A DARU-Net-based method for digital holographic phase denoising

Yuan Chen[a,b,*], Zheng Wen[c], Peichao Li[c], Guangming Zhang[d], Jingjie Guo[c], Ming Dong[b,c], Hongwei Ma[b,c]

[a]*College of Sciences, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China*

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

[c]*College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi,710054, China*

[d]*School of Engineering, Liverpool John Moores University, Liverpool, L3 3AF, United Kingdom*

**Email: chenyuan1030@126.com*

Abstract: In digital holographic measurement, noise is inevitably introduced during the acquisition process, which severely degrades phase quality and hinders accurate recovery of the true phase and three-dimensional morphology of the specimen. Thus, a phase denoising method for digital holography based on dilated attention residual U-Net (DARU-Net) is proposed in this paper. This method adopts U-Net as the backbone and incorporates multiple dilated convolutional blocks (DCBs) with varying dilation rates to enlarge the receptive field, thereby enabling multi-scale feature extraction. A hybrid attention block, combining channel and spatial attention, is integrated into the skip connections to highlight key phase information via adaptive learning of feature weights. A simulated dataset containing various fringe densities and noise levels is constructed for model training and evaluation. In addition, a transmission digital holographic optical path imaging system is constructed to collect holograms of the specimens for experimental validation. Both simulation and experimental results demonstrate that the proposed method outperforms other comparison methods in terms of noise reduction performance across all evaluation metrics, and significantly improves phase unwrapping accuracy. Overall, the proposed method effectively suppresses noise interference and exhibits notable advantages in noise robustness and phase fidelity.

Keywords: digital holography; DARU-Net; phase denoising; hybrid attention block; dilated convolutional block

1 Introduction

Digital holography enables simultaneous recording of both the intensity and phase information of the measured object, and has been widely applied in medical imaging, profilometry, and defect detection [1-4]. In practice, however, when coherent light illuminates a rough object surface, the superposition of scattered wavefronts from individual surface points gives rise to granular speckle and parasitic fringes, which constitute the dominant sources of coherent noise in digital holographic systems [5]. In addition, mechanical vibrations and environmental disturbances can introduce phase perturbations, ultimately degrading the signal-to-noise ratio of the phase data and compromising accurate recovery of the true phase and three-dimensional morphology of the specimen. Therefore, phase denoising is essential for precise phase retrieval.

Conventional methods for speckle suppression in digital holography are primarily divided into two categories: optical and numerical methods [6]. However, they have inherent limitations in practical applications. Optical methods suffer from high system complexity, poor real-time performance, limited noise suppression capability, and difficulty in meeting the requirements of dynamic phase measurement [7,8]. Numerical methods, on the other hand, are highly dependent on parameter selection, entail high computational cost, lack generalizability, and are prone to losing key phase information [9-11]. Recently, deep learning methods have attracted widespread attention in the field of digital holographic phase denoising, owing to the powerful feature extraction and adaptive learning capabilities of convolutional neural networks. Hao et al. [12] applied the FFDNet

model to fringe pattern denoising in electronic speckle pattern interferometry, achieving favorable denoising speed, although the denoising accuracy still leaves room for improvement. Chen et al. [13] proposed a Swin-UNet-based denoising network that employs the Transformer architecture to smooth noisy images while preserving both shallow and deep features, thereby yielding good denoising performance. Fang et al. [14] presented a convolutional Fourier neural network architecture that combines the advantages of convolutional neural networks (CNNs) and Fourier neural operators to extract features simultaneously in both spatial and frequency domains, enabling effective denoising under severe speckle noise and reducing phase unwrapping error. Yan et al. [15] proposed a CNN-based wrapped phase denoising method that separately denoises the numerator and denominator of the arctangent function, applies the arctangent operation to the filtered results to recover a clean wrapped phase, and subsequently extracts accurate phase using a straightforward phase unwrapping algorithm. Fang et al. [16] proposed a conditional generative adversarial network combining U-Net and DenseNet for wrapped-phase denoising. Montresor et al. [17] used a residual network to estimate speckle noise in phase data and subtracted the estimate from the noisy input to obtain the denoised phase. In related fringe-pattern processing, Gurrola-Ramos et al. [18] proposed a U-Net-based fringe-pattern denoising network that combines dense connections and residual learning to recover fringe structures. Ram et al. [19] combined deep learning with a signal subspace method to estimate phase derivatives from noisy interferograms, and Ram et al. [20] used a conditional generative adversarial network for fringe pattern normalization. In summary, deep learning-based denoising methods offer significant advantages over traditional methods in term of computational efficiency, robustness, and adaptive feature learning, showing great promise for practical applications.

Nevertheless, most existing deep learning-based phase denoising methods rely heavily on large, high-quality datasets, and insufficient training data often degrades model generalizability. Moreover, most models prioritize noise suppression over the preservation of key phase features, such as edges and fine fringe structures. Therefore, this paper focuses on digital holographic phase map denoising and develops DARU-Net based on the U-Net encoder-decoder architecture. Compared with the U-Net, which uses conventional convolutions for feature extraction and directly fuses the corresponding encoder and decoder features through skip connections, DARU-Net introduces a dilated convolutional block with parallel branches of different dilation rates in the encoder to extract phase features at multiple scales. Meanwhile, channel and spatial attention blocks are embedded in the skip connections to adaptively weight the features transferred to the decoder. The network takes a noisy phase map as input, predicts the noise residual, and subtracts it from the input phase map to obtain the denoised phase map. A simulated dataset containing various fringe densities and noise levels is constructed using random matrix enlargement (RME) and Gaussian function superposition (GFS) for network training and evaluation. In addition, a transmission digital holographic optical path imaging system is established to collect holograms of the specimens, thereby verifying the effectiveness of the proposed method. Furthermore, ablation experiments are conducted under simulated data splits and training configurations to evaluate the respective contributions of residual learning, the dilated convolutional block, and the hybrid attention block (HAB) to phase denoising performance, thereby verifying the effectiveness of the individual components and their combined design in DARU-Net.

2 Phase dataset preparation

When training deep learning models for phase denoising, it is essential to construct a dataset containing a sufficiently large number of noise phase samples. To enable the network to handle complex phase variations, the training data must cover a diverse range of phase distributions and fringe jump patterns. However, owing to experimental constraints, it is difficult to obtain a large number of real-world phase maps as ground-truth labels. Therefore, two numerical simulation methods, namely RME and GFS are adopted to generate the training dataset in this paper [21].

The RME method generates diverse phase distributions by upsampling a randomly initialized matrix via linear interpolation. The specific parameters are set are as follows: the initial matrix size is drawn from the range [2,8]; phase values are distributed over the intervals [10,30], [30,40] and [40,50] in a ratio of 4:3:3; the initial matrix is subsequently upsampled to 128×128 by linear interpolation to produce the absolute phase distribution [22,23]. The GFS method simulates the phase surface of the object wave by superposing randomly generated Gaussian functions [24,25]. The parameter settings are as follows: the number of Gaussian components ranges over [5,9], the standard deviation over [10,20], and the amplitude over [0,1]. After superposition, the resulting phase map is linearly rescaled to an absolute phase height within [10,50], thereby yielding the final absolute phase distribution.

To simulate the noise conditions encountered in practice, speckle noise, Gaussian noise, and their mixture are independently added to each simulated absolute phase map. The speckle intensity factor ranges over [0,0.5] and the Gaussian noise standard deviation over [0,1], further enhancing dataset diversity and ensuring that the trained model has good generalization ability. The final dataset comprises 20000 sample pairs, each consisting of a noisy phase map as the input and the corresponding noise residual as the target label. Two representative examples of noisy and noise-free phase maps from the dataset are illustrated in Fig. 1.

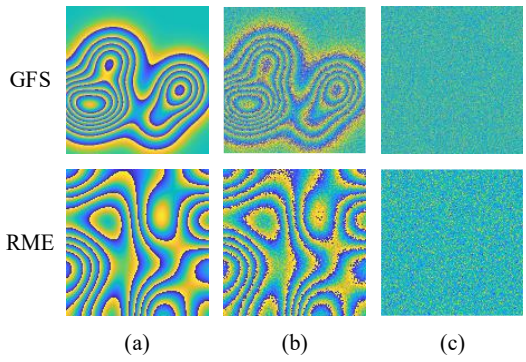


Fig. 1. Two representative examples of phase data: (a) noise-free phase; (b) noisy phase; (c) noise residual.

3 DARU-Net network model

3.1 Phase denoising workflow

This paper proposes a phase denoising method based on DARU-Net, adopting U-Net as the backbone architecture. Since the wrapped phase is confined to the interval (-π, π), directly recovering the clean phase from the noisy input is generally challenging. Therefore, a residual learning strategy is adopted: the noisy phase is fed into the network as input, the network is trained to predict the noise residual, and the denoised phase is obtained by subtracting the predicted residual from the input. The overall workflow is illustrated in Fig. 2.

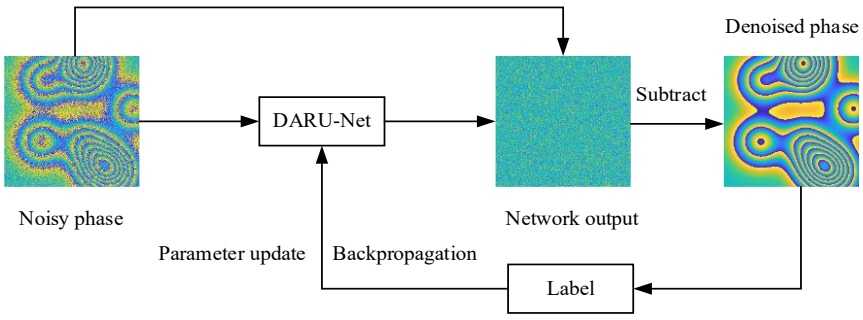


Fig. 2. Overall phase denoising workflow of the proposed DARU-Net.

3.2 Network architecture

The DARU-Net adopts an encoder–decoder architecture for phase denoising, as illustrated in Fig. 3. The encoder consists of alternating DCBs and pooling layers, which progressively extract multi-scale phase features. Unlike the conventional convolutional block in the U-Net, each DCB comprises five dilated convolutional layers that effectively enlarge the receptive field and thereby capture increasingly representative high-level features. The decoder reconstructs the phase features through bilinear interpolation-based upsampling, followed by two convolutional layers (Conv3×3+BN+ReLU). Encoder and decoder stages are linked by skip connections: the upsampled feature map at each decoder level is spatially matched and concatenated along the channel dimension with the corresponding encoder feature map after enhancement by a HAB. This design promotes bidirectional feature exchange across semantic levels, stabilizes network training, and enhances the model's sensitivity to noise patterns across different phase regions, thereby providing richer feature representations.

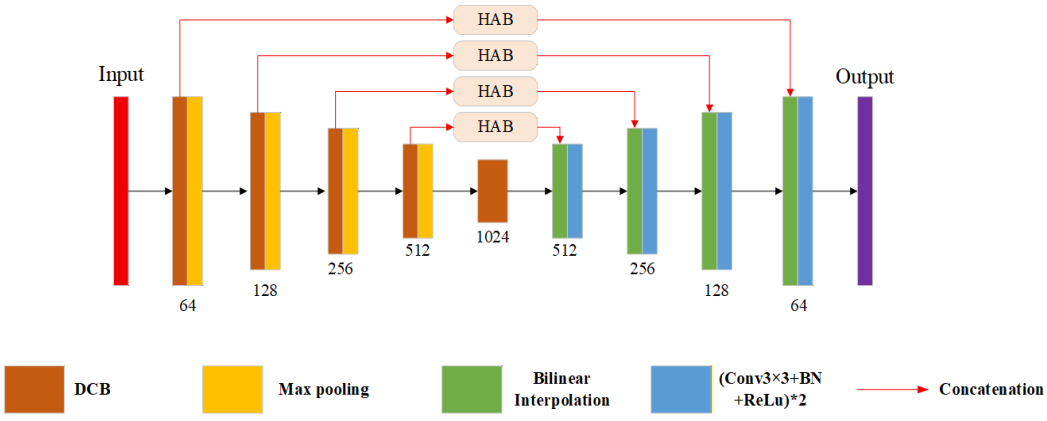


Fig. 3. Overall architecture of the proposed DARU-Net.

The structure of the DCB is illustrated in Fig. 4. This module first applies a standard 3×3 convolution to the input feature map to extract preliminary phase features, and then feeds the output into five parallel dilated convolutional branches with dilation rates of 1, 2, 3, 4 and 5, respectively. Each branch consists of a dilated convolution, batch normalization (BN), a ReLU activation, and a Dropout layer. BN accelerates training and improves model stability; ReLU introduces nonlinearity to enhance representational capacity; Dropout randomly deactivates a subset of neurons to mitigate overfitting and improve generalization. The outputs from the five branches are concatenated and fused by a 1×1 pointwise convolution. Subsequently, the module input features are added to the fused features element-wise via a residual connection. This design effectively alleviates the gradient vanishing issue in deep neural networks, thereby accelerating convergence and improving training

stability.

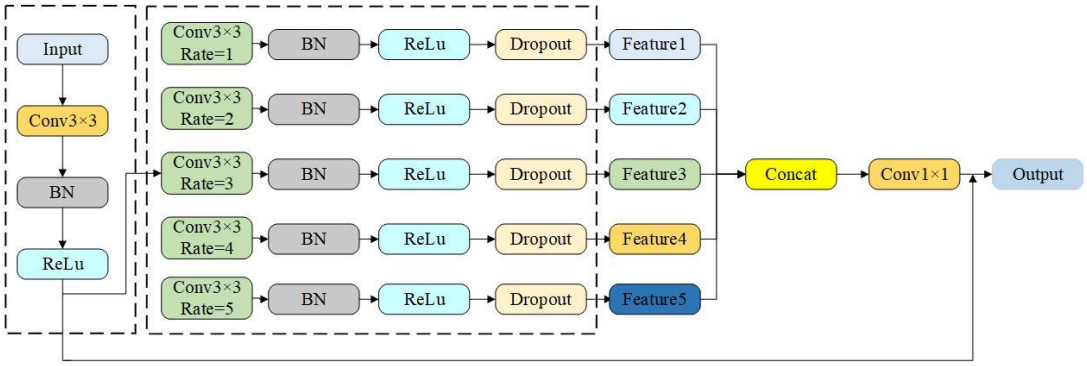


Fig. 4. Structure of the DCB.

The HAB structure is illustrated in Fig. 5. This module combines a channel attention module (CAM) and a spatial attention module (SAM) to adaptively weight the channel and spatial dimensions of the feature maps, thereby enhancing informative phase features while suppressing redundant noise. The implementation process consists of two stages. (1) CAM: global average pooling and global max pooling are applied to the input feature map of size H×W×C to compress the spatial dimensions, yielding two global feature vectors F_1 (1×1×C) and F_2 (1×1×C). These two vectors are then fed into a shared multilayer perceptron (MLP) to adjust the channel dimensions. The MLP adopts an up-then-down strategy to reduce computational cost. Finally, the normalized channel attention weights Mc are generated via a sigmoid activation. (2) SAM: average pooling and max pooling are applied along the channel dimension of the channel-attended feature map (H×W×C), yielding two single-channel feature maps X_1 (H×W×1) and X_2 (H×W×1). These two maps are concatenated along the channel dimension and fused by a 3×3 convolution layer. The normalized spatial attention weights Ms are then obtained via a sigmoid activation function. These weights are multiplied element-wise with the feature map to produce a weighted output, thereby effectively enhancing the representational of phase information.

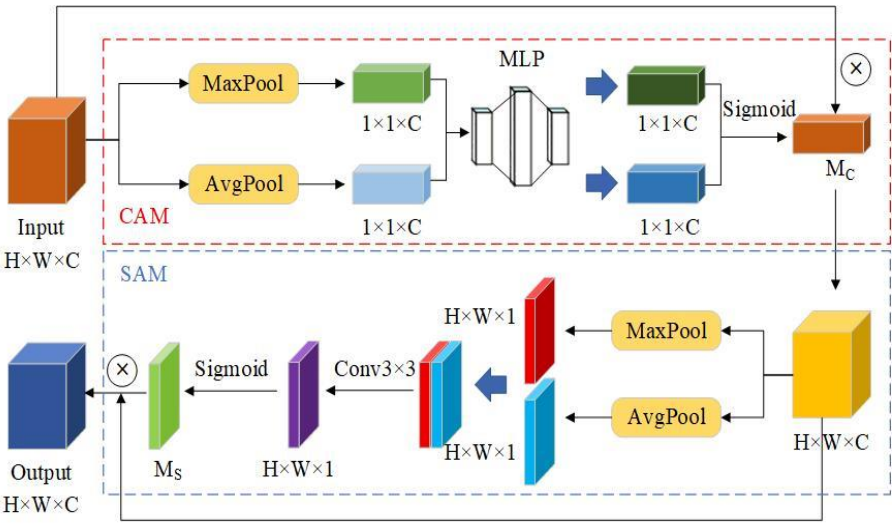


Fig. 5. Structure of the HAB.

3.3 Training configuration and test results analysis

The training parameters for the proposed DARU-Net model are as follows: Huber loss is used as the objective function, and the model is trained using the Adam optimizer. The initial learning rate is set to 5×10^{-4}, with a decay factor of 0.9 and a minimum threshold of 1×10^{-5}. The model is

trained for 100 epochs with a batch size of 32. The simulated dataset comprises 20000 sample in total, randomly split into training, validation, and test subsets at a 7:2:1 ratio. Fig. 6 presents the training and validation loss curves of the DARU-Net model. As the number of training epochs increases, both the training and validation losses decrease rapidly in the early stage and then gradually stabilize. The two loss curves exhibit similar trends, with close final loss values and no pronounced fluctuations, indicating that the proposed model achieves a relatively stable convergence process under the current dataset and training settings.

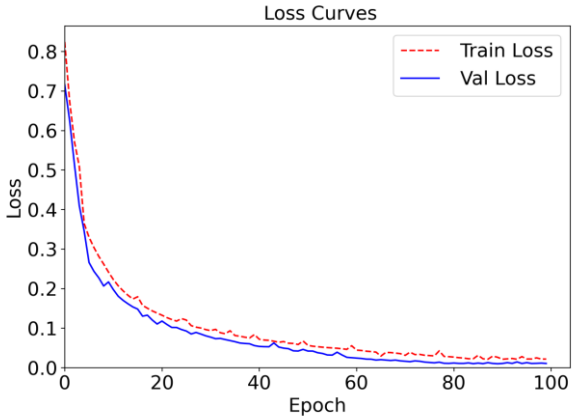


Fig. 6. Training and validation loss curves.

To evaluate the denoising performance of the proposed method, we compare it with four other methods: NLM [26], BM3D [27], CNN [15], and ResU-Net [28]. Fig. 7 gives the denoising results of the different methods on four test samples. The four samples correspond to Gaussian noise, speckle noise, mixed noise I, and mixed noise II, respectively. Mixed noise I combines relatively strong Gaussian noise with relatively weak speckle noise, whereas mixed noise II employs the opposite combination. All methods process identical noisy inputs.

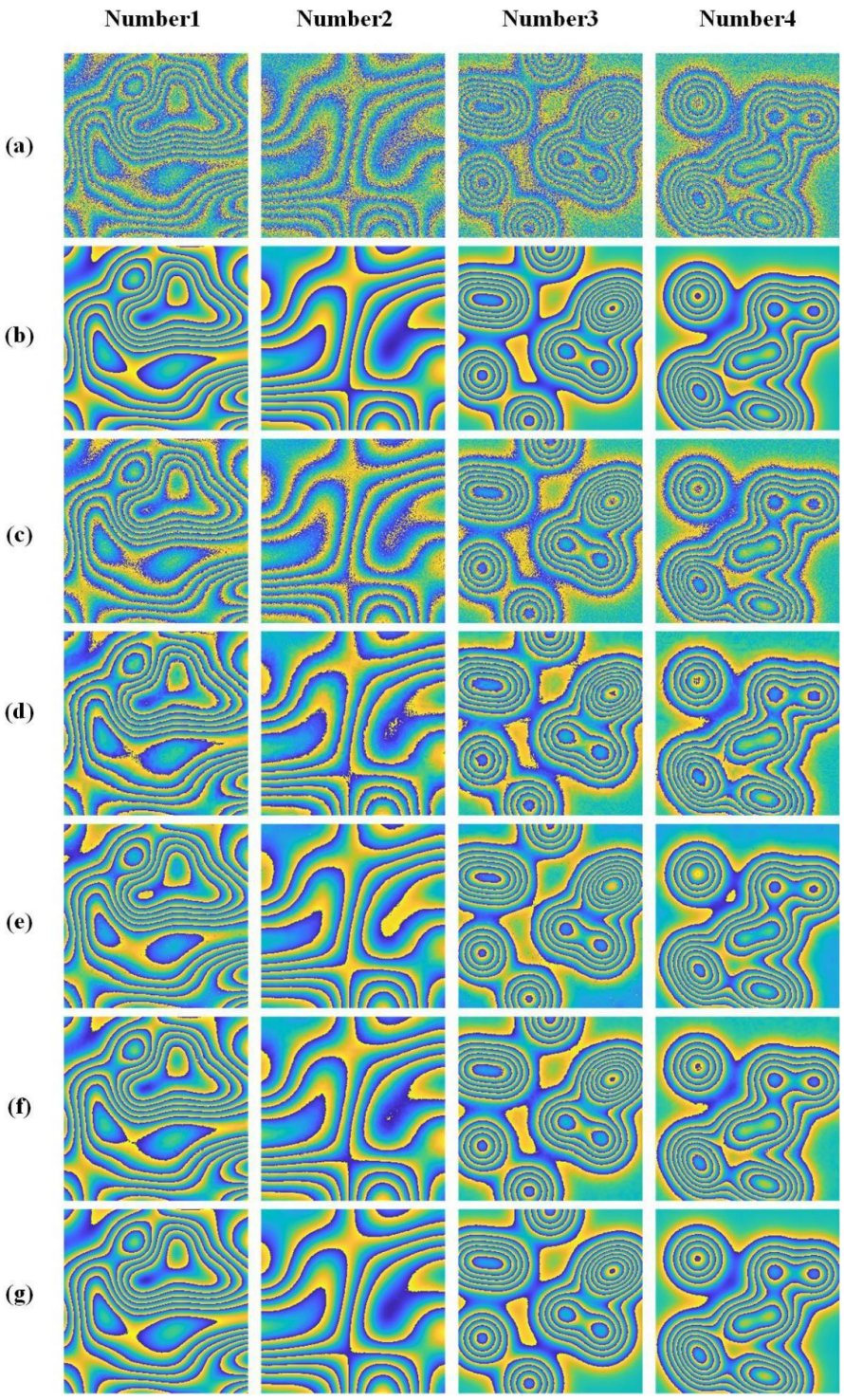


Fig. 7. Comparison of denoising results obtained by different methods on four test samples: (a) input; (b) ground truth; (c) NLM; (d) BM3D; (e) CNN; (f) ResU-Net; (g) proposed method.

As can be seen, NLM performs poorly denoising results in regions with dense wrapped-phase fringes, leaving noticeable residual noise and thus failing to effectively suppress noise. BM3D tends to distort and degrade the fringe structures in the wrapped phase, which may affect the accuracy of subsequent phase unwrapping. CNN and ResU-Net demonstrate superior denoising capability and can better preserve fine details of the wrapped phase, but the denoised phase contours remain insufficiently smooth. In contrast, the proposed method achieves the best results across all four test

samples, exhibiting particularly strong performance in areas with dense fringe patterns. The overall phase contours are recovered more smoothly and completely, thereby validating its effectiveness and robustness in phase denoising.

To quantitatively assess the denoising performance of the five methods, four evaluation metrics are adopted: peak signal-to-noise ratio (PSNR) [29], structural similarity index (SSIM) [21], root mean square error (RMSE) [30], and equivalent number of looks (ENL) [28].

(1) PSNR is used to measure the distortion between a denoised image and a reference image. A higher PSNR value indicates closer similarity to the reference, corresponding to lower distortion and better fidelity. The calculation formula is:

$$\mathrm{MSE} = \frac{1}{M \times N} \times \sum_{i=1}^{M} \sum_{j=1}^{N} \left[I_x(i,j) - I_y(i,j) \right]^2 \tag{1}$$

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}} \right) \tag{2}$$

where MSE is the mean squared error between the reference and denoised images, and $M \times N$ is the image size. I_x and I_y are the grayscale values of the reference and denoised images at (i, j), respectively. MAX is the maximum possible pixel value (255 for 8-bit images, 1 for normalized images).

(2) SSIM is used to measure the similarity of two images in terms of brightness, contrast, and structural information. The closer the value is to 1, the more similar the two images are. Its formula is:

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

where μ_x and μ_y are the pixel mean values of the reference and denoised images, respectively; σ_{x}^2 and σ_y^2 are the variances of the two images; σ_{xy} is their covariance; and c_1 and c_2 are stabilizing constants.

(3) RMSE is used to measure the pixel-level numerical deviation between the denoised and reference images. A smaller RