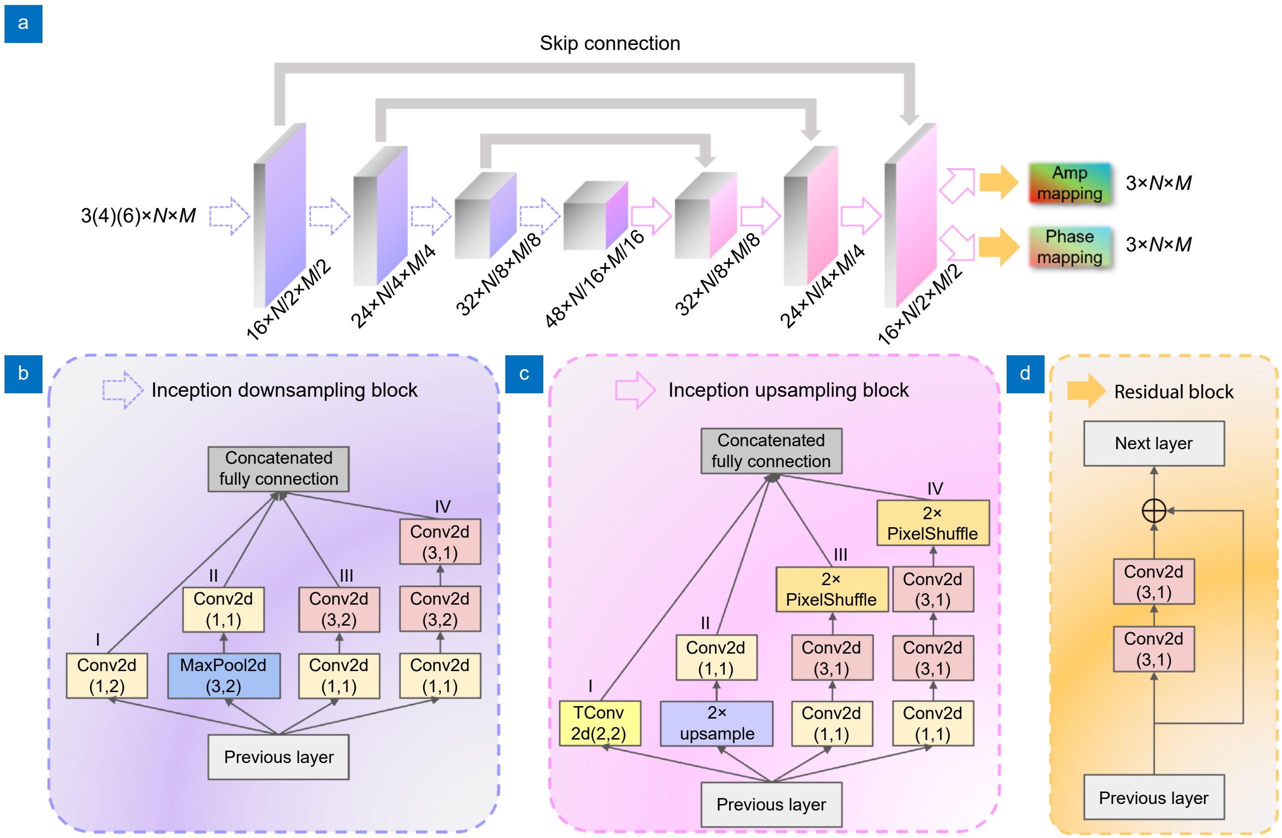


**Fig. 2 |** Neural network architecture of IncepHoloRGB. (**a**) The overall U-Net architecture. (**b**) Inception downsampling block architecture. (**c**) Inception upsampling block architecture. (**d**) Residual block architecture.
*($k$, $s$) in the blocks represents the kernel size $k$ and the stride $s$.

network parameters but scarcely degrades the representation ability.

The Inception sampling block, either downsampling in Fig. 2(b) or upsampling in Fig. 2(c), consists of 4 paths to implement multi-scale feature extraction and processing. Note that InstanceNorm and PReLU layers are sequentially added before every 3×3 convolution layer, which are omitted in Fig. 2(b–d) due to the purpose of simplification. Path Ⅰ connects input and output features through single 1×1 (2×2 transposed) convolutional kernel with stride of 2 in downsampling (upsampling) block. Max pooling (bicubic upsample) operation is added in path Ⅱ as alternative methods to capture (recovery) features followed by a 1×1 convolution utilized to adjust the number of channels. In path Ⅲ and Ⅳ, 1×1 convolutions are used to not only keep or reduce channel numbers to remove computational bottlenecks but also further strengthen the nonlinear mapping relationship before the expensive 3×3 convolutions. In order to further expand computational potential at various scales, we choose one 3×3 convolution in path Ⅲ with receptive field size of 3 and two 3×3 convolutions in path Ⅳ with receptive field size of 5. It is worth noting that conventional zero-padded transposed convolutions are replaced by the combination of convolutions and pixel shuffle in Inception upsampling block to realize finer upsampling results, since a great number of zero parameters are ineffective and dilute the features of the original information. Afterwards, features of different paths are concatenated in the channel dimension and then fully connected to form the output of the entire block. The core concept of Inception sampling blocks is the extensive use of 1×1 convolutions, which reduces computational costs and avoids the computational overhead caused by directly using large convolution kernels. Finally, the residual block, as shown in Fig. 2(d), consists of two 3×3 convolutions and skip connection. It is used to further enhance the expressive capacity of DNN before the final amplitude or phase mapping, and to improve the optimization and training efficiency of the network model through residual learning[44]. By a carefully crafted design, we increase the width of the network with little computational budget while keeping the depth constant to enhance the accurate representation and learning ability of IncepHoloRGB.

## Multi-wavelength propagation module

Existing unsupervised CGH approaches merely generate red, green and blue channel holograms through three models respectively in order to realize full-color display. The reason is that holograms are wavelength dependent and if we change the specific wavelength the hologram must be recalculated. However, considering tensors have independent channel dimension in deep learning framework, wavelength can be fully compatible with tensor's channel dimension to achieve mutually non-interfering parallel computing. Therefore, we propose a differentiable multi-wavelength propagation module that can simultaneously support diffraction field propagation for RGB wavelength channels, and utilize bandwidth limitations[45] to ensure high computational accuracy even at longer distances. The formula is as follows:

$$\begin{aligned} & u\left(x, y, \lambda_i, z_n\right)=\mathcal{F}^{-1}\left\{\mathcal{F}\left[u\left(x_0, y_0, \lambda_i\right) \cdot H \cdot H_{\text{filter}}\right]\right\}, \\ & H=\exp \left[\mathrm{j} \frac{2\pi}{\lambda_i} z_n \sqrt{1-\lambda_i^2\left(f_x^2+f_y^2\right)}\right], \\ & H_{\text{filter}}=f_x<\frac{1}{\lambda_i}\left[\left(\frac{2 z_n}{M \cdot p}\right)^2+1\right]^{-1/2} \\ & \& \quad f_y<\frac{1}{\lambda_i}\left[\left(\frac{2 z_n}{N \cdot p}\right)^2+1\right]^{-1/2}, \end{aligned} \tag{2}$$

where $u\left(x_0, y_0, \lambda_i\right)$ is the complex amplitude distribution of input optical field and $u\left(x, y, \lambda_i, z_n\right)$ is the output field, $H$ is the transfer function, $f_x$ and $f_y$ are the spatial frequencies, $H_{\text{filter}}$ is the spectrum filtering function, $p$ represents the pixel pitch, $\lambda_i (i=\mathrm{R}, \mathrm{G}, \mathrm{B})$ is the wavelength of RGB laser respectively, $z_n (n=1,2,3\ldots)$ represents the distance between different layers, which is simplified as a single variable in 2D scenarios. To eliminate the errors caused by circular convolution and ensure calculation accuracy, zero padding operation is adopted during sampling the target optical field.

## Hybrid loss function for color perception

The selection of loss function determines whether the training of network model meets the requirements of reconstruction quality. In our task, training a full-color CGH model is not just a copy of previous single-wavelength models since each wavelength channel is equally important. In order to further eliminate color deviation, it is necessary to design a color-aware loss function for full-color holographic reconstruction, which must have constraints for each wavelength so that every color component owns high reconstruction fidelity. Besides, the information of RGB images is more complicated than grayscale images, so how to make the network converge quickly is also a problem that must be considered. Therefore, a hybrid loss function for color perception is proposed, which consists of three parts: the negative Pearson correlation coefficient (NPCC) loss, vision-aware (VA) loss and focal frequency loss (FFL)[46].

NPCC loss enables the network to approach convergence rapidly in fewer epochs. Considering the image format $[C, N, M]$, where $C=3$ represents three wavelengths corresponding to red (R), green (G) and blue (B), $N$ and $M$ denote the vertical and horizontal size of the image, NPCC loss between the target image $I$ and the predicted image $\hat{I}$ can be formulated as:

$$\mathcal{L}_{\text{NPCC}}\left(\hat{I}, I\right) = \frac{1}{C}\sum_{k=1}^{C}\frac{1+\text{NPCC}_k\left(\hat{I}_k, I_k\right)}{2} ,$$

$$\text{NPCC}_k\left(\hat{I}_k, I_k\right) = -\frac{\sum_{j=1}^{N}\sum_{i=1}^{M}\left(\hat{I}_{k_{i,j}}-\bar{\hat{I}}_k\right)\left(I_{k_{i,j}}-\bar{I}_k\right)}{\left[\sum_{j=1}^{N}\sum_{i=1}^{M}\left(\hat{I}_{k_{i,j}}-\bar{\hat{I}}_k\right)^2\left(I_{k_{i,j}}-\bar{I}_k\right)^2\right]^{1/2}} \tag{3}$$

