Research Article

ADVANCED IMAGING

DWS-Net: a depth-wise separable convolutional neural network for robust phase-only hologram encoding

Shu-Feng Lin,[a] Jingwei Chen,[a] Dayong Wang,[a] Jie Zhao,[a] Lu Rong,[a] Yunxin Wang,[a] Yu Zhao,[b,*] and Chao Ping Chen[c,*]

[a]School of Physics and Optoelectronic Engineering, Beijing University of Technology, Beijing, China
[b]College of Information Engineering, Yangzhou University, Yangzhou, China
[c]State Key Laboratory of Avionics Integration and Aviation System-of-Systems Synthesis, School of Integrated Circuits, Shanghai Jiao Tong University, Shanghai, China

Abstract. Neural-network-based computer-generated hologram (CGH) methods have greatly improved computational efficiency and reconstruction quality. However, they will no longer be suitable when the CGH parameters change. A common phase-only hologram (POH) encoding strategy based on a neural network is presented to encode POHs from complex amplitude holograms for different parameters in a single training. The neural network is built up by introducing depth-wise separable convolution, residual modules, and complex-value channel adaptive modules, and it is randomly trained with traditional training input image datasets, a built wavelength pool, and a reconstructed distance pool. The average peak signal-to-noise ratio/structural similarity index measure (PSNR/SSIM) for the proposed network encoded POHs can reach 29.19 dB/0.83, which shows 117.35% and 144.12% improvement compared with the double-phase encoding strategy. The variances of the PSNR and SSIM in different reconstructed distances and different wavelengths are increased by 86.83% and 80.65%, respectively, compared with traditional networks in this strategy. Such a method makes it possible to encode multiple complex amplitude holograms of arbitrary wavelength and arbitrary reconstructed distance without the need for retraining, which is friendly to digital filtering or other operations within the CGH generation process for CGH designing and CGH debugging.

Keywords: computer-generated hologram; complex amplitude hologram encoding; convolutional neural network.

Received May 27, 2025; revised manuscript received Aug. 9, 2025; accepted Aug. 14, 2025; published online Sep. 22, 2025.

© The Authors. Published by Chinese Laser Press under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: 10.3788/AI.2025.10012]

1. Introduction

Holography can modulate arbitrary three-dimensional light fields using compact planar devices, which has been widely used in the applications of three-dimensional displays[1–3], near-eye displays[4], light field manipulation[5], etc. However, the computer-generated holograms (CGHs) based on accurate calculations of diffraction propagation have complex values, which should be encoded into the amplitude or phase value to be loaded on spatial light modulators (SLMs) for actual light modulation.

Traditional CGH generation methods typically include ray-tracing-based methods[6], point-based methods[7], layer-based propagation methods[8], etc. Such methods figure out complex amplitude holograms, which makes it impossible for them to be loaded onto the SLMs directly. Therefore, complex encoding is needed to process the complex amplitude holograms, as shown in Fig. 1(a). First, the detour phase method is initially proposed to encode both the amplitude and phase information by adjusting the size and position of sub-cell apertures for a complex-amplitude encoding[9]. However, such a method needs multiple pixels as a single sub-cell for dynamic display using current SLMs, which causes narrow diffraction angles and low diffraction efficiency. The down-sampling method is also proposed to

*Address all correspondence to Yu Zhao, zhaoyu@yzu.edu.cn; Chao Ping Chen, ccp@sjtu.edu.cn

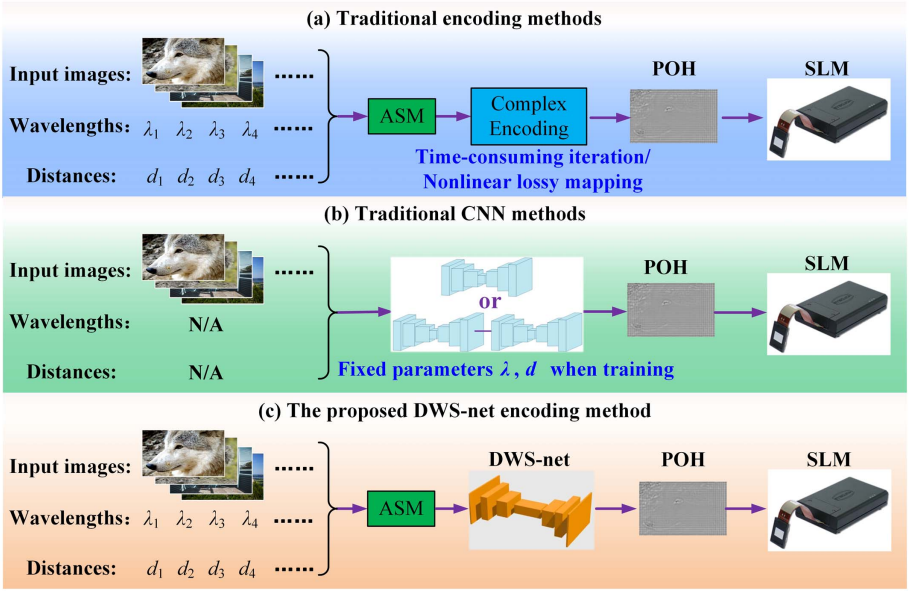

Fig. 1 Phase-only hologram generation strategy of (a) the traditional encoding methods, (b) the traditional CNN methods, and (c) the proposed DWS-Net encoding method.

realize phase-only hologram (POH) encoding by alleviating the edge effect of kinoform hologram reconstruction to increase the high-frequency component[10]. But this approach decreases the resolution of the input image. Additionally, the double-phase method encodes the complex amplitude into two sets of phase information and utilizes a checkerboard form for the phase-only encoding[11]. However, the periodic arrangement of the double phase increases the sampling interval, leading to periodic extension of the reconstructed images and diffraction energy dispersal. Besides, the error diffusion method encodes POHs by diffusing the pixel errors between the POH and complex amplitude hologram to adjacent pixels[12], which increases the computation time in encoding due to the pixel-by-pixel iteration. In addition to the direct phase encoding methods, the GS algorithm[13] can achieve POH generation by constraining the input–output of diffraction propagation. Furthermore, phase retrieval methods based on gradient descent[14] or Wirtinger[15] can directly generate POHs, which can obtain a reconstruction effect close to the complex amplitude holograms. While these types of approaches are flexible enough to encode complex amplitude holograms, their application is limited by the encoding efficiency caused by the time-consuming iterative process or the reconstruction quality caused by the nonlinear lossy mapping.

With the advancement of computer technology, convolutional neural networks (CNNs) have rapidly developed and been introduced to directly calculate POHs by an optimized network[16]. Therefore, CNN-based hologram algorithms, such as convolutional residual networks[17], HoloNet[18], tensor holography[19], and holoencoders[20], have been gradually developed. Such methods establish the neural network and train it by supervised or unsupervised methods, which can directly generate POHs rapidly by achieving nonlinear transformation of input images. The CNN-based methods increase the generation speed of CGHs from seconds to milliseconds. To increase the reconstruction quality of the CNN-based method, the complex-valued convolutional neural network (CCNN) is introduced to generate the CGHs by directly processing complex amplitude values[21]. In addition, more and more functional networks have been proposed, such as the multi-depth hologram, 3D hologram generation, and color hologram generation. The multi-depth hologram generation network[22], depth generative holography network[7], and other networks[23–25] have been successively proposed to directly generate POHs of multiple-depth scenes and even holograms of 3D scenes. The tensor holography network is proposed by utilizing large-scale datasets and occlusion-aware point-based mapping to address high computational costs and occlusion issues, achieving real-time, high-resolution 3D color POH synthesis[19]. Moreover, some researchers combine the neural network-based CGH method with the Taylor Rayleigh–Sommerfeld point cloud gridding (PCG) method to achieve 47% faster computation than traditional PCG on GPUs[26]. Although the integration of more functions can be achieved through the nonlinear transformation of neural networks with a superior performance, the current network usually needs to be trained for specific scenarios. The CGH parameters, such as wavelengths and distances, are fixed in the training process, as shown in Fig. 1(b). It will face a retraining problem when the CGH parameters of the hologram generation process are changed or if a digital filtering operation is performed in the CGHs. Digital filtering or other operations are usually performed on the hologram in the actual holographic display, which also results in the generated POH becoming a complex amplitude hologram again.

To enhance the adaptability of CNN-based algorithms, most features of the neural network to generate holograms are stripped except for the function of encoding the complex amplitude holograms into POHs with high generalization capabilities in this work. A common POH encoding method is presented by establishing a depth-wise separable convolutional neural network (DWS-Net), as shown in Fig. 1(c). When the network is applied in practical processing, it can encode the

complex amplitude hologram not only with different input images but also with different wavelengths and propagation distances. The DWS-Net is built up by introducing the depth-wise separable convolution (DWSC), residual modules (RMs), and complex-value channel adaptive modules (CCAMs). Based on the traditional CCNN architecture, the DWSC is introduced to extract features and reduce network parameters in the down-sampling process, and the RM and CCAM are introduced to optimize the up-sampling process. The introduction of such modules could enhance the ability of feature extraction from the complex-amplitude hologram across different wavelengths and reconstructed distances. In the training process, diffraction propagation-based physical models are employed before and after the input and output of the network for unsupervised training. By constructing a wavelength pool and a reconstructed distance pool, the random training mechanism is adopted to randomly select each value in the corresponding pools for each training image in each epoch. Since only three values are primary parameters in full-color display, but infinite values are possible for the reconstructed distance, a different strategy is employed for the corresponding pool construction, which is complete construction for the wavelength pool and partial construction for the reconstructed distance pool. Due to the generalization ability of networks, by the random training mechanism, the number of values in those pools could be a limited quantity covering a desired range rather than all the values that are possible. Then, the proposed method enables the neural network as a common POH encoder processing complex amplitude holograms with arbitrary wavelengths and arbitrary reconstructed distances by just one training, which can effectively improve the generalization ability of neural networks and reduce the risk of retraining in hologram processing.

2. Method for the Hologram Encoding

2.1. Network Architecture

The network architecture of the proposed DWS-Net for POH encoding is illustrated in Fig. 2, which follows the typical U-Net architecture of the CCNN. It consists of three identical down-sampling operations and three diverse up-sampling operations.

Each down-sampling operation employs the improved DWS-Net, which is composed of depth-wise convolution and point-wise convolution, as shown by the blue color and magenta color progresses in Fig. 2, respectively. In the first down-sampling operation, the depth-wise convolution expands the single input channel into 16 output channels using 16 independent kernels for extracting diverse local features. Subsequently, the point-wise convolution uses a 1×1 kernel to fuse the information from these 16 channels for enhancing the feature representation. In the second down-sampling operation, the depth-wise convolution distributes the 16 input channels into groups, with each group using two independent kernels to expand the output to 32 channels for refining and enriching the features. The point-wise convolution executes the same processing as it does in the first down-sampling process. The third down-sampling operation repeats the process of the second down-sampling operation and finally fuses the information from 64 channels.

In the up-sampling operations, the RM and CCAM are introduced to enhance the feature extraction capabilities for different wavelengths and different reconstructed distances in this work. The first up-sampling operation is first executed by the complex-value deconvolution (C-DeConv), as shown by the yellow progress in Fig. 2, which compresses the channels of feature maps and expands the resolution of the feature maps. The

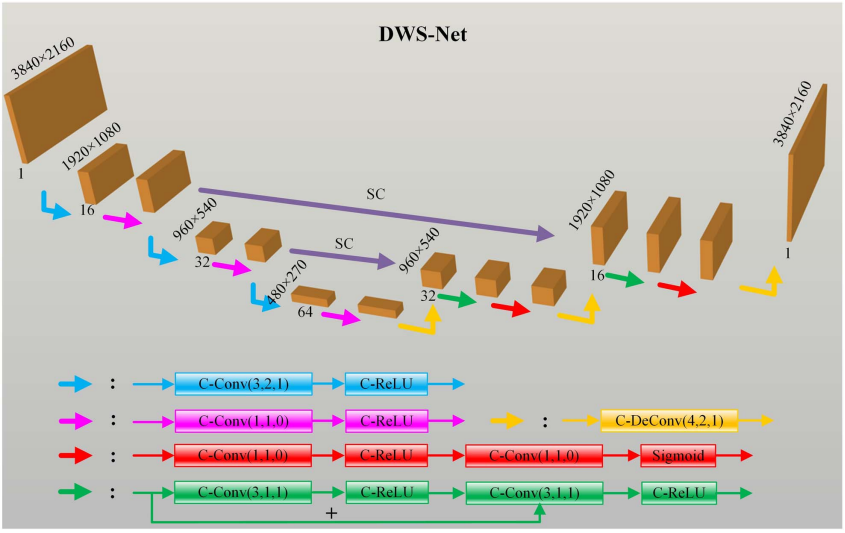

Fig. 2 Architecture diagram of the proposed DWS-Net for POH encoding, including the depth-wise convolution process of down-sampling (in blue color), point-wise convolution process of down-sampling (in magenta color), complex de-convolution process in up-sampling (in yellow color), introduced RM in up-sampling (in green color), and introduced CCAM in up-sampling (in red color).

expanded feature maps are fused with the third layer feature maps in the down-sampling operations via a skip connection. Afterward, the fused feature maps are processed by a residual block structure to reduce the performance degeneration of feature extraction in a deep net, as shown by the green progress in Fig. 2. Thereafter, the CCAM introduced here is used to enhance the weight for each significant feature, as shown by the red progress in Fig. 2. For the second up-sampling operation, the feature maps undergo the same processing with the first up-sampling operation. For the third up-sampling operation, a single-channel phase-only feature map is compressed by just a single C-DeConv similar to the previous up-sampling operation, which produces a POH output with the same resolution as the input complex amplitude hologram.

The entire down-sampling process gradually compresses the input features and extracts multi-scale key information from the complex amplitude hologram. The up-sampling process effectively maintains the intricate relationship between amplitude and phase information through optimization and fusion. This proposed DWS-Net-based method significantly improves the generalization ability of the network and provides more stable phase recovery of the POH.

The RM is implemented by combining a 3×3 C-DeConv, a complex-value-ReLU (C-ReLU) activation function, a 3×3 complex-value convolution (C-Conv), and a skip connection, as shown by the green progress in Fig. 2. When processing the complex-valued features, the RM can reduce the performance degeneration of feature extraction by skip connecting with the input of this RM. In this way, the network ensures that both the amplitude and phase information of the complex-value signal are retained and propagated. This allows the network to better learn the complex relationships between amplitude and phase.

The CCAM, shown by the red progress in Fig. 2, contains a 1×1 C-Conv process, a C-ReLU process, a 1×1 C-Conv process again, and a sigmoid activation process, sequentially. The key process of the CCAM is the sigmoid nonlinear transformation to allocate higher weight to the more important feature. Here, we customize a complex-value sigmoid function to distribute the weight, which is defined as

$$\mathrm{Sigmoid}(X) = \frac{1}{1 + \exp(-|X|)}, \tag{1}$$

where $X = \alpha + j\beta$, and α and β represent the real and imaginary parts of current complex-value feature, respectively. Then, the input of the sigmoid function is the modulus of X, which is calculated as

$$|X| = \sqrt{\alpha^2 + \beta^2}. \tag{2}$$

Then, it can map the modulus into [0,1]. Such a process enables the network to focus more on features with a smaller modulus of each complex-value feature to weaken the impact of these features. This mechanism is similar to channel attention, but it retains the intricate relationship between amplitude and phase, making it particularly suitable for feature optimization and extraction in hologram encoding tasks involving complex amplitude signals.

