

Super-resolution reconstruction of the holographically reconstructed image based on improved ESRGAN

Yuan Chen^{1,2,*}, Sitian Li³, Hongwei Ma^{2,3}, Peichao Li³, Guangming Zhang⁴, Jingjie Guo³, Ming Dong^{2,3}

1College of Sciences, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

2Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Detection and Control, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

3College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

4School of Engineering, Liverpool John Moores University, Byrom Street, L3 3AF, United Kingdom

**Email: chenyan1030@126.com*

Abstract: In digital holographic measurement, various noise sources, including dark current noise, readout noise, and photon shot noise from the CCD, can adversely affect the recording of interference fringes by introducing voltage fluctuations and degrading fringe contrast. Furthermore, optical system imperfections such as stray light, component defects and air turbulence lead to superposition and distortion of the fringe pattern, causing spatial deformation and temporal jitter. These combined effects ultimately compromise the resolution of the reconstructed holographic image. Therefore, in this paper, a super-resolution reconstruction method for the holographically reconstructed images based on improved ESRGAN is proposed, by optimizing the network structure and loss function, and introducing a noise reduction module to enhance the denoising ability and detail recovery performance of this method. Firstly, the DIV2K public dataset is converted into a digital holographically simulated dataset by means of the off-axis Fresnel holographic recording method. Then, the experimental dataset is constructed by reconstructing the holograms of the samples collected using the reflective off-axis optical path system based on Fresnel diffraction method. Finally, the constructed dataset is used for network training and testing, and the experimental results are compared and analyzed with those of the other four methods. The experimental results show that the proposed method outperforms the other methods in five metrics: PSNR, SSIM, RMSE, LPIPS and BRISQUE, especially in detail recovery and texture enhancement, and the reconstruction results are more in line with human visual perception characteristics. Therefore, the improved ESRGAN method proposed in this paper has significant advantages in the super-resolution reconstruction quality of the holographically reconstructed images, and provides a new technical approach for high-precision holographic imaging.

Keywords: Super-Resolution, Digital Holography, ESRGAN

1 Introduction

In digital holographic measurement [1], when a laser is chosen as the light source, its high coherence will lead to random interference of light waves interacting with rough object surfaces or optical components, forming speckle noise [2]. When the beam modulation is not uniform as it passes through the collimating beam expander, beam splitter, and other optical components, the wavefront will produce local delays or advances, leading to phase plane distortion and wavefront distortion. This introduces false phase changes, resulting in the loss of high-frequency details in the reconstructed image and blurring the image [3-4]. In addition, some non-targeted reflected light from the experimental environment enters the optical path and is superimposed on the hologram to generate background noise. Therefore, many super-resolution reconstruction methods have been

developed to overcome the problem of image quality degradation caused by light source coherence and uneven intensity distribution [5-7].

In recent years, deep learning-based super-resolution reconstruction methods have been demonstrated to have significant advantages, which effectively improve the detail recovery ability of holograms by automatically learning image features through a large amount of data training [8-10]. Among them, the convolutional neural network (CNN) based method has gradually become a mainstream one. As a representative of early deep learning methods, the super-resolution convolutional neural network (SRCNN) uses a three-layer convolutional structure to achieve end-to-end super-resolution reconstruction. Dong [11] et al. proposed an end-to-end mapping learning method based on deep networks to directly model the nonlinear relationship between low- and high-resolution images, but this method suffered from the limitations of low computational efficiency and high computational volume. Chen [12] et al. developed a super-resolution method based on the edge fusion diffusion model, which effectively solves the problem of high-frequency information loss and limited generation diversity in single image super-resolution reconstruction by combining the image edge fusion technique with the improved SENet module. Liu et al [13] presented an eight-layer convolutional neural network structure, which significantly improved the characterization of complex data structures by increasing the network depth, and showed better performance in single image super-resolution reconstruction compared to the traditional three-layer network structure. Jin et al [14] applied generative adversarial networks (GANs) to super-resolution image reconstruction, and generated high-resolution images through adversarial training. Compared with traditional methods, this method performs better in detail restoration and texture authenticity, and its reconstruction results are more consistent with the characteristics of human visual perception. Wang et al [15] proposed the enhanced SRGAN (ESRGAN) model by improving the network architecture, adversarial loss and perceptual loss of SRGAN, and using residual-in-residual dense block (RRDB) as the basic network construction unit without batch normalization. Lim et al [16] developed an enhanced deep super-resolution (EDSR) network, which removed the redundant modules in the traditional ResNet, and significantly improved the characterization and reconstruction performance of the model. Chen et al [17] proposed the MMRSRGAN model and two local image super-resolution methods by optimizing and improving the ESRGAN, which effectively enhanced the super-resolution reconstruction accuracy of camera calibrated images. Saharia et al [18] presented an SR3 (super-resolution via repeated refinement) method that combines a denoising diffusion probabilistic model with the U-Net architecture to achieve super-resolution reconstruction through an iterative refinement process. Wang et al [19] proposed an underwater video super-resolution method (UWVSR) based on coarse-grained and fine-grained feature alignment, which effectively improved the super-resolution reconstruction performance of underwater videos. Kim et al [20] developed a single image super-resolution method based on a deep VGG-net-inspired 20-layer convolutional network, which significantly improved the reconstruction accuracy and visualization through residual learning and the very high learning rate training strategy.

Aiming at the problems of high computational complexity and noise sensitivity of the ESRGAN method, an improved ESRGAN method is proposed to enhance the reconstruction performance of holographically reconstructed images in this paper. Firstly, the BM3D algorithm is used to suppress the noise interference in the reconstructed image. Secondly, the depth of the network is optimized to improve the computational efficiency and maintain the reconstruction quality by reducing the number of layers of the RRDB. Finally, the reflective off-axis digital

holographic optical system is constructed to collect the holograms of the test samples for experimental research, and the reconstruction results are compared with those of the commonly used methods to verify the effectiveness of the proposed method.

2 Improved ESRGAN method

2.1 Generator

In this paper, the original generator structure is improved as follows: firstly, the activation function of RRDB is replaced from ReLU to ComplexReLU, so that it has the ability to process the features in the complex domain; secondly, the balance between computational efficiency and reconstruction quality is achieved by reducing the number of RRDB layers. The complex convolution process is shown in Fig. 1, and the improved generator structure is shown in Fig. 2: (1) the underlying features are extracted by the initial convolution layer; (2) the multi-scale features are captured by employing multiple RRDB modules in the depth convolution; (3) the resolution is improved and the amount of computation is reduced by a sub-pixel convolutional up-sampling module; (4) a high-resolution image is output by the convolution layer.

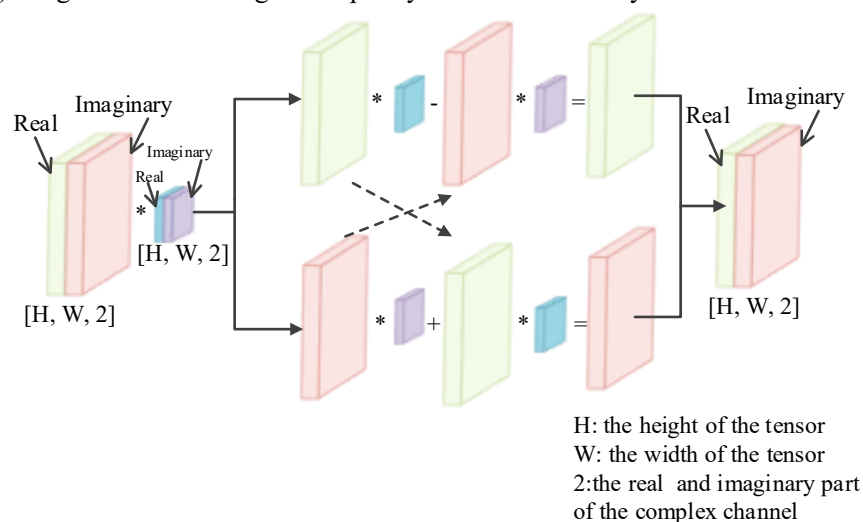
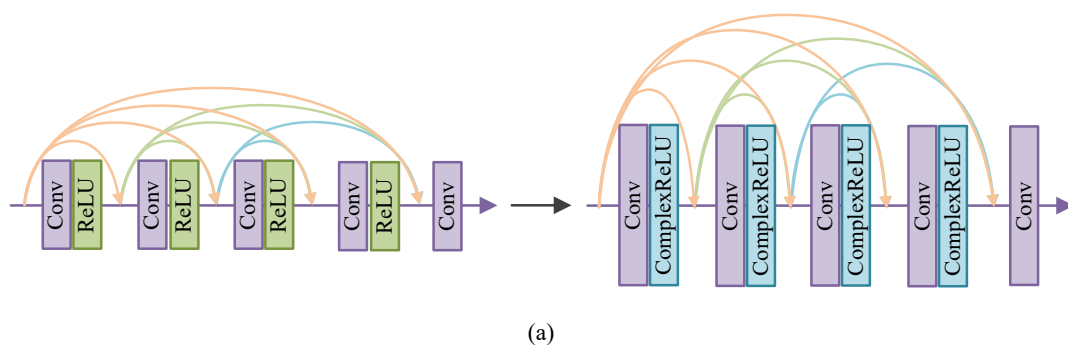


Fig.1 Schematic diagram of the complex convolution process.



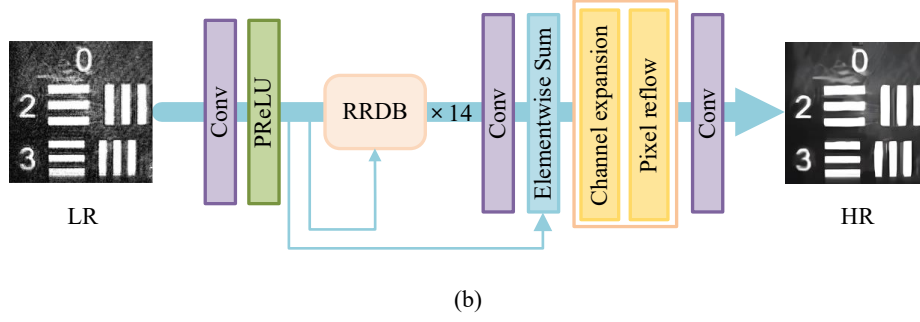


Fig.2 Improved generator structure: (a) RRDB module; (b) generator network structure.

In this paper, ESRGAN is used to design the discriminator, and its network structure is shown in Fig. 3, which is composed of 8 layers of convolution, and each layer consists of 3×3 convolutional layers and LeakyReLU activation function ($\alpha=0.2$) module. The last layer is a single-channel convolution, and the discriminative probability of each patch is output by Sigmoid.

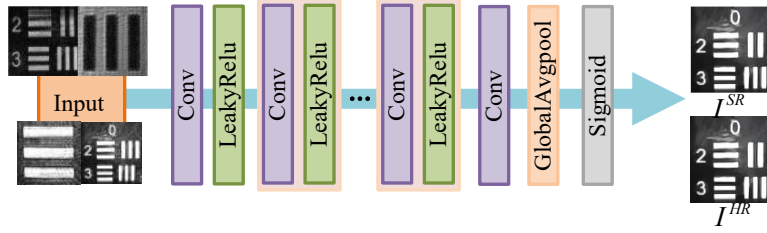


Fig.3 Discriminator network structure.

The GAN optimizes the network model through adversarial training of the generator and the discriminator. Where the discriminator learns to distinguish between real and generated images, and the generator tries to generate more realistic images to fool the discriminator. Through repeated iterations of both, the performance of the generator and the discriminator is gradually improved, and eventually the network is able to generate the high-resolution images.

2.2 Network structure

The BM3D network (NN-BM3D) is constructed, which includes block-matching (BM), 3D collaborative filtering and aggregation module, and uses convolutional network to realize end-to-end training. In this paper, NN-BM3D and ESRGAN are used to construct an improved ESRGAN model, and its network structure is shown in Fig. 4, which adopts NN-BM3D for noise reduction of the input images, and then the improved ESRGAN for super-resolution reconstruction, and finally outputs the high-resolution (HR) images. The network structure is described as follows:

(1) NN-BM3D noise reduction: shallow CNN is used to extract block features, replace the original pixels to calculate the distance (including 2-layer convolution (Conv) and 1-layer activation function (ReLU)), calculate the cosine similarity of the extracted feature maps to generate groups of similar blocks, and stack the similar blocks into 3D groups G_i according to the indices. For each 3D group, the 3D convolution simulation is used to transform the blocks (including 2-layer 3D convolution (Conv3d) and 1-layer activation function (LeakyReLU)), and the threshold filtering is learned using soft thresholding instead of hard thresholding. The filtered 3D groups are inverted into 2D blocks and weighted aggregation is performed: their weights are learned by a small network (consisting of 2-layer Conv2d, 1-layer globally averaged pooling layer, and weight normalization), and finally the block weights are added to the image.

(2) Super-resolution reconstruction: the noise-reduced image is input to the network, and the first convolution layer is used to extract the edge information of the image (including a 3×3

convolution layer with a step size of 1). The second to 16th convolution layers are used to extract the deep features by the improved RRDB layer. The 17th layer performs upsampling using sub-pixel convolution, and reconstructs the low-resolution (LR) image as a HR image. First, the features of the LR image are extracted and the number of channels is expanded through convolution. Then, the traditional upsampling information loss is avoided through pixel rearrangement and the image quality is processed and optimized. Finally, the two-layer convolution is used to convert the feature maps into the RGB image and output the HR image.

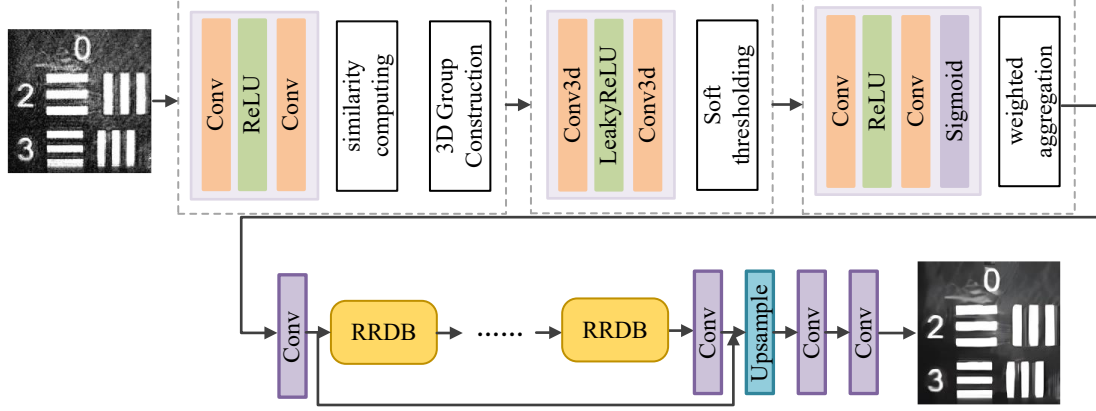


Fig.4 Improving the network structure of ESRGAN.

2.3 Loss functions

The loss function is taken from the ESRGAN, including adversarial loss, perceptual loss and pixel-wise L1 loss functions.

(1) The adversarial loss function is the core of GAN, which is expressed as:

$$L_G^{\text{adv}} = -E_{x \sim p_{\text{data}}(x)}[\log(1 - D_{\text{ra}}(x, G(z)))] - E_{z \sim p_z(z)}[\log D_{\text{ra}}(G(z), x)] \quad (1)$$

$$L_D^{\text{adv}} = -E_{x \sim p_{\text{data}}(x)}[\log D_{\text{ra}}(x, G(z))] - E_{z \sim p_z(z)}[\log(1 - D_{\text{ra}}(G(z), x))] \quad (2)$$

where $D_{\text{ra}}(x, G(z))$ is the relative authenticity score of the discriminator between the real high-resolution image x and the generated high-resolution image $G(z)$ and $D_{\text{ra}}(G(z), x)$ is the relative authenticity score of the discriminator between the generated image $G(z)$ and the real image x .

(2) The perceptual loss function extracts the high-level features of the image through the pre-trained network and is used to measure the perceptual similarity between the generated image and the real image, which is expressed as:

$$L_{\text{perceptual}} = E_{x \sim p_{\text{data}}(x), z \sim p_z(z)} \left[\sum_i \frac{1}{C_i H_i W_i} \|\phi_i(x) - \phi_i(G(z))\| \right] \quad (3)$$

where $\phi_i(x)$ is the feature map of the real image x at layer i of the pre-trained network VGG19; $\phi_i(G(z))$ is the feature map of the generated image $G(z)$ at layer i of the pre-trained network; C_i, H_i and W_i are the number of channels, height and width of the feature map, respectively.

(3) The L1 regularization loss function is the sum of the absolute values of the pixel-level differences between the generated image and the real image, and is used to constrain the pixel-level differences between the generated image and the real HR image, which is calculated by:

$$L_{L1} = E_{x \sim p_{\text{data}}(x), z \sim p_z(z)} [\|x - G(z)\|_1] \quad (4)$$

The total loss function of the improved ESRGAN is:

$$L_G = \lambda_{\text{adv}} L_G^{\text{adv}} + \lambda_{\text{perceptual}} L_{\text{perceptual}} + \lambda_{L1} L_{L1} \quad (5)$$

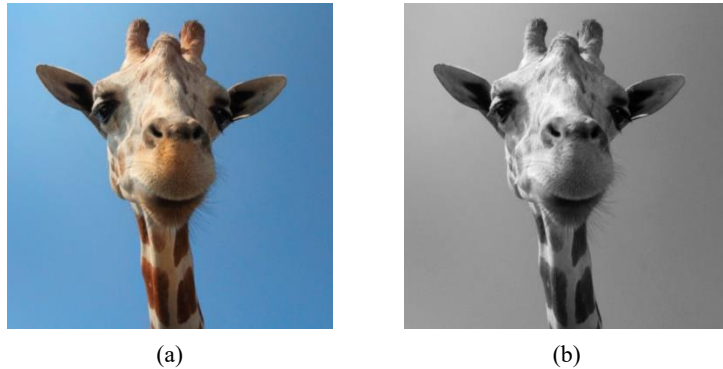
where λ_{adv} , $\lambda_{\text{perceptual}}$, and λ_{L1} are the weights for adversarial loss, perceived loss and L1 loss, respectively. The loss function of the discriminator is solely the adversarial loss, which is used to accurately distinguish between the real and generated images, and to maximize the probability of correctly identifying the real images. In contrast, the loss function of the generator integrates three different loss terms to maximize the probability of the discriminator misjudging the generated images. To ensure the stability of the network structure, the weights are adjusted in this paper in segments to determine the optimal values. Initially, λ_{adv} , $\lambda_{\text{perceptual}}$ and λ_{L1} are set to 0.1, 0.5 and 1.0, respectively. After dynamic parameterization, the final weights are determined to be $\lambda_{\text{adv}} = 1.0$, $\lambda_{\text{perceptual}} = 1.0$ and $\lambda_{L1} = 0.01$, respectively.

3 Experimental verifications

All the experiments in this paper are completed on a computer with a 13th Gen Intel(R) Core (TM) i9-13900KF CPU, 32G RAM memory and an NVIDIA GeForce RTX 4090 GPU. The Python 3.7 programming language and the PyTorch 2.1 framework are used to train and test the network models for the improved ESRGAN, ESRGAN, SRCNN, VDSR (very deep super-resolution network) and improved Real-ESRGAN(IR-ESRGAN) [21], within the PyCharm integrated development environment. The angular spectrum method (ASM) is performed within this environment. The simulation dataset preparation is implemented using MATLAB 2015b.

3.1 Dataset preparation

In this paper, the simulated images and real images are used to construct the dataset. A total of 1000 LR images and 1000 HR images are selected from the public data set DIV2K [22], which are first converted to grayscale images. Then, the optical propagation process is simulated based on the principle of Fresnel diffraction and the off-axis holographic recording method is used to generate the digital holograms. Finally, the holographically reconstructed images are obtained using Fresnel diffraction reconstruction. The simulated results are shown in Fig. 5.



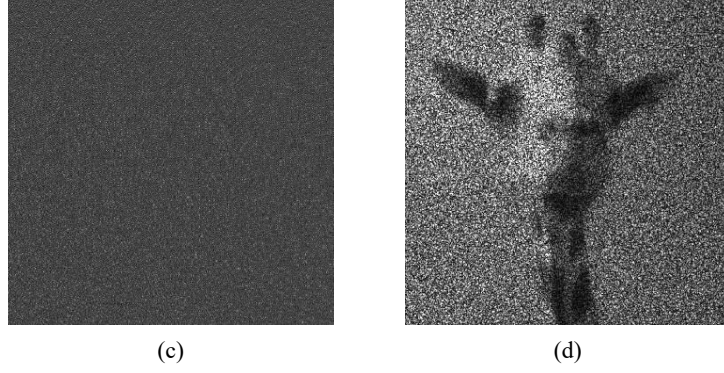


Fig.5 The simulated results: (a) original image; (b) grayscale image; (c) hologram; (d) reconstructed image.