VA loss is proportionally composed of multi-scale structural similarity (MS-SSIM)[47] loss and mean square error (MSE) loss to enhance visual perception based on multi-scale image features and pixel-level deviations:

$$\mathcal{L}_{\text{VA}}\left(\hat{I}, I\right) = \alpha\cdot\mathcal{L}_{\text{MS-SSIM}}\left(\hat{I}, I\right) + (1-\alpha)\cdot\mathcal{L}_{\text{MSE}}\left(\hat{I}, I\right) , \tag{4}$$

where $\alpha$ is the scale factor. Note that MS-SSIM loss is also color-aware.

FFL concentrates on frequency information recovery to refine spectrum characteristics, which is defined as:

$$\begin{aligned}\mathcal{L}_{\text{FFL}}\left(\hat{I}, I\right) &= \frac{1}{C}\sum_{k=1}^{C}\text{FFL}_k\left(\hat{I}_k, I_k\right) ,\\ \text{FFL}_k\left(\hat{I}_k, I_k\right) &= \frac{1}{MN}\sum_{j=1}^{N}\sum_{i=1}^{M}w_{k_{i,j}}\\ &\quad\cdot\left|\text{DFT2}\left(\hat{I}_{k_{i,j}}\right)-\text{DFT2}\left(I_{k_{i,j}}\right)\right|^2 ,\end{aligned} \tag{5}$$

where DFT2 means 2D discrete Fourier transform (DFT) and $w_{k_{i,j}} = \left|\text{DFT2}\left(\hat{I}_{k_{i,j}}\right)-\text{DFT2}\left(I_{k_{i,j}}\right)\right|$ is the dynamic weight of different frequency component.

Finally, the hybrid target loss function is derived as:

$$\mathcal{L}_{\text{tar}}\left(\hat{I}, I\right) = \mathcal{L}_{\text{NPCC}}\left(\hat{I}, I\right) + \mathcal{L}_{\text{VA}}\left(\hat{I}, I\right) + \mathcal{L}_{\text{FFL}}\left(\hat{I}, I\right) . \tag{6}$$

## Results and discussion

We demonstrate the holographic display of the proposed IncepHoloRGB model in both 2D and 3D mode. The computational platform is based on Intel Core i7-13700KF CPU and NVIDIA GeForce RTX 3090 GPU, deploying PyTorch 2.5.1 with Python 3.10 to implement IncepHoloRGB. AdamW algorithm[48] is used for optimizing network model. Specific physical parameters are defined as follows: the SLM's pixel pitch is 8 μm, and the wavelengths for the red, green, and blue lasers are 638 nm, 520 nm and 450 nm, respectively. To accurately match the SLM's resolution of 1920×1080, bicubic interpolation is adopted before the 4$^{\text{th}}$ downsampling and 2$^{\text{nd}}$ upsampling blocks. The target images in dataset are resized to 1600×900 according to practical display region and padded with zeros to 1920×1080.

We first present IncepHoloRGB in 2D POH generation. To verify that our proposed multi-wavelength propagation module can calculate exact diffraction field in full-color channels even with long range, the propagation distance is set to 0.35 m. The initial learning rate is 0.001 and the maximum training epoch is 30, finally storing the network parameters when the loss value is minimized on the test set. Ablation study is conducted to test the availability of hybrid loss function for color perception on DIV2K dataset (containing 800 training images and 100 test images with 2K resolution), which is shown in Table 1. The structural similarity (SSIM), peak signal-to-noise ratio (PSNR) and root mean square error (RMSE) are utilized to quantitatively evaluate the reconstruction quality of different loss functions. NPCC loss alone makes the network converge soon but the final image fidelity is limited even with quite low loss value. Combined with VA loss, all the image assessment metrics are improved significantly, manifesting its contributions on visual fidelity. Training with NPCC, VA and FFL loss achieves the highest reconstruction quality owing to the fact that the network is optimized in both spatial and frequency domain.