2.2. Training Strategy of the Proposed Network

Traditional neural network-based methods typically introduce the angular spectrum method (ASM) as the diffraction reconstruction module to optimize the network parameters of the specially designed single neural network or multiple cascading networks, as shown in the lower part of Fig. 3(a). The nonlinear transformation from input to output is determined by the CGH parameters fed to the ASM, which are usually fixed parameters. When the network is trained, we can directly use this network to predict a corresponding POH by inputting the target image, as shown in the upper part of Fig. 3(a). When the CGH parameters change or digital filtering is introduced, the trained network should be retrained. This is a common situation when designing holograms and debugging holograms. The proposed method strips most of the functions except the function of POH encoding from complex amplitude holograms. When the network is trained, the POHs can be directly predicted by feeding with the complex amplitude holograms, as shown in the upper part of Fig. 3(b).

The training strategy of the proposed DWS-Net is illustrated in the lower part of Fig. 3(b). The propagation models of forward and backward ASM are introduced to generate complex amplitude holograms and reconstruct the POH predicted by the DWS-Net for the network training, respectively. The corresponding input image in the DWS-Net training is the 2D image of the DIV2K training dataset. Multiple CGH parameters are constructed into two parameter pools, and the CGH parameters for forward and backward ASM are randomly selected from the parameter pools. The error loss is obtained by comparing the reconstructed image of the predicted POH with the original

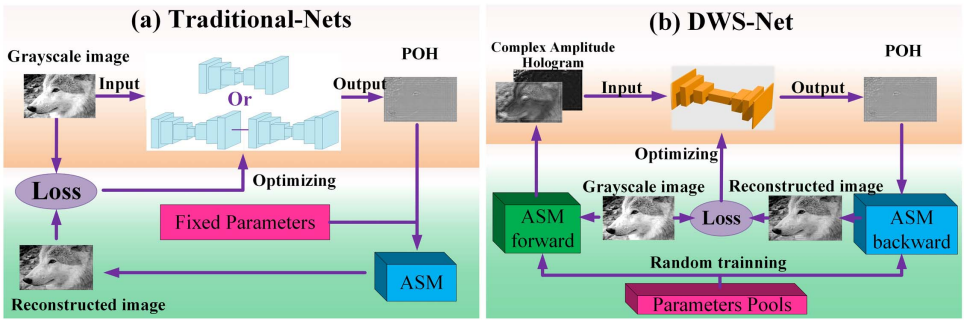

Fig. 3 Network and training strategy comparison of (a) the traditional methods and (b) the proposed method. The upper parts are the usage, and the lower parts are the training process.

2D image of the dataset. The network is optimized using the Adam optimizer based on the error loss.

First, the forward ASM for generating the complex amplitude holograms corresponding to the input of the DWS-Net is carried out as follows:

$$CH = F^{-1}[F(I_0) \cdot H_1], \tag{3}$$

where CH is the complex amplitude hologram after ASM propagation, I_0 is the complex amplitude input with the training set image as amplitude and a zero matrix as phase, F represents the 2D Fourier transform, and F^{-1} represents the 2D inverse Fourier transform. H_1 represents the transfer function of ASM, which is calculated as follows:

$$H_1 = \begin{cases} \exp\left(jkz_1\sqrt{(1-\lambda^2 f_x^2-\lambda^2 f_y^2)}\right), & \text{if } \sqrt{f_x^2+f_y^2} < \frac{1}{\lambda} \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where k denotes the wavenumber and z_1 represents the propagation distance. λ indicates the wavelength, while f_x and f_y represent the spatial frequencies along the x and y directions, respectively. To simulate the accurate diffraction propagation process, the zero-padding and band-limitation processes are applied before the Fourier transform to perform diffraction calculations.

Second, the calculated complex amplitude hologram is fed into DWS-Net for POH prediction. After iterative optimization and training, the DWS-Net could learn and extract the features from the complex amplitude holograms. A high reconstruction quality POH can be transformed from its corresponding complex-value CGH by the proposed DWS-Net.

Third, the POH predicted by the DWS-Net is then backward propagated using the ASM algorithm for reconstruction, which is the inverse process of the first process.

Fourth, the error loss is calculated by comparing the difference of the intensity of the reconstructed image with the corresponding training set input image. The loss function used in this work is the mean squared error (MSE). Let P and P_{target} represent the reconstructed image of the POH predicted by the DWS-Net and the original image from the training dataset, respectively, both having the same dimensions. The MSE is computed as follows:

$$\text{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(P_{ij} - P_{\text{target}_{ij}}\right)^2, \tag{5}$$

where m is the height (number of rows) of the image, n is the width (number of columns) of the image, and P_{ij} and $P_{\text{target}_{ij}}$ present the grayscale values of the images P and P_{target} at the coordinates of (i, j).

Finally, based on the error loss, the gradients of each network parameter are calculated through backpropagation of the CNN, and the DWS-Net is optimized by the Adam optimizer.

It is worth noting that random training for different wavelengths and different reconstructed distances is adopted to enhance the generalization ability of the network for complex amplitude hologram encoding. In this work, a wavelength vector pool for the red, green, and blue (RGB) color channels, as well as a reconstructed distance vector pool, is set up for random training. The value number in the reconstructed distance vector pools could be a limited quantity covering a desired range rather than listing all the values that are possible. For each training image in every epoch, different wavelengths and reconstructed distances are randomly selected in their corresponding pools. Then, different complex amplitude holograms corresponding to different wavelengths and reconstructed distances are fed into the DWS-Net for random training, which could increase the generalization capability for different cases of encoding.

3. Simulation and Experimental Results

To verify the effectiveness of the proposed method, the numerical simulation and optical experiment are carried out in this section. This work utilizes the DIV2K dataset to construct the training set for neural network training, comprising 700 training images and 100 test images (800 total). The training wavelength vector pool is set with three values of 473, 532, and 660 nm, corresponding with the optical setup. And the reconstructed distance vector pool is set with 11 equally spaced values in a range from 180 to 230 mm. The DWS-Net training is conducted on a workstation equipped with an Intel(R) Xeon(R) Gold 6226R CPU at 2.90 GHz (dual processors), 128 GB of RAM, and an NVIDIA RTX A6000 GPU (47.5 GB of VRAM), running the Windows 11 operating system. The model's learning rate was set to 0.001, and training for 40 epochs took approximately 5 h.

To facilitate validation, the CGH parameters are the same both in the digital reconstruction simulation of the POH encoded by DWS-Net and the optical experiments for holographic displays, which are based on the holographic display prototype as shown in Fig. 4. The RGB light sources are provided by three lasers with wavelengths of 473, 532, and 660 nm. A polarizer is used to make sure there is a uniform polarization state for the illumination. The expanded beam is illuminated onto the SLM (HOLOEYE GAEA-2) for phase modulation, with a pixel pitch of 3.74 μm and a resolution of 3840 × 2160. The encoded POH is loaded on the SLM to modulate the wavefront of the incident light according to a pre-determined optical pattern. Finally, the recorded image is reconstructed at the preset plane captured by a digital camera model NIKON D7100. The digital simulation process corresponding to such an optical process uses angular

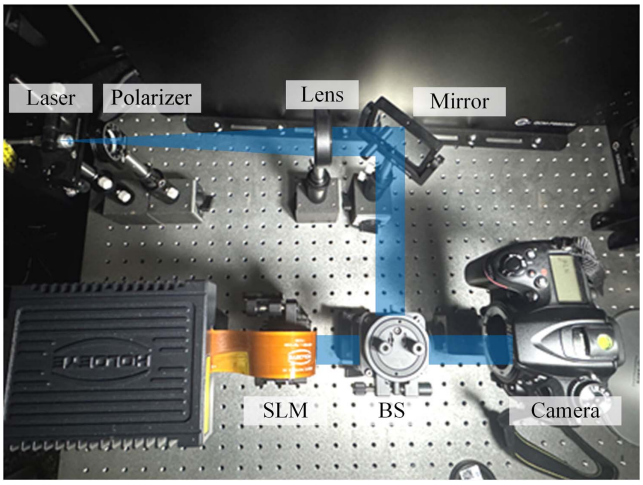

Fig. 4 Schematic diagram of the optical setup for the holographic display prototype.

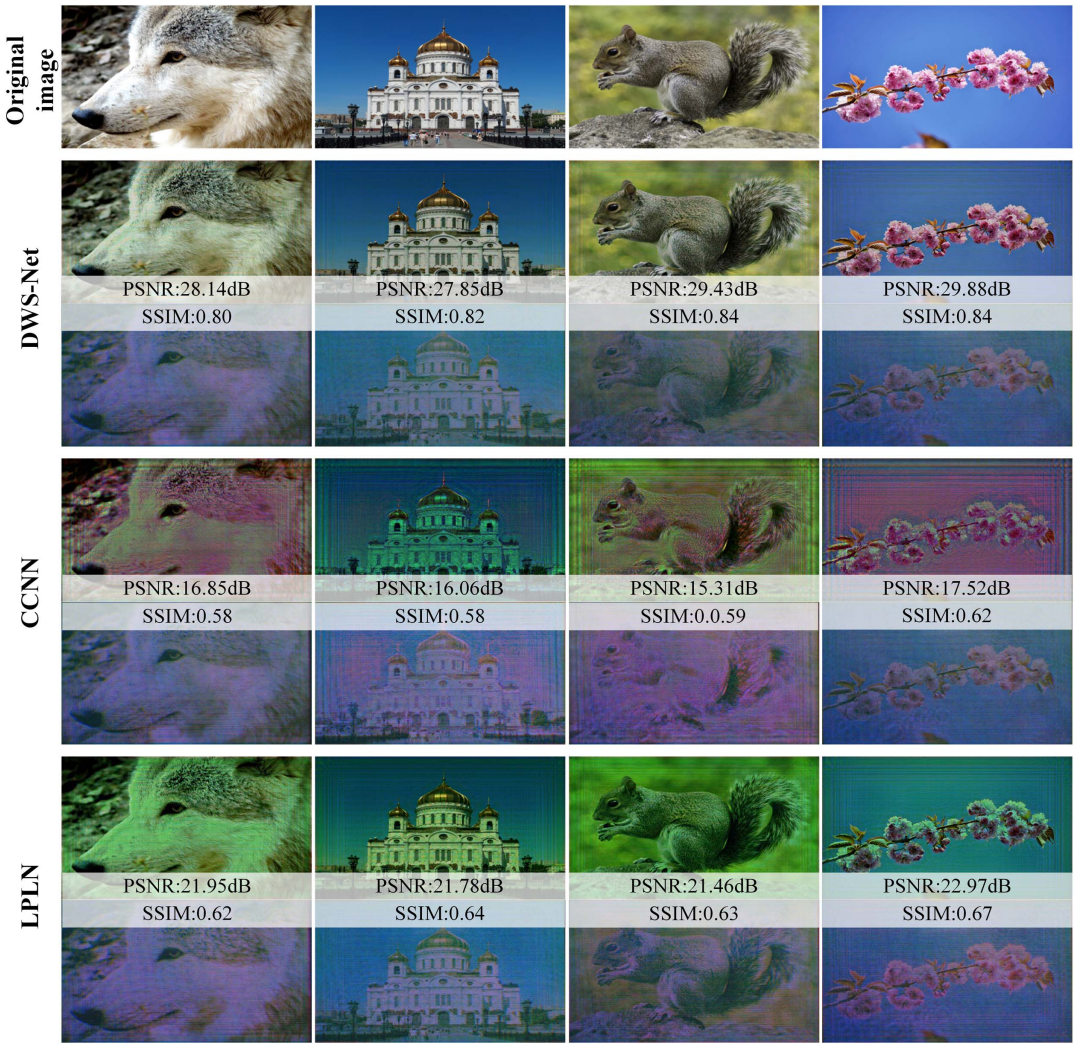

Fig. 5 Comparison of digital simulation (upper rows of each method) and optical experiment (lower rows of each method) of the POH reconstruction encoded by different networks at different wavelengths (blue: 473 nm, green: 532 nm, red: 660 nm) under a fixed distance ($z = 222$ mm).

spectrum diffraction propagation to produce a digitally reconstructed image.

3.1. Evaluation of POH Encoding for the Proposed Network

Because the proposed method is designed to encode a complex-value hologram into a POH, in general, only complex networks can accomplish this task. Therefore, we compare the proposed network with the CCNN and the layered phase learning network (LPLN) in the proposed strategy. Specifically, the comparison tests all three networks under the same input data, training strategies, network-parameter settings, and simulation environment to evaluate their performance in encoding the complex amplitude holograms. The test images for the comparison are two animals in the testing dataset of the DIV2K. The peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used as evaluation metrics.

First, the digital simulation and optical reconstruction of the POHs encoded by different networks in different wavelengths are shown in Fig. 5, in which the reconstructed distance is fixed at 222 mm away from the hologram plane. To evaluate the impact of the image category/content on encoding quality, we selected four sets of images in our test set, and our models had different effects on textures, faces, or sparse geometric patterns. In order to facilitate the observation of the network's joint encoding ability for full-color images, the simulation and optical reconstruction results of RGB channels for four different types of input images are digitally fused into color images. Because the energy ratio of the three channels of the optical experiment is not accurately adjusted, the results of the optical experiment are dark and purple. The PSNR and SSIM values for the three color channels are averaged and marked in Fig. 5. It is easy to find that the average PSNR and SSIM of the reconstruction encoded by the proposed DWS-Net remain at a better level compared with the CCNN and LPLN. Moreover, the encoding quality in each color channel and the equalization across different colors are better than those of the other two methods. However, the reconstruction quality of the encoding task by the traditional CCNN and LPLN is uneven. The joint full-color encoding effect of the traditional CCNN is reddish and purple, which proves that its encoding ability in the green channel is weaker than that of

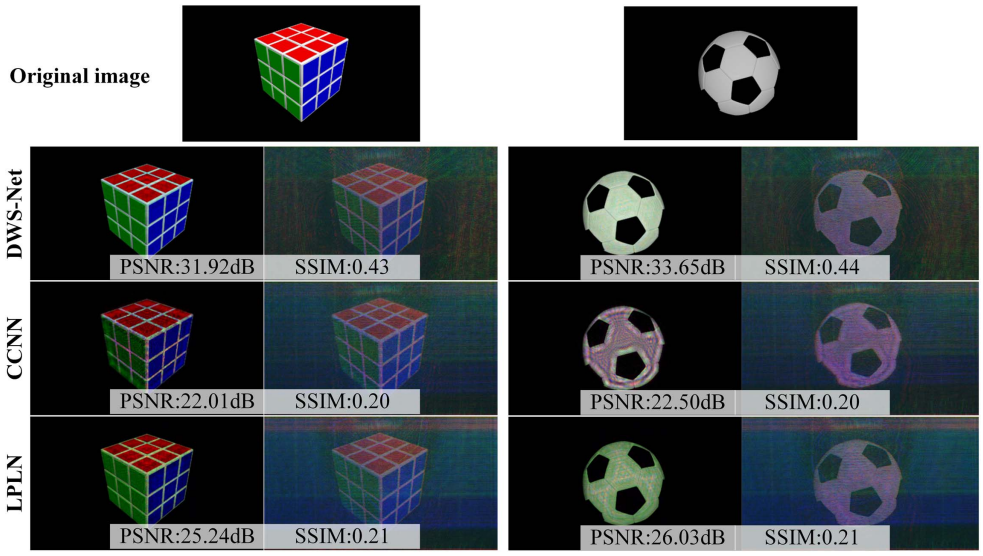

Fig. 6 Comparison of digital simulation (left column of each image) and optical experiment (right column of each image) of full-color encoding on sparse images.

the red and blue channels. The joint full-color encoding effect of the LPLN is yellowish, which proves that its encoding ability in the blue channel is weak. These results confirm the improved reconstruction quality of the proposed DWS-Net in multi-wavelength complex amplitude hologram encoding and the robustness of the three channels.

To further evaluate the encoding effect of the network on a sparse image, two sets of independent patterns without backgrounds are used as input images to generate complex amplitude holograms and encode them into POHs by the encoding network. The reconstructed results are shown in Fig. 6. It can be seen that the average PSNR and SSIM of the three channels reconstructed by the proposed DWS-Net are much higher than those of the traditional CCNN and LPLN methods. The traditional CCNN and LPLN methods not only have poor equilibration in the three channels but also have uneven reconstruction images and seriously poor reconstruction quality in the low-frequency part. To illustrate the effectiveness of the proposed method, we use the blue color channel as examples, and the reconstructed distance range covers from 180 to 230 mm. Figure 7 shows only three random distance results without training. It can be found that the proposed DWS-Net consistently delivers stable, high-quality reconstruction results at various distances. The other two networks exhibit noticeable degradation in reconstruction quality, with lower PSNR and SSIM values and evident image color distortions, indicating a limited capability in handling these distance conditions.

In addition, the detailed PSNR and SSIM of the reconstruction results to the POHs encoded by these methods with different wavelengths and different reconstructed distances are also plotted in Fig. 8. The tested distances are set at 1 mm intervals in the range from 180 to 230 mm, which is much more than the number in the parameter pools under the trained progress. At different test distances along the horizontal axis in each plot of Fig. 8, the PSNR/SSIM values for different wavelengths are represented by RGB dots. To better illustrate the volatility of the reconstructed quality of these methods at different wavelengths and at different reconstructed distances, the average value of each case is plotted with a solid line in each plot. The deviation from the average value line represents the robustness of the coding by this method at different reconstructed distances and at different wavelengths. It can be seen that the average PSNR and SSIM for DWS-Net are as high as 30.83 dB and 0.87, respectively, as shown in Figs. 8(a) and 8(d). It also maintains a small deviation both along different reconstructed distances and different wavelengths. The variances for PSNR and SSIM values are 2.19 and 0.00096, while the average PSNR/SSIM of the traditional CCNN method and LPLN method are 23.15 dB/0.66 and 26.24 dB/16.46, as shown in Figs. 8(b)–8(f). In addition, the values are more decentralized around the average value, and the quantitative variances of PSNR/SSIM for the traditional CCNN method and LPLN method are 24.29/0.00824 and 16.46/0.01872, respectively. That is to say, the PSNR/SSIM robustness of the DWS-Net for different reconstructed distances and wavelengths has been improved 90.98%/88.35% compared to the CCNN and 86.70%/94.87% compared to the LPLN. These results show that DWS-Net not only produces higher reconstruction quality across all wavelengths and distances but also provides more consistent and robust results, whereas there are some worse values both in the CCNN and LPLN at untrained sampling points, even though the LPLN is proposed for multiple reconstructed distances. Therefore, the proposed DWS-Net neural network can encode the complex amplitude hologram with different wavelengths and different reconstructed distances just under a single training, showing stronger adaptability and higher coding efficiency.

3.2. *Evaluation of POH Encoding for the Proposed Strategy*

In the actual holographic display application, some digital filtering or other operations are usually performed on the hologram. As a result, the generated POH becomes a complex amplitude hologram again. Operations such as digital filtering are random, and the hologram generation network cannot be trained for all the digital processing cases. As a result, strategies similar to

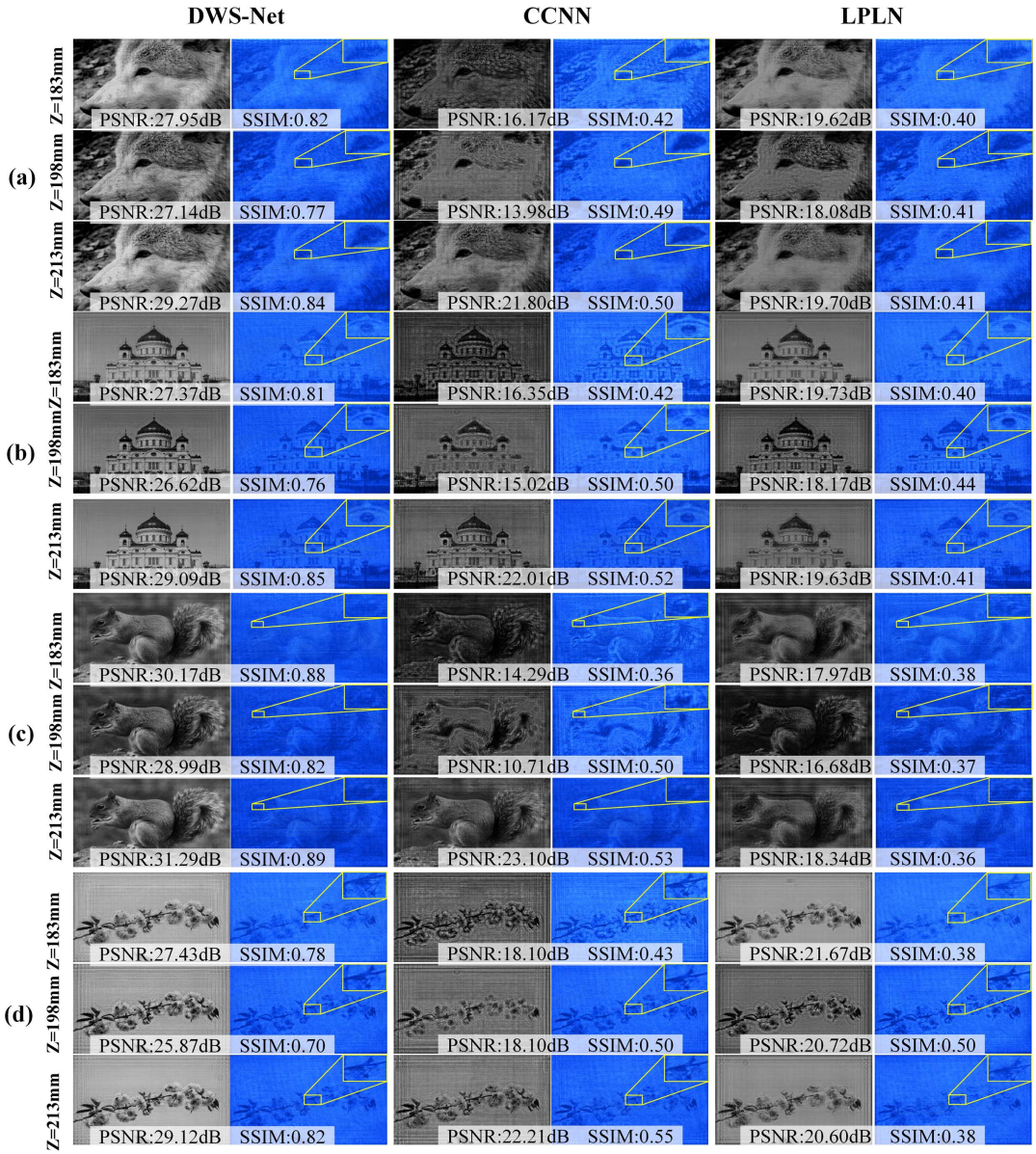

Fig. 7 Comparison of digital simulation and optical experiment of the POH reconstruction encoded by different networks with reconstructed distances at 183, 198, and 213 mm in blue ($\lambda = 473$ nm).

traditional coding methods have to be used to encode the complex amplitude holograms into POHs again. The proposed strategy can directly encode the complex amplitude hologram the same way as the traditional encoding method, no matter if there is a digital process or not, which not only can solve the problem of neural network retraining and relieve the nonlinear lossy mapping of the traditional encoding method but also can improve the reconstruction quality. To verify this performance, the experiments are also carried out.

The iterative methods can theoretically approach the ideal results infinitely on the basis of increasing the iterative time, which is not considered for comparison in this work. Two sets of results for comparing the encoding capacity between the proposed strategy with the double-phase encoding method are shown in Fig. 9. The complex amplitude holograms are first calculated by the ASM and then are transformed into the Fourier domain for digital filtering. Figures 9(a)–9(f) show the three cases of digital filtering of two sets of input images. The whole size of the frequency spectra, shown in Figs. 9(a) and 9(d), represents that the complex amplitude holograms are originally calculated by the ASM. The other spectra overlapped with blue masks are the cases of digital filtering by removing three out of four and five out of six of the outer high frequencies, as shown in Figs. 9(b)–9(f), respectively. Following, the filtered spectra are inversely transformed into the spatial domain as filtered complex amplitude holograms. Figures 9(g)–9(l) are the simulated reconstructed results directly using the complex amplitude holograms inversely transformed from Figs. 9(a)–9(f). It can be seen that the reconstruction quality is deteriorated due to the digital filtering, however, which should be set as the destination for the DWS-Net encoding and double-phase encoding. Because the complex amplitude hologram cannot be loaded on

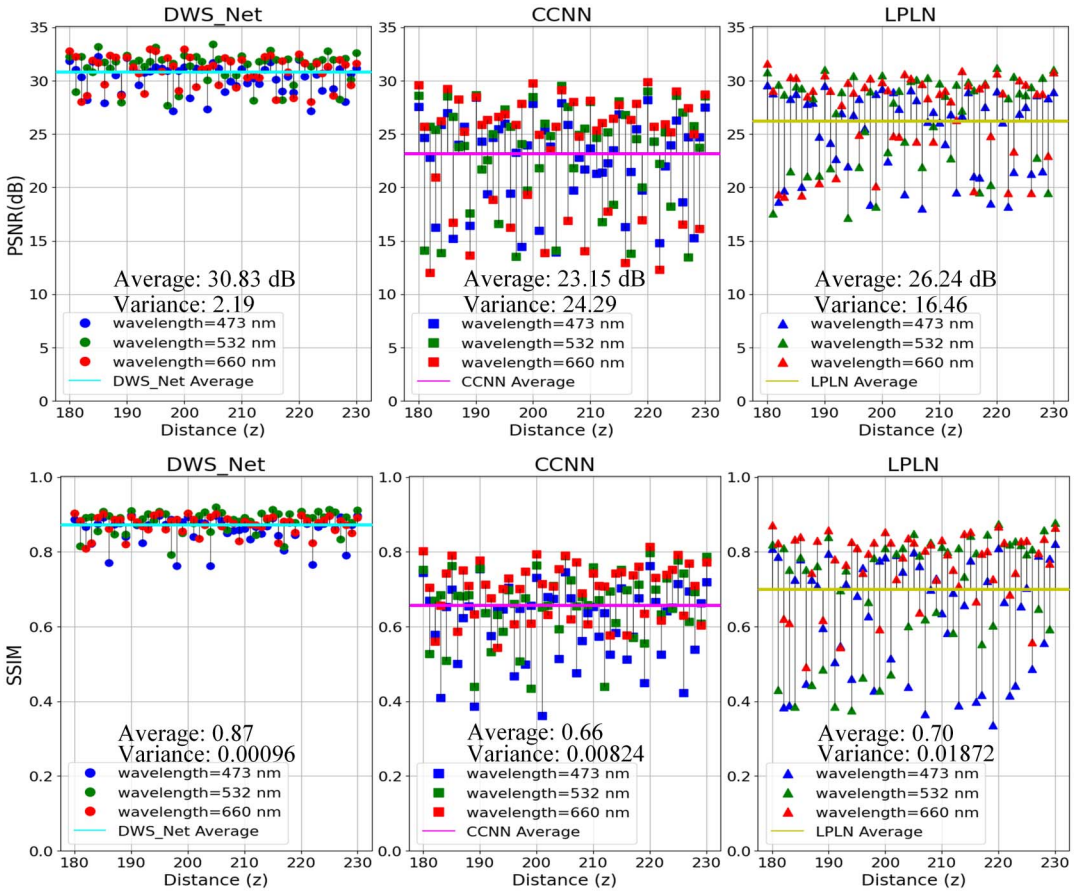

Fig. 8 Average PSNR and SSIM values of POHs corresponding to the two set images in Figs. 5 and 7, which are encoded by different networks with different reconstructed distances and different wavelengths. The straight line in each group represents the average value of all wavelengths and all reconstructed distances and the variance of the PSNR and SSIM was calculated for the two images across all wavelengths and distances.

the SLM, the proposed strategy and the traditional encoding method can overcome this issue. After encoding the complex amplitude holograms, the reconstructed images encoded by DWS-Net are shown in Figs. 9(m)–9(r), and the reconstructed images encoded by double-phase encoding are shown in Figs. 9(s)–9(x). We can find that the reconstruction quality of the POHs encoded by the proposed DWS-Net is almost the same as that of the filtered complex amplitude hologram, and the reconstruction PSNR value between the filtered complex amplitude hologram and the encoded POH by the DWS-Net can exceed 30 dB with an exceeding SSIM 0.90, as labeled on Figs. 9(m)–9(r), while the average reconstruction PSNR and SSIM values for the double-phase encoding are just 13.03 dB and 0.19, respectively, as labeled in Figs. 9(s)–9(x). Therefore, it can be testified that the proposed encoding strategy based on the neural network is effective and has higher encoding performance than that of the traditional double-phase encoding method.

To further verify the encoding performance of the DWS-Net for filtered complex amplitude holograms in different reconstructed distances, the holograms calculated with reconstructed distances from 180 to 230 mm and the wavelength for blue, green, and red colors are evaluated for the proposed DWS-Net. The PSNR and SSIM values are plotted in Fig. 10. It can be found that the PSNR and SSIM values for the two test images remain consistently around 30 dB. This indicates the excellent performance of DWS-Net in the task of encoding digitally filtered complex amplitude holograms. It is noteworthy that the proposed DWS-Net was not specifically trained for the case of digitally filtered complex amplitude holograms. Instead, it was trained on complex amplitude holograms that were not filtered, and it can still effectively handle the encoding task for the digitally filtered holograms.

3.3. Ablation Study

In this section, the ablation experiment is conducted by comparing with four cases. Specifically, the CCNN is the standard convolution without DWSC and RM+CCAM. The proposed DWS-Net contains both the DWSC and RM+CCAM on the standard CCNN. To evaluate the effects of DWSC and RM+CCAM, we designed four cases of the ablation experiment for comparison. The experiments of all the four cases are carried out under identical training conditions, using the same dataset for training. The quantitative evaluation results are given using the average PSNR and SSIM values of the reconstructed images corresponding to the encoded POHs across different wavelengths and various reconstructed distances. The POHs are encoded by the complex

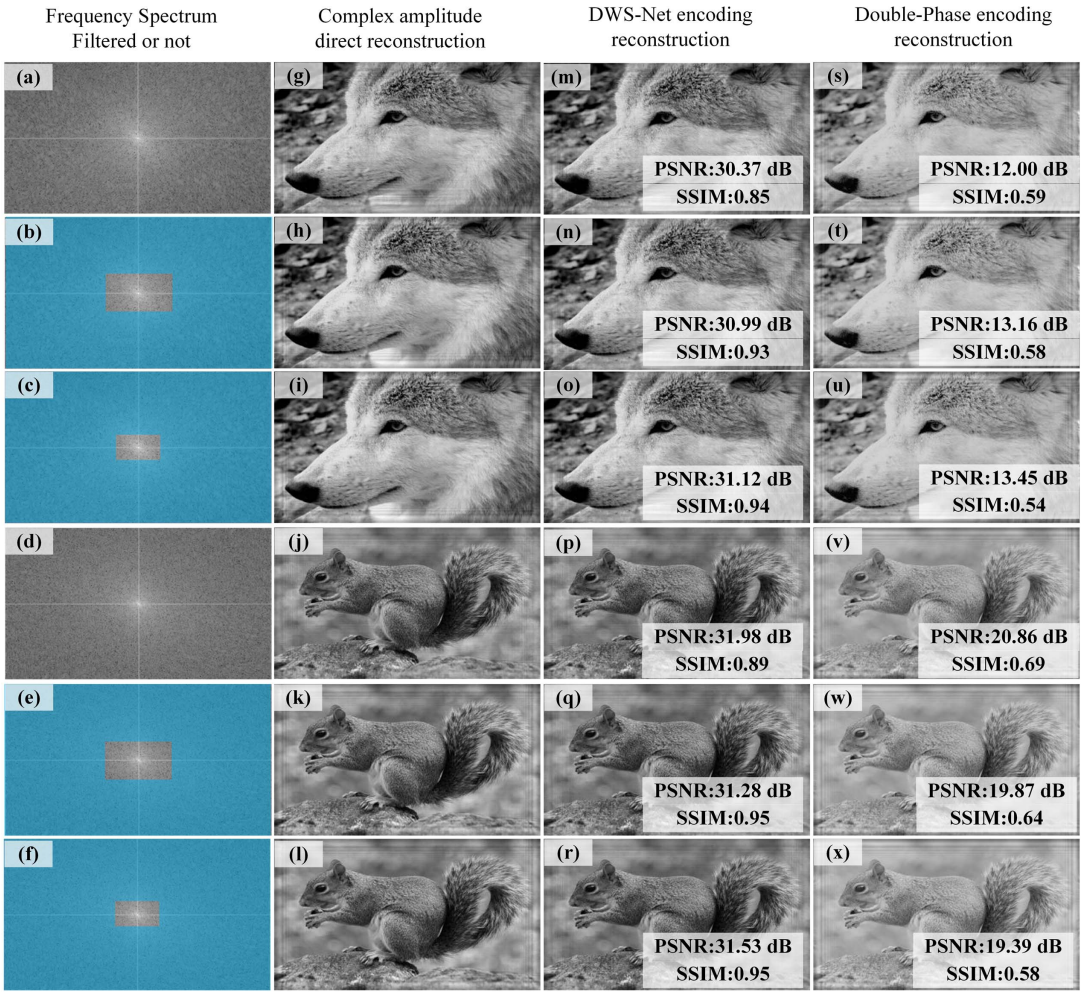

Fig. 9 Simulation results of different encoding capacities on original holograms (the first and fourth rows), and holograms after digital filtering by removing three out of four (the second and fifth rows) and five out of six (the third and sixth rows) of the outer high frequencies, comparing by reconstructed with the original complex amplitude value (the second column), encoded by the DWS-Net (the third column), and the double-phase encoding method (the fourth column).

amplitude holograms of 100 images that are not in the training dataset.

To ensure a fair comparison, all experiments maintain the same network depth, number of channels, and training procedures but only change the network architecture. First, for each test image, we calculate the average PSNR and SSIM values across different reconstructed distances under the red, green, and blue color wavelengths. Then, we compute the overall mean of these averages across all 100 test images, as shown in Table 1. The number of network parameters under different conditions is also counted. To clearly discuss the generalization capability of the four cases under different input images, different wavelengths, and different reconstructed distances, the variances (Var.) of both the PSNR and SSIM are also calculated to reveal the changes in the PSNR and SSIM under different situations rather than a single average value. The results of this evaluation are presented in Table 1. It can be seen that, even though the introduction of RM&CCAM increases the number of network parameters, the reconstruction quality has been improved by a 1 dB improvement in the PSNR and a 0.04 gain in SSIM, while the DWSC can dramatically reduce the number of network parameters to 74.15% compared to the traditional CCNN. In addition, the reconstruction quality has also been improved by 1.66 dB in the PSNR and 0.07 in the SSIM. For the Var., it has a great improvement from 6 to 1.32 in the PSNR and from 0.0031 to 0.0004 in the SSIM. However, there are still low value points along different distances, but this has a little effect on the variance. For the case of CCNN+RM&CCAM+DWSC, which is the proposed DWS-Net, it has almost the same network parameters as the traditional CCNN even though it is integrated with RM&CCAM and DWSC. The average PSNRs and SSIMs achieve 29.19 dB and 0.83, respectively, but also the Var. has been improved to 0.79 and 0.0006 corresponding to the PSNR and SSIM. What is exciting is that the introduction of DWSC has significantly reduced the volatility of the data from 6 to 1.32 and from 15.58 to 0.79 in the PSNR, from 0.0031 to 0.0004 and from 0.0135 to 0.0006 in SSIM, which show 78.00% and 94.93% improvement in the PSNR and 87.10% and 95.56% improvement in the SSIM for the two cases of DWSC integration, respectively. In general, the proposed

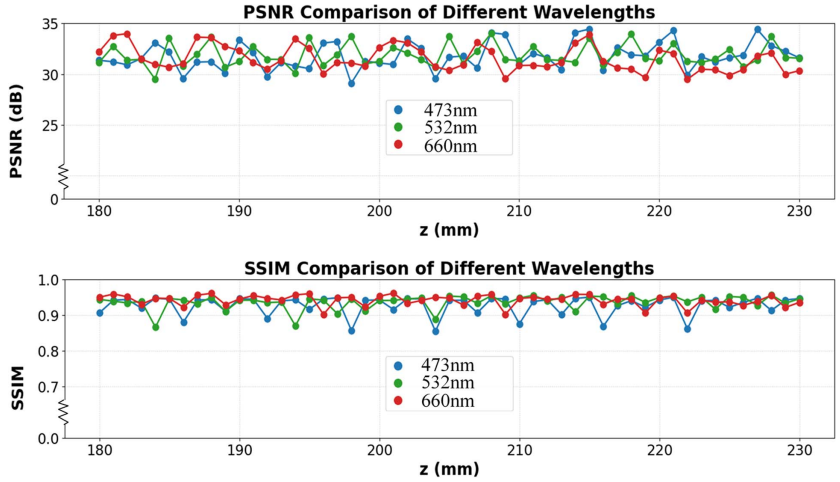

Fig. 10 Average reconstruction (a) PSNR and (b) SSIM values corresponding to the two set images for the DWS-Net encoding the digitally filtered complex amplitude holograms in different reconstructed distances and different wavelengths.

Table 1 Comparison of Network Performance Metrics.

	DWSC	RM&CCAM	Network parameter	PSNR/Var.	SSIM/Var.	Time (ms)
CCNN	×	×	129,122	26.77 dB/6.00	0.75/0.0031	6.46
CCNN+RM&CCAM	×	✓	178,098	27.76 dB/15.58	0.79/0.0135	8.44
CCNN+DWSC	✓	×	95,746	28.43 dB/1.32	0.82/0.0004	14.27
CCNN+RM&CCAM+DWSC (DWS-Net)	✓	✓	144,722	29.19 dB/0.79	0.83/0.0006	16.68

method integrates the DWSC, RM, and CCAM into the traditional CCNNs with a 2.42 dB and 0.08 improvement in a high PSNR and SSIM level and 86.83%/80.65% improvement in the variance of the PSNR/SSIM. These data show that the proposed DWS-Net has high consistency and robustness in POH encoding at different reconstruction distances and different wavelengths. In addition, we evaluated the average inference time required by each network to encode a single image under fixed distance and wavelength conditions. The results show that the conventional CCNN achieves the fastest inference speed, with an average of only 6.46 ms. With the introduction of the DWSC module, the inference time increases to 8.44 ms. Further integration of the RM and CCAM modules raises the time to 14.27 ms. In comparison, our proposed method achieves enhanced performance with an inference time of 16.68 ms. Although the computational cost slightly increases, it results in improved encoding quality and robustness. It can be seen that the average inference time of the proposed DWS-Net is increased by introducing the RM&CCAM and DWSC from 6.64 to 16.68 ms. However, compared with the average inference time of the traditional CCNN and LPLN, 4.93 and 2.24 ms, the increased inference time is acceptable on the basis of the improvement in encoding performance.

4. Conclusion

In this paper, we present a novel network architecture for encoding complex amplitude holograms into POHs. By incorporating grouped DWSC, RM, and CCAM into a traditional CCNN, the network removes most of the redundant features typically used in hologram generation, retaining only the essential capabilities for efficiently encoding complex amplitude holograms into high-generalization POHs. The proposed DWS-Net is a versatile encoding network that requires only a single training with less training data to handle complex amplitude holograms for different wavelengths and reconstructed distances. At varying reconstructed distances, the average reconstruction PSNRs/SSIMs for POHs of 100 test images are 28.49 dB/0.81 in blue, 29.66 dB/0.84 in green, and 29.43 dB/0.83 in red, achieving high-quality, highly generalizable full-color holographic reconstruction. The variance of the reconstruction for 100 POHs can still remain at 0.79 and 0.0006 corresponding to the PSNR and SSIM. For the proposed network, the variances of the PSNR and SSIM are increased by 86.83% and 80.65% compared to the traditional CCNN in this strategy. For the encoding strategy, the PSNR and SSIM of the reconstruction have been improved by 117.35% and 144.12%, respectively, compared with the traditional double-phase encoding method, which demonstrates that the proposed method shows significant improvements in both encoding ability and generalization performance and stability. Especially, the DWS-Net has high robustness of POH encoding at different reconstructed distances and different wavelengths. Furthermore, this approach effectively addresses the retraining issue when the CGH parameters change in traditional neural networks, and it is friendly to digital

filtering or other operations on the CGH generation process for CGH designing and CGH debugging.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62275006, 62205283, 62220106005, and 62175004) and the Ministry of Industry and Information Technology of China (No. GO0300164/001).

References

1. P. A. Blanche, "Holography, and the future of 3D display," *Light Adv. Manuf.* **2**, 446 (2021).
2. C. P. Chen, Y. Su, and C. G. Jhun, "Recent advances in holographic recording media for dynamic holographic display," *J. Opt. Photonics* **1**, 1 (2014).
3. Y. Xiong, *et al.*, "Coherent backlight system for flat-panel holographic 3D display," *Opt. Commun.* **296**, 41 (2013).
4. L. Zhou, *et al.*, "See-through near-eye displays enabling vision correction," *Opt. Express* **25**, 2130 (2017).
5. W. Ouyang, *et al.*, "Ultrafast 3D nanofabrication via digital holography," *Nat. Commun.* **14**, 1716 (2023).
6. T. Nishitsuji, *et al.*, "Review of fast calculation techniques for computer-generated holograms with the point-light-source-based model," *IEEE Trans. Ind. Inf.* **13**, 2447 (2017).
7. P. W. M. Tsang, *et al.*, "Review of fast methods for point-based computer-generated holography," *Photonics Res.* **6**, 837 (2018).
8. H. Pang, *et al.*, "Accurate hologram generation using layer-based method and iterative Fourier transform algorithm," *IEEE Photonics J.* **8**, 1 (2016).
9. B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," *Appl. Opt.* **5**, 967 (1966).
10. H. Ko and H. Y. Kim, "Deep learning-based compression for phase-only hologram," *IEEE Access* **9**, 79735 (2021).
11. Y. Wu, *et al.*, "Generation of phase-only Fourier hologram based on double phase method and quantization error analysis," *IEEE Photonics J.* **12**, 7000913 (2020).
12. K. Liu, Z. He, and L. Cao, "Pattern-adaptive error diffusion algorithm for improved phase-only hologram generation," *Chin. Opt. Lett.* **19**, 050501 (2021).
13. Y. Wu, *et al.*, "Adaptive weighted Gerchberg-Saxton algorithm for generation of phase-only hologram with artifacts suppression," *Opt. Express* **29**, 1412 (2021).
14. S. Liu and Y. Takaki, "Optimization of phase-only computer-generated holograms based on the gradient descent method," *Appl. Sci.* **10**, 4283 (2020).
15. P. Chakravarthula, *et al.*, "Wirtinger holography for near-eye displays," *ACM T. Graphic.* **38**, 213 (2019).
16. Y. Wang, *et al.*, "Hardware implementations of computer-generated holography: a review," *Opt. Eng.* **59**, 102413 (2020).
17. R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," *Appl. Opt.* **57**, 3859 (2018).
18. Y. Peng, *et al.*, "Neural holography with camera-in-the-loop training," *ACM. T. Graphic.* **39**, 1 (2020).
19. L. Shi, *et al.*, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature* **591**, 234 (2021).
20. J. Wu, *et al.*, "High-speed computer-generated holography using an autoencoder-based deep neural network," *Opt. Lett.* **46**, 2908 (2021).
21. C. Zhong, *et al.*, "Real-time high-quality computer-generated hologram using complex-valued convolutional neural network," *IEEE Trans. Vis. Comput. Graph.* **30**, 3709 (2023).
22. J. Lee, *et al.*, "Deep neural network for multi-depth hologram generation and its training strategy," *Opt. Express* **28**, 27137 (2020).
23. C. Zhong, *et al.*, "Real-time 4K computer-generated hologram based on encoding conventional neural network with learned layered phase," *Sci. Rep.* **13**, 19372 (2023).
24. X. Yan, *et al.*, "Generation of multiple-depth 3D computer-generated holograms from 2D-image-datasets trained CNN," *Adv. Sci.* **12**, 2408610 (2024).
25. X. Yan, *et al.*, "Recoding double-phase holograms with the full convolutional neural network," *Opt. Laser Technol.* **174**, 110667 (2024).
26. Y. Zhao, *et al.*, "Holographic voice-interactive system with Taylor Rayleigh-Sommerfeld based point cloud gridding," *Opt. Laser Eng.* **179**, 108270 (2024).