To obtain the real holograms, the constructed reflective off-axis digital holographic experimental setup (as shown in Fig.6) is used to collect 50 digital holograms of test samples (as shown in Fig.7), which are reconstructed into the real images based on Fresnel diffraction, with a reconstruction distance of 800mm, and they are randomly rotated in the range of $-90^\circ \sim 90^\circ$ to generate a total of 500 holographically reconstructed images.

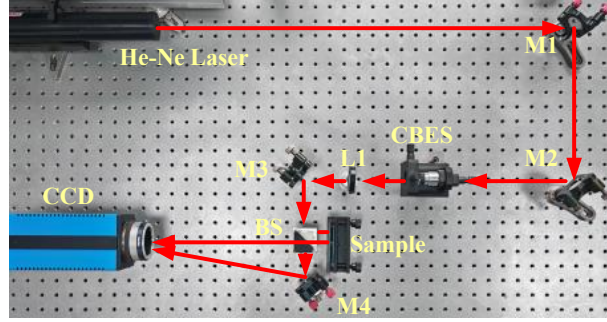


Fig.6 Experimental setup.

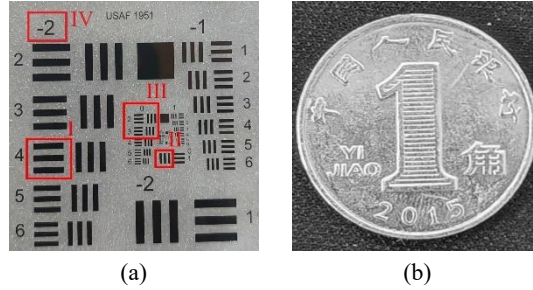


Fig.7 Test samples: (a) optical resolution plate; (b) coin.

3.2 Image quality evaluation metrics

In this paper, peak signal-to-noise ratio (PSNR) [23], structural similarity index (SSIM) [24], root mean squared error (RMSE) [25], learned perceptual image patch similarity (LPIPS) [26], and blind/referenceless image spatial quality evaluator (BRISQUE) [27] are used to quantitatively evaluate the image quality after super-resolution reconstruction.

(1) PSNR is used to measure the pixel-level distortion of an image, which is calculated as follows:

$$MSE = \frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} [I_{i,j} - K_{i,j}]^2 \quad (6)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (7)$$

where MSE is the mean square error of the real image I and the generated image K ; $I(i, j)$ and $K(i, j)$ are the pixel values at the respective coordinates; $H \times W$ is the image resolution; MAX is the maximum possible values of the image.

(2) SSIM is used to measure the image similarity in terms of structure, brightness and contrast, which is more in line with human vision perception, and is calculated as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

where μ is the mean value of the image, σ is the variance of the image, σ_{xy} is the covariance of the images x and y , and C_1 and C_2 are constants.

(3) RMSE is used to measure the pixel-level accuracy in the image reconstruction task, and the smaller the RMSE value, the better the image quality. It is calculated using the following formula:

$$RMSE = \sqrt{\frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} [I(i, j) - K(i, j)]^2} \quad (9)$$

(4) LPIPS has a high computational complexity and relies on pre-trained models, but it is more in line with human visual perception. The smaller the value of LPIPS, the better the image quality, which is calculated as:

$$LPIPS(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\phi_l(x)_{h,w} - \phi_l(y)_{h,w})\|_2^2 \quad (10)$$

where $\phi_l(x)$ and $\phi_l(y)$ are the feature maps for images x and y in the l -layer of the pre-trained network, respectively; w_l is the weight of the l -layer; H_l and W_l are the height and width of the feature maps, respectively.

(5) BRISQUE is combined with CNN to evaluate the image quality, and the smaller the BRISQUE value, the better the reconstruction. The normalization coefficient (MSCN) is calculated as follows:

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L w(k, l) \cdot I(i+k, j+l) \quad (11)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w(k, l) \cdot [I(i+k, j+l) - \mu(i, j)]^2} \quad (12)$$

$$MSCN(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (13)$$

where w is a Gaussian weighted window, I is the pixel value of the image, and C is a minimal constant ensuring the denominator is non-zero.

3.3 Network training and testing

The 1000 sets of LR-HR simulated images are cropped and rotated to construct a dataset containing 4000 sets of LR-HR holographically reconstructed images, which is divided into training and testing sets at a ratio of 9:1 to train and test the improved ESRGAN model. The model training

parameters are set as follows: the number of input samples per batch is 16; the initial learning rate of the Adam optimizer is 0.005 and is updated iteratively with a multiplicity of 0.50; the iteration is stopped when the loss reaches $1e-5$; the convolution compensation is set to 1; the depth convolution adopts a 3×3 convolution kernel, and the point-by-point convolution adopts a 1×1 convolution kernel. The network has converged after 125 training iterations, and the training loss function curve is shown in Fig. 8.

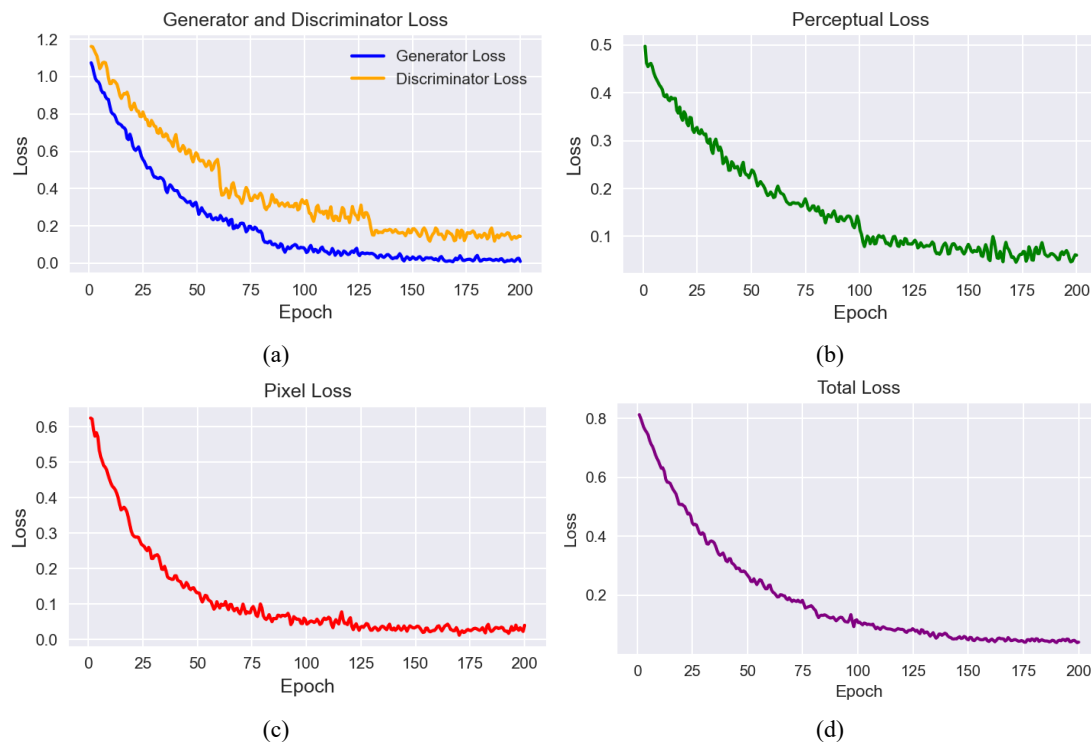


Fig.8 Network training loss curves: (a) generative adversarial loss; (b) perceptual loss; (c) pixel loss (L1 loss); (d) total loss.

As can be seen from Fig. 8, the model constructed in this paper exhibits good convergence characteristics. To evaluate the network performance, 400 samples are selected from the original 500 experimental data and mixed with the test set samples to test the network, and the test results are then quantitatively analyzed using the PSNR and SSIM metrics, the calculation results of which are shown in Fig. 9.

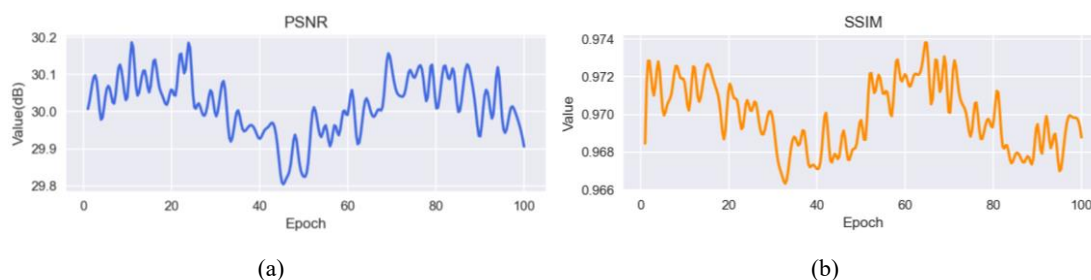


Fig.9 Two metrics for the test results of the improved ESRGAN model: (a) PSNR; (b) SSIM.

The mean values of PSNR and SSIM are 30.024 and 0.970, respectively. As can be seen from Fig. 9, both metrics exhibit smooth amplitude fluctuations and high mean values, indicating that the reconstructed images generated by the improved ESRGAN model are less distorted and more consistent with the perceptual characteristics of human vision in terms of texture detail, edge sharpness and structural integrity.

3.4 Simulation results and analysis

The effectiveness of the improved ESRGAN model is verified by comparing its results with those of the ESRGAN, ASM, SRCNN, VDSR and IR-ESRGAN methods using the simulation dataset.

In the simulation test, the camera is first used to shoot the optical resolution plate and complete the preprocessing operation. Then, based on the Fresnel diffraction principle, the propagation of light is simulated and the digital hologram is generated using the off-axis holographic recording method. Finally, the holographically reconstructed image is obtained using the Fresnel diffraction reconstruction technique. The data simulation results are presented in Figs. 10(a) and (b), respectively, while the reconstruction results are shown in Figs. 10 (c)-(h).

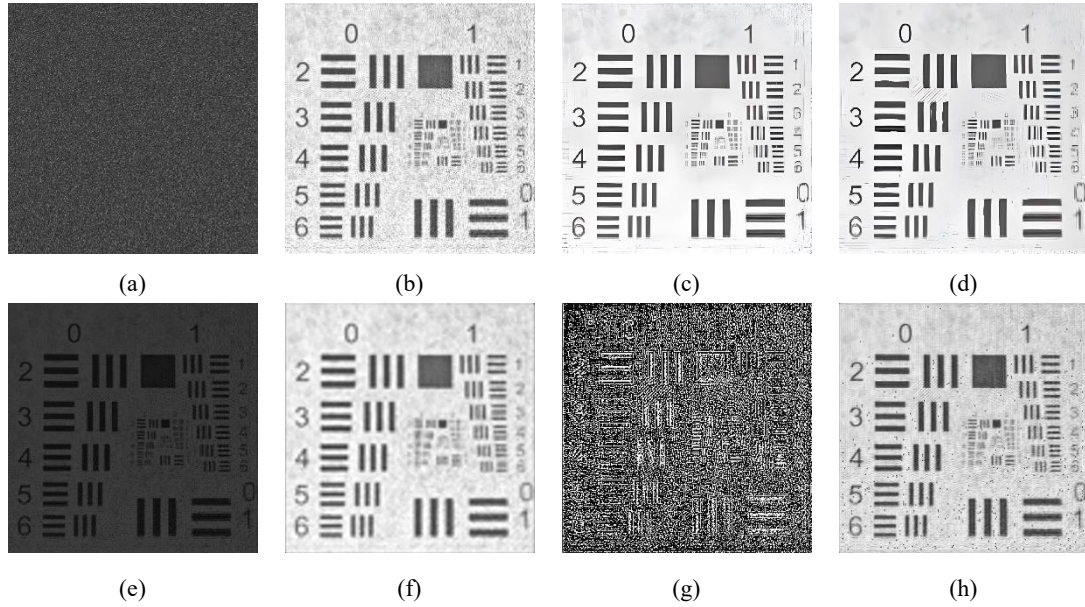


Fig.10 Test results: (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 1 Calculation results of five metrics for Figs.10(c)-10(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	19.324	18.678	4.068	19.005	4.955	17.574
SSIM	0.6158	0.570	0.201	0.592	0.010	0.599
RMSE	9.293	9.467	10.344	9.515	10.402	9.365
LPIPS	0.407	0.404	0.344	0.553	1.017	0.556
BRISQUE	81.771	85.241	85.547	82.883	164.153	115.56
Time(s)	3.491	4.253	13.497	9.588	64.146	25.346
Resolution	1152×1152	1152×1152	576×576	288×288	286×286	288×288

As can be seen from Fig. 10 and Table 1, the method proposed in this paper yields the best results when the added noise is consistent.

3.5 Experimental results and analysis

The improved ESRGAN model is tested using the experimental dataset, and the results are compared with those of the ESRGAN, ASM, SRCNN, VDSR and IR-ESRGAN methods to verify its effectiveness.

3.5.1 Analysis of the reconstruction results for optical resolution plate

The experimental results of the holographically reconstructed image for Region I in Fig. 7 (a) are shown in Fig. 11. It can be seen that compared to other methods, the improved ESRGAN method significantly improves the clarity of the image, makes the detailed information clearer and eliminates the noise points in the reconstructed image, enhancing the sense of hierarchy. To quantitatively analyze the reconstruction effect of these five methods, the above five indicators are used to evaluate the image quality after super-resolution reconstruction, and the calculation results are shown in Table 2.