The final version of IncepHoloRGB is trained on DF2K dataset, which combines DIV2K and Flickr2K (2650 2K training images) augmented with vertical and horizontal flipping. Figure 3 compares the numerically simulated reconstruction results of IncepHoloRGB with other conventional CGH methods. Note that the proposed multi-wavelength propagation module is used to extend the typical algorithms so that all the methods can generate holograms of RGB wavelengths simultaneously for fair comparison. Both iterative GS (Fig. 3(a)) and SGD (Fig. 3(b)) algorithms run for 100 times but still suffer from speckle noise by introduced random phase profile, among which SGD attains high PSNR but there

**Table 1 |** Ablation study with different loss functions trained on DIV2K (full-color channels).

| | NPCC | NPCC+VA | **NPCC+VA+FFL** |
|---|---|---|---|
| Minimum loss value (↓) | **0.0040** | 0.0281 | 0.0264 |
| SSIM (↑) | 0.49 | 0.83 | **0.86** |
| PSNR (↑) | 10.36 | 25.59 | **27.54** |
| RMSE (↓) | 0.3270 | 0.0557 | **0.0449** |


**Fig. 3 |** Comparisons of numerically reconstructed full-color images with conventional CGH methods with details magnification. (**a**) GS algorithm. (**b**) SGD algorithm. (**c**) DPAC algorithm. (**d**) IncepHoloRGB.

are conspicuous defects visually in magnified details, leading to a degradation on SSIM. The reconstructed image of non-iterative DPAC method is dim and has obvious artifacts throughout the entire scene, especially wavy stripes at the boundaries (Fig. 3(c)). In contrast, IncepHoloRGB achieves the best visual impression with clear and accurate detailed information.

We also compare IncepHoloRGB with other deep learning-based CGH methods in Table 2. All the models are modified to employ the high-accuracy propagation module described in Eq. (2) for fair comparison, tested on 100 DIV2K images. Instead of generating full-color holograms through three separate network models like Holo-Encoder[27] and HoloNet[17], our method outputs RGB POHs simultaneously via a unified framework with much fewer parameters and computation costs which outperforms previous deep learning models, while maintaining high fidelity for all color channels. IncepHoloRGB can generate RGB POHs with 0.88 SSIM and 29.00 PSNR in 15.7 ms/RGB frames, which is converted to frame rate of 191, holding great potential for achieving real-time and high-quality holographic display (a dynamic holographic video is provided in Supplementary information Section 4).

Figure 4 demonstrates the 2D holographic display results (more numerically simulated reconstruction results are provided in Supplementary information Section 2 and time-multiplexed full-color holographic display system used for optical experiments is detailed in Supplementary information Section 3). Note that our results of holographic experiments in Fig. 4(c) are all panoramic displays of the initial target images and directly captured

**Table 2 |** Quantitative comparisons of different CGH network models (full-color channels).

| Network model | Hologram resolution (↑) | Parameters (↓) | Model memory (MB) (↓) | SSIM (↑) | PSNR (↑) | Calculation time / frame (ms) (↓) | Generation frame rate (FPS) (↑) |
|---|---|---|---|---|---|---|---|
| Holo-Encoder | 1920×1072 | $7.4\times10^5\times3=2.22\times10^6$ | 2.85×3=8.55 | 0.49 | 21.97 | 19.7 | 51 |
| HoloNet | 1920×1072 | $2.87\times10^6\times3=8.61\times10^6$ | 10.95×3=32.85 | 0.80 | 27.41 | 21.8 | 46 |
| **IncepHoloRGB** | **1920×1080** | $\mathbf{1.09\times10^6}$ | **4.18** | **0.88** | **29.00** | **15.7/3=5.23** | **191** |

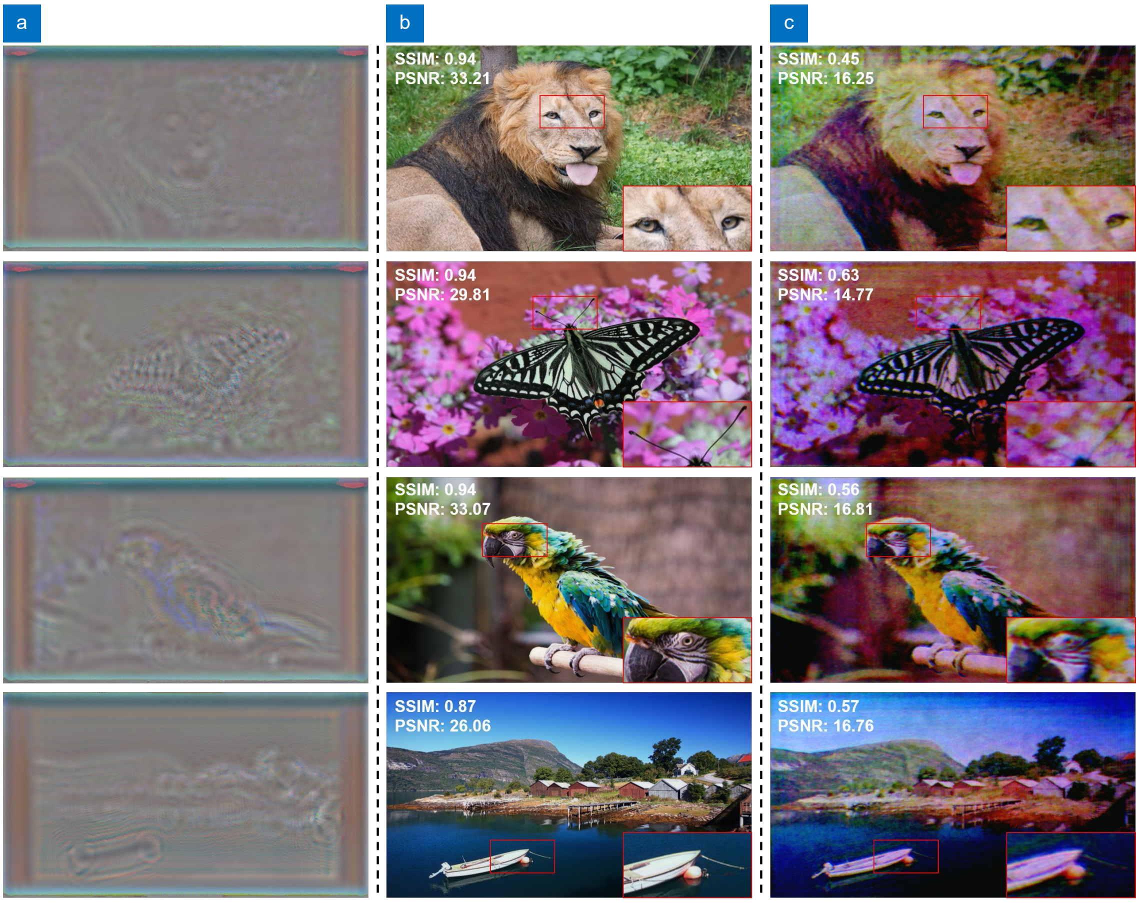


**Fig. 4 |** Holographic display of 2D scenes from IncepHoloRGB. (**a**) RGB POHs generated by IncepHoloRGB. (**b**) Numerically simulated reconstruction from (a) with details magnification in red boxes. (**c**) Optically experimental reconstruction from (a) with details magnification in red boxes.

from camera without any unnecessary cropping. From top to bottom, the lion's pupils, the feelers of the butterfly, the parrot's cheek patches near the eyes and the ropes at the bow of the boat are easily distinguishable in both numerical simulations (Fig. 4(b)) and practical experiments (Fig. 4(c)), representing the decent generalization ability and consistency for complicated scenarios.

For 3D POH generation, Boosting Monocular Depth[49] model is employed to yield depth map corresponding to DF2K images, establishing a large-scale RGBD dataset. Note that our framework is compatible with any 2D-3D depth estimation methods and here we choose this model because it can balance the consistency in low resolution depth and detailed representation in high-resolution depth. Due to the complexity and difficulty of holographic display in 3D scenes compared to 2D, dataset and training strategy are further enhanced. Firstly, the depth map is segmented averagely and randomly shuffled in all depth layers to enforce the network to learn all the features of images in dataset at each layer. Taking into account both computational resources and feasibility of simulations and experiments, the depth map is divided into 3 layers with distance of 0.35 m, 0.36 m and 0.37 m in our proof-of-concept setup. Secondly, a two-stage training method is adopted. IncepHoloRGB is pre-trained with low-resolution MIT-CGH-4K dataset[26] (including 3800 RGBD training images and 100 RGBD test images with a resolution of 384 × 384) with 0.001 learning rate within 10 epochs, and then retrained with high-resolution DF2K RGBD dataset to strengthen details with 0.0005 learning rate within 20 epochs. Figure 5 depicts the 3D holographic display results (more numerically simulated reconstruction results are provided in Supplementary information Section 2). The numerical simulation of all-in-focus image, shown in Fig. 5(a), exhibits the accurate reconstruction of the original RGB image. Specifically, with the target RGB image and predicted depth map (Fig. 5(b)) as RGBD input, IncepHoloRGB outputs 3D RGB POHs (Fig. 5(c)), which display precise depth focusing (in-focus detail and out-of-focus blur from near to far distances). Benefiting from the proposed depth-traced superimposition method and

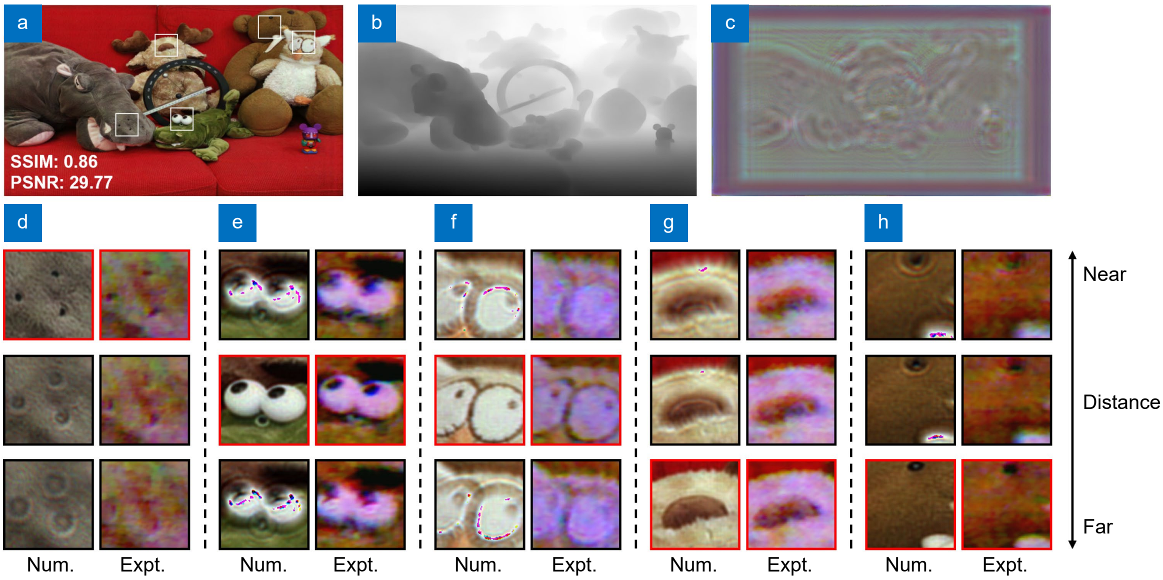


**Fig. 5 |** Holographic display of 3D scenes from IncepHoloRGB. (**a**) Numerically simulated reconstruction of all-in-focus image. (**b**) Depth map predicted by Boosting Monocular Depth model. (**c**) RGB POHs generated by IncepHoloRGB. (**d–h**) Numerically simulated and optically experimental reconstruction from (c) in different depths with details magnification (red boxes are in-focus parts and black boxes are out-of-focus parts).

3D enhanced training, IncepHoloRGB can generate high-fidelity 3D POHs even in intricate 3D scenes, achieving intuitive depth-of-field effects and accurate front-rear spatial relationships in both simulations and experiments.

## Conclusions

In summary, IncepHoloRGB is proposed to calculate high-fidelity FHD RGB computer-generated holograms simultaneously supporting both 2D and 3D full-color reconstruction. Combined with Inception sampling blocks and multi-wavelength propagation module, this lightweight network model significantly optimizes computational efficiency through parallel computing while ensuring image quality, attaining 0.88 SSIM and 29.00 PSNR with 191 FPS in POH reconstruction, which proves the pivotal role in dynamic holographic display. Further incorporated with depth-traced superimposition and 3D enhanced training, IncepHoloRGB achieves authentic 3D holographic reconstruction with in-focus and out-of-focus visual perception, showing the versatile applications for next-generation stereo display approach.

The current network design draws inspiration from Inception. Emerging evidence suggests that exploring concise architectures such as vision mamba (ViM)[50] can further reduce the network parameters while ensuring comparable representation learning capability. Due to hardware limitations, the depth map of current IncepHoloRGB 3D model is divided into 3 layers, but this method has no upper limit and can be extended to arbitrary number of layers without increasing the network scale. It requires additional time and computational costs only during training but no extra cost in hologram generation. As an unsupervised model, IncepHoloRGB owns the ample potential to train for more diverse 3D modeling types besides RGBD images and incorporate with innovative 3D reconstruction techniques like neural radiance fields (NeRF)[51] to realize full-angle-of-view holographic display, or combined with novel holographic camera[52,53] to realize real-time capture and display of real-world 3D scenes. It also provides a universal framework for processing multi-wavelength optical data with amplitude and phase values, possibly utilized in phase imaging[54] and metasurface design[55–58].

## References

1. Chang CL, Bang K, Wetzstein G et al. Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective. *Optica* **7**, 1563–1578 (2020).
2. Ge X, Chen S, Lin K et al. Deblurring, artifact-free optical coherence tomography with deconvolution-random phase modulation. *Opto-Electron Sci* **3**, 230020 (2024).
3. Blinder D, Birnbaum T, Ito T et al. The state-of-the-art in computer generated holography for 3D display. *Light Adv Manuf* **3**, 35 (2022).
4. Pi DP, Liu J, Wang YT. Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display. *Light Sci Appl* **11**, 231 (2022).
5. Yu PP, Liu YF, Wang ZQ et al. Ultrahigh-density 3D holographic

projection by scattering-assisted dynamic holography. *Optica* **10**, 481–490 (2023).

6. Yang YQ, Forbes A, Cao LC. A review of liquid crystal spatial light modulators: devices and applications. *Opto-Electron Sci* **2**, 230026 (2023).
7. Gao H, Wang YX, Fan XH et al. Dynamic 3D meta-holography in visible range with large frame number and high frame rate. *Sci Adv* **6**, eaba8595 (2020).
8. Zhang CL, Zhang DF, Bian ZP. Dynamic full-color digital holographic 3D display on single DMD. *Opto-Electron Adv* **4**, 200049 (2021).
9. Liu YC, Xu K, Fan XH et al. Dynamic interactive bitwise meta-holography with ultra-high computational and display frame rates. *Opto-Electron Adv* **7**, 230108 (2024).
10. Xu K, Liu YC, Fan XH et al. Simultaneous dynamic display of meta-hologram and meta-nanoprinting with high frame rate. *Laser Photonics Rev* **19**, 2400815 (2025).
11. Wang XE, Fan XH, Liu YC et al. 3D nanolithography via holographic multi-focus metalens. *Laser Photonics Rev* **18**, 2400181 (2024).
12. Zhu ZB, Li YF, Wang JF et al. Reconfigurable origami chiral response for holographic imaging and information encryption. *Opto-Electron Sci* **4**, 240026 (2025).
13. Li ZS, Sun JS, Fan Y et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electron Sci* **2**, 220023 (2023).
14. Sui X, He ZH, Chu DP et al. Non-convex optimization for inverse problem solving in computer-generated holography. *Light Sci Appl* **13**, 158 (2024).
15. Gerchberg RW. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1972).
16. Chakravarthula P, Peng YF, Kollin J et al. Wirtinger holography for near-eye displays. *ACM Trans Graph* **38**, 213 (2019).
17. Peng YF, Choi S, Padmanaban N et al. Neural holography with camera-in-the-loop training. *ACM Trans Graph* **39**, 185 (2020).
18. Maimone A, Georgiou A, Kollin JS. Holographic near-eye displays for virtual and augmented reality. *ACM Trans Graph* **36**, 85 (2017).
19. Chen LZ, Tian SZ, Zhang H et al. Phase hologram optimization with bandwidth constraint strategy for speckle-free optical reconstruction. *Opt Express* **29**, 11645–11663 (2021).
20. Lee B, Kim D, Lee S et al. High-contrast, speckle-free, true 3D holography via binary CGH optimization. *Sci Rep* **12**, 2811 (2022).
21. Wu Y, Wang J, Chen C et al. Generation of phase-only fourier hologram based on double phase method and quantization error analysis. *IEEE Photonics J* **12**, 7000913 (2020).
22. Yu X, Zhang HM, Zhao Z et al. On the use of deep learning for computer-generated holography. *iScience* **28**, 112507 (2025).
23. Horisaki R, Takagi R, Tanida J. Deep-learning-generated holography. *Appl Opt* **57**, 3859–3863 (2018).
24. Lee J, Jeong J, Cho J et al. Deep neural network for multi-depth hologram generation and its training strategy. *Opt Express* **28**, 27137–27154 (2020).
25. Liu SC, Chu DP. Deep learning for hologram generation. *Opt Express* **29**, 27373–27395 (2021).
26. Shi L, Li BC, Kim C et al. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).
27. Wu JC, Liu KX, Sui XM et al. High-speed computer-generated holography using an autoencoder-based deep neural network. *Opt Lett* **46**, 2908–2911 (2021).
28. Shui XH, Zheng HD, Xia XX et al. Diffraction model-informed neural network for unsupervised layer-based computer-generated holography. *Opt Express* **30**, 44814–44826 (2022).
29. Liu KX, Wu JC, He ZH et al. 4K-DMDNet: diffraction model-driven network for 4K computer-generated holography. *Opto-Electron Adv* **6**, 220135 (2023).
30. Zheng HD, Peng JC, Wang Z et al. Diffraction model-driven neural network trained using hybrid domain loss for real-time and high-quality computer-generated holography. *Opt Express* **31**, 19931–19944 (2023).
31. Zhong CL, Sang XZ, Yan BB et al. Real-time high-quality computer-generated hologram using complex-valued convolutional neural network. *IEEE Trans Vis Comput Graph* **30**, 3709–3718 (2024).
32. Yuan GZQ, Zhou M, Liu F et al. Physics-aware cross-domain fusion aids learning-driven computer-generated holography. *Photonics Res* **12**, 2747–2756 (2024).
33. Wang X, Liu XL, Jing T et al. Phase-only hologram generated by a convolutional neural network trained using low-frequency mixed noise. *Opt Express* **30**, 35189–35201 (2022).
34. Dong ZX, Xu C, Ling YY et al. Fourier-inspired neural module for real-time and high-fidelity computer-generated holography. *Opt Lett* **48**, 759–762 (2023).
35. Yu GW, Wang J, Yang H et al. Asymmetrical neural network for real-time and high-quality computer-generated holography. *Opt Lett* **48**, 5351–5354 (2023).
36. Liu QW, Chen J, Yao YW et al. Frequency aware high-quality computer-generated holography via multilevel wavelet learning and channel attention. *Opt Lett* **49**, 5559–5562 (2024).
37. Song XL, Dong JQ, Liu MH et al. Real-time intelligent 3D holographic photography for real-world scenarios. *Opt Express* **32**, 24540–24552 (2024).
38. Yan XP, Liu XL, Li JQ et al. Recoding double-phase holograms with the full convolutional neural network. *Opt Laser Technol* **174**, 110667 (2024).
39. Conklin AA, Kumar S. Solving the big computing problems in the twenty-first century. *Nat Electron* **6**, 464–466 (2023).
40. Fang QW, Zheng HD, Xia XX et al. Diffraction model-driven neural network with semi-supervised training strategy for real-world 3D holographic photography. *Opt Express* **32**, 45406–45420 (2024).
41. Szegedy C, Liu W, Jia YQ et al. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (IEEE, 2015). http://doi.org/10.1109/CVPR.2015.7298594.
42. Szegedy C, Vanhoucke V, Ioffe S et al. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2818–2826 (IEEE, 2016). http://doi.org/10.1109/CVPR.2016.308.
43. 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–MICCAI 2015* 234–241 (Springer, 2015). http://doi.org/10.1007/978-3-319-24574-4_28.
44. He KM, Zhang XY, Ren SQ et al. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on*

*Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016). http://doi.org/10.1109/CVPR.2016.90.

45. Matsushima K, Shimobaba T. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Opt Express* **17**, 19662–19673 (2009).
46. Jiang LM, Dai B, Wu W et al. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 13899–13909 (IEEE, 2021). http://doi.org/10.1109/ICCV48922.2021.01366.
47. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In *Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers* 1398–1402 (IEEE, 2003). http://doi.org/10.1109/ACSSC.2003.1292216.
48. Loshchilov I, Hutter F. Decoupled weight decay regularization. In *Proceedings of the 7th International Conference on Learning Representations* (OpenReview. net, 2019).
49. Miangoleh SMH, Dille S, Mai L et al. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9680–9689 (IEEE, 2021). http://doi.org/10.1109/CVPR46437.2021.00956.
50. Zhu LH, Liao BC, Zhang Q et al. Vision mamba: efficient visual representation learning with bidirectional state space model. In *Proceedings of the Forty-first International Conference on Machine Learning* (OpenReview. net, 2024).
51. Mildenhall B, Srinivasan PP, Tancik M et al. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun ACM* **65**, 99–106 (2022).
52. Wang D, Li ZS, Zheng Y et al. Liquid lens based holographic camera for real 3D scene hologram acquisition using end-to-end physical model-driven network. *Light Sci Appl* **13**, 62 (2024).
53. Li ZS, Liu C, Li XW et al. Real-time holographic camera for obtaining real 3D scene hologram. *Light Sci Appl* **14**, 74 (2025).
54. Li ZS, Sun JS, Fan Y et al. Deep learning assisted variational Hilbert quantitative phase imaging. *Opto-Electron Sci* **2**, 220023 (2023).
55. Ha YL, Luo Y, Pu MB et al. Physics-data-driven intelligent optimization for large-aperture metalenses. *Opto-Electron Adv* **6**, 230133 (2023).
56. Zhu RC, Wang JF, Qiu TS et al. Direct field-to-pattern monolithic design of holographic metasurface via residual encoder-decoder convolutional neural network. *Opto-Electron Adv* **6**, 220148 (2023).
57. Zhang RZ, Guo YH, Zhang F et al. Dual-layer metasurface enhanced capacity of polarization multiplexing. *Laser Photonics Rev* **18**, 2400126 (2024).
58. Ha YL, Wang LJ, Guo YH et al. Monolithic integrated multi-parameter light-field sensor based on tailored disordered nanostructures. *Adv Opt Mater* **13**, 2402509 (2025).

## Acknowledgements

We are grateful for financial supports from National Natural Science Foundation of China (Grant No. 62205117, 52275429), National Key Research and Development Program of China (Grant No. 2021YFF0502700), Young Elite Scientists Sponsorship Program by CAST (Grant No. 2022QNRC001), West Light Foundation of the Chinese Academy of Sciences (Grant No. xbzg-zdsys-202206), Hubei Natural Science Foundation Innovative Research Group Project (2024AFA025).

## Author contributions

X. Yu, W. Xiong and H. Gao conceived the study; X. Yu developed and trained IncepHoloRGB; X. Yu, Z. L. Teng, X. H. Fan designed and built the experimental system; X. Yu, T. C. Liu performed the depth estimation method and established DF2K RGBD dataset; X. Yu, Z. L. Teng conducted the experiments and tested IncepHoloRGB. All authors discussed and commented on the manuscript.

## 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.2025.250130

Scan for Article PDF