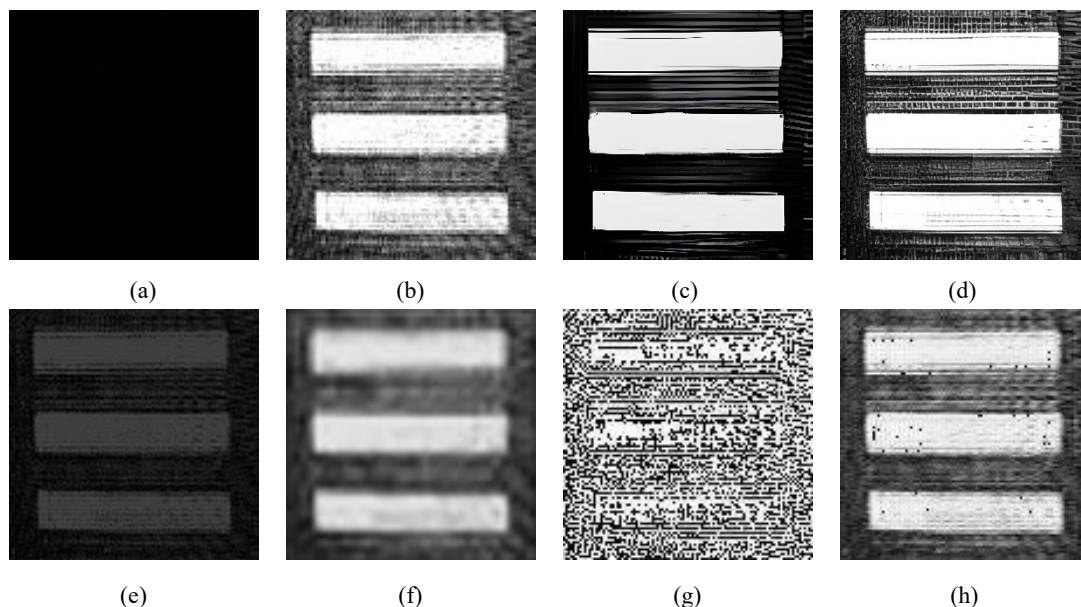
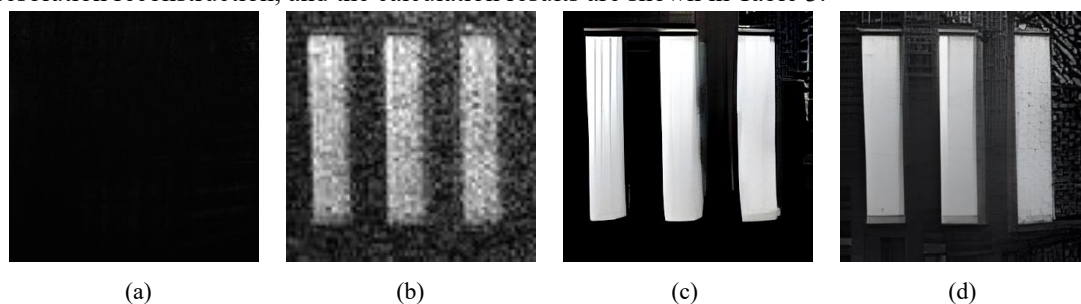


Fig.11 Experimental results of Region I: (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 2 Calculation results of five metrics for Figs.11(c)-10(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	21.757	18.724	6.524	17.146	5.108	17.216
SSIM	0.832	0.589	0.204	0.485	0.017	0.680
RMSE	8.433	9.018	10.248	10.094	9.894	9.182
LPIPS	0.267	0.416	0.315	0.576	0.867	0.433
BRISQUE	26.258	51.543	79.037	39.628	169.793	109.54
Time(s)	3.491	4.253	13.497	9.588	64.146	26.764
Resolution	428×428	428×428	214×214	105×105	107×107	107×107

The experimental results of the holographically reconstructed image for Region II in Fig. 7 (a) are shown in Fig. 12. The above five indicators are used to evaluate the image quality after super-resolution reconstruction, and the calculation results are shown in Table 3.



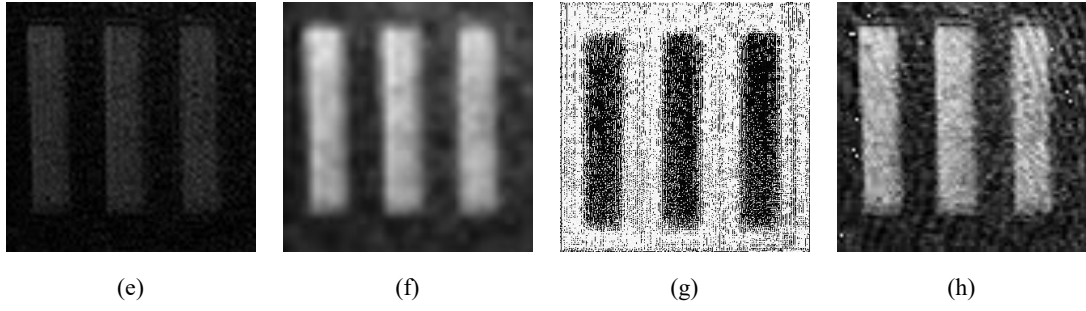


Fig.12 Experimental results of Region II: (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 3 Calculation results of five metrics for Figs.12(c)-11(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	20.384	18.845	9.743	19.797	4.376	17.722
SSIM	0.662	0.467	0.200	0.526	0.026	0.390
RMSE	9.648	9.708	10.367	9.734	10.388	9.767
LPIPS	0.191	0.272	0.261	0.585	0.295	0.131
BRISQUE	29.681	89.671	72.671	35.464	126.671	104.934
Time(s)	3.481	4.253	13.497	9.588	64.146	23.560
Resolution	192×192	192×192	164×164	81×81	152×152	82×82

The experimental results of the holographically reconstructed image for Region III in Fig. 7 (a) are shown in Fig. 13. The above five indicators are used to evaluate the image quality after super-resolution reconstruction, and the calculation results are shown in Table 4.

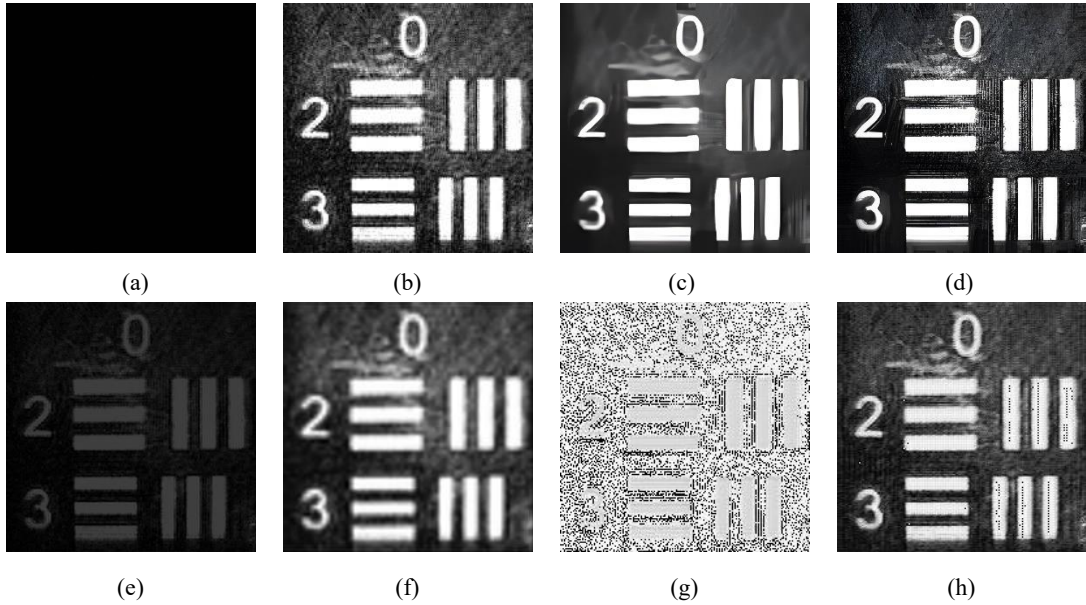


Fig.13 Experimental results of Region III (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 4 Calculation results of five metrics for Figs.13(c)-12(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	22.284	20.261	8.413	21.027	4.891	21.341
SSIM	0.838	0.722	0.226	0.631	0.048	0.749
RMSE	8.752	8.949	10.287	8.991	10.573	8.848
LPIPS	0.133	0.150	0.279	0.359	0.917	0.296

BRISQUE	13.094	76.061	72.591	38.884	141.592	93.421
Time(s)	3.490	4.253	13.497	9.588	64.146	23.753
Resolution	768×768	768×768	384×384	192×192	192×192	192×192

The experimental results for the holographically reconstructed image of Region IV in Fig. 7 (a) are shown in Fig. 14. The above five indicators are used to evaluate the image quality after super-resolution reconstruction, and the calculation results are shown in Table 5.

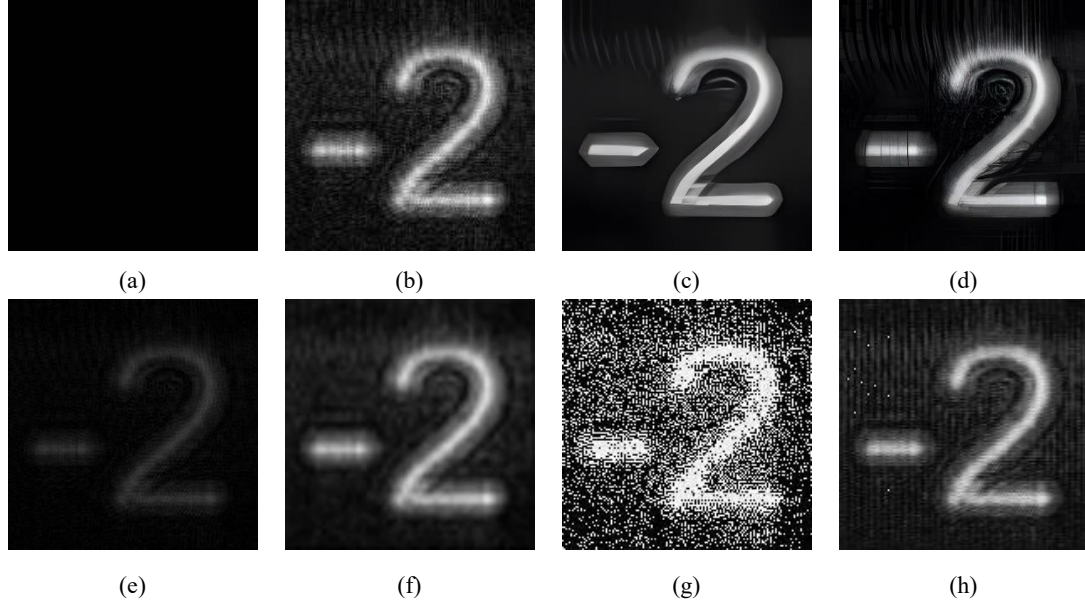


Fig.14 Experimental results of Region IV: (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 5 Calculation results of five metrics for Figs.14(c)-12(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	23.954	21.103	14.113	22.837	7.081	22.509
SSIM	0.568	0.546	0.244	0.536	0.021	0.482
RMSE	8.613	9.438	10.196	8.044	10.108	8.968
LPIPS	0.225	0.351	0.375	0.497	0.8895	0.300
BRISQUE	13.094	75.660	79.893	39.724	154.072	102.652
Time(s)	4.470	5.453	12.387	10.578	63.899	25.862
Resolution	596×596	596×596	298×298	147×147	149×149	149×149

It can be observed from Tables 2-5 that compared to the other four methods, the values of PSNR and SSIM corresponding to the super-resolution reconstruction image obtained using the proposed method are the largest, while the value of LPIPS is the smallest, indicating that the reconstructed image is less distorted and more in line with the perceptual characteristics of human vision in terms of texture details, edge sharpness and structural integrity. In addition, the value of RMSE is the smallest, demonstrating that the reconstructed image is of higher quality and closer to the real image. The value of BRISQUE is also the smallest, demonstrating that the super-resolution reconstruction is more effective. Furthermore, the computational time of the proposed method is the shortest. Therefore, the proposed method outperforms the other four methods in terms of super-resolution reconstruction of holographically reconstructed images.

3.5.2 Analysis of the reconstruction results for coin

The experimental results of the holographically reconstructed image for the coin are shown in

Fig. 15. The above five indicators are used to evaluate the image quality after super-resolution and the calculation results are shown in Table 6.

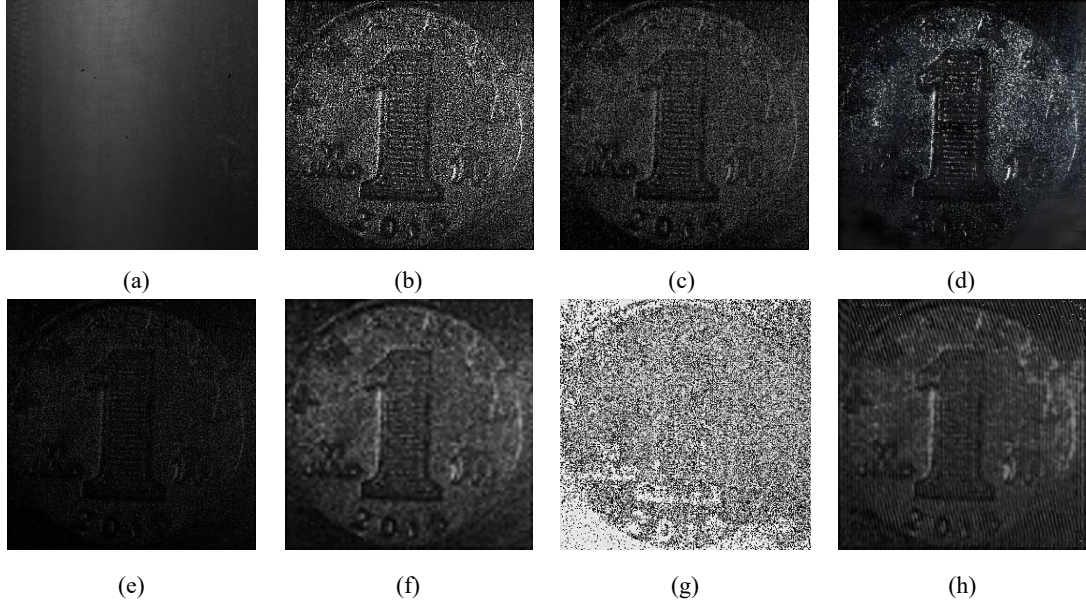


Fig.15 Experimental results for coin: (a) hologram; (b) reconstructed image; (c) improved ESRGAN; (d) ESRGAN; (e) ASM; (f) SRCNN; (g) VDSR; (h) IR-ESRGAN.

Table 6 Calculation results of five indexes for Figs.15(c)-11(h).

	Improved ESRGAN	ESRGAN	ASM	SRCNN	VDSR	IR-ESRGAN
PSNR	15.332	12.671	12.006	13.446	5.915	13.087
SSIM	0.574	0.211	0.214	0.149	0.009	0.219
RMSE	9.850	10.012	9.592	9.997	10.397	10.248
LPIPS	0.168	0.351	0.353	0.626	0.746	1.180
BRISQUE	43.482	79.953	91.284	46.907	159.940	94.925
Time(s)	6.553	7.047	13.808	31.215	63.035	26.206
Resolution	2218×2218	2218×2218	532×532	264×264	264×264	264×264

It can be seen from Table 6 that the values of PSNR and SSIM the super-resolution reconstruction image obtained using the proposed method are the largest compared to the other four methods. Meanwhile, the values of RMSE, LPIPS, and BRISQUE are the smallest, and the computational time is the shortest. This further proves that the method has significant advantages in terms of reconstruction accuracy, visual quality and computational efficiency.

4 Conclusions

In this paper, a super-resolution reconstruction method for holographically reconstructed images based on improved ESRGAN is proposed, which involves optimizing the network structure and loss function, as well as introducing the noise reduction model to enhance its denoising ability and detail recovery performance. Firstly, the noise reduction model is used to denoise the holographically reconstructed image. Then, the image is input into the generator for super-resolution reconstruction. Finally, the proposed method, ESRGAN, ASM, SRCNN, VDSR and IR-ESRGAN are used to perform super-resolution reconstruction of holographically reconstructed images, and five metrics are used to quantitatively evaluate the reconstruction results. Simulation and experimental results show that the PSNR value of the reconstructed image obtained using the proposed method is by 7.71% higher than that of ESRGAN, by 2.97% higher than that of SRCNN

and by 4.23% higher than of IR-ESRGAN. The SSIM value is by 3.87% higher than ESRGAN and by 25.85% higher than SRCNN. The RMSE value is at least 0.62% lower than ESRGAN and at least 0.89% lower than SRCNN. The LPIPS value is at least 11.33% lower than ESRGAN and at least 31.41% lower than the other four methods. The BRISQUE value is at least 45.62% lower than ESRGAN and at least 7.31% lower than SRCNN. The above results demonstrate that the reconstructed images obtained using the proposed method are less distorted and more consistent with the perceptual characteristics of human vision in terms of texture detail, edge sharpness and structural integrity, and has higher computational efficiency. Therefore, the improved ESRGAN method offers significant advantages for the super-resolution reconstruction of holographically reconstructed images.

Funding. National Natural Science Foundation of China (Grant No. 52175518).

Disclosures. The authors declare no conflicts of interest.

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

References

- [1] Zeng T, Zhu Y, Lam E Y. Deep learning for digital holography: a review[J]. *Optics Express*, 2021, 29(24): 40572-40593.
- [2] Chen Y, Fan Y, Zhang G, et al. Speckle noise suppression of a reconstructed image in digital holography based on the BM3D improved convolutional neural network[J]. *Applied Optics*, 2024, 63(22): 6000-6011.
- [3] Chen Z Q, Fu S. Optical wavefront distortion due to supersonic flow fields[J]. *Chinese Science Bulletin*, 2009, 54: 623-627.
- [4] Tsang P W M, Chow Y T, Poon T C. Generation of phase-only Fresnel hologram based on down-sampling[J]. *Optics Express*, 2014, 22(21): 25208-25214.
- [5] Chen B, Zhang J, Huang L, et al. High-resolution reconstruction of spectrum-overlapped off-axis holography by deflecting reference beam of Gaussian symmetry[J]. *Applied Physics Letters*, 2023, 123(11): 111109.
- [6] Zhang W, Ren Y, Feng Y, et al. Deep learning bimodal frequency-selective digital holography by using a confocal dual-beam setup[J]. *New Journal of Physics*, 2024, 26(12): 123021.
- [7] Chang X, Jiang S, Hu Y, et al. Pixel super-resolved lensless on-chip sensor with scattering multiplexing[J]. *ACS Photonics*, 2023, 10(7): 2323-2331.
- [8] Park S, Kim Y, Moon I. Automated phase reconstruction and super-resolution with deep learning in digital holography[J]. *Optics & Laser Technology*, 2024, 176: 111030.
- [9] Chen H, Huang L, Liu T, et al. EFIN: enhanced Fourier imager network for generalizable autofocusing and pixel super-resolution in holographic imaging[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2023, 29(4: Biophotonics): 1-10.
- [10] Shaked N T, Micó V, Trusiak M, et al. Off-axis digital holographic multiplexing for rapid wavefront acquisition and processing[J]. *Advances in Optics and Photonics*, 2020, 12(3): 556-611.
- [11] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(2): 295-307.
- [12] Chen Z, Zhang L, Zhang X. Edge fusion diffusion for single image super-resolution[J]. *Knowledge-Based Systems*, 2025: 113241.
- [13] Liu P F, ZHAO H C, LIU M D. Image super-resolution reconstruction based on convolutional neural network[J]. *Computer Engineering and Applications*, 2019, 55(09): 197-202.

- [14] Jin Y, Zhang Y, Cen Y, et al. Pedestrian detection with super-resolution reconstruction for low-quality image[J]. Pattern Recognition, 2021, 115: 107846.
- [15] Wang X, Yu K, Wu S, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]. 15th European Conference on Computer Vision (ECCV). Munich, Germany. 2018: 63-79.
- [16] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. IEEE conference on computer vision and pattern recognition workshops. Honolulu, USA. 2017: 136-144.
- [17] Chen H, Li H, Yao C, et al. Image super-resolution based on improved ESRGAN and its application in camera calibration[J]. Measurement, 2025, 242: 115899.
- [18] Saharia C, Ho J, Chan W, et al. Image super-resolution via iterative refinement[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 4713-4726.
- [19] Wang J, Lu H. Coarse-to-fine grained alignment video super-resolution for underwater camera[J]. IEEE Transactions on Consumer Electronics, 2024.
- [20] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, Nevada, USA. 2016: 1646-1654.
- [21] Chen G, Chen J, Liao H, et al. Super-resolution reconstruction of off-axis digital holography based on Real-ESRGAN[J]. Applied Optics, 2025, 64(9): 2245-2254.
- [22] Li H, Yang Y, Chang M, et al. SRDiff: single image super-resolution with diffusion probabilistic models[J]. Neurocomputing, 2022, 479: 47-59.
- [23] Sun J, Yuan Q, Shen H, et al. A single-frame and multi-frame cascaded image super-resolution method[J]. Sensors, 2024, 24(17): 5566.
- [24] Huan H, Zou N, Zhang Y, et al. Remote sensing image reconstruction using an asymmetric multi-scale super-resolution network[J]. The Journal of Supercomputing, 2022, 78(17): 18524-18550.
- [25] Chen Y, Wang Q, Zhang G M, et al. Phase unwrapping in digital holography based on SRDU-Net[J]. Optics Communications, 2024, 573: 131055.
- [26] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, Utah, USA. 2018: 586-595.
- [27] Zhu H, Li L, Wu J, et al. Generalizable no-reference image quality assessment via deep meta-learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1048-1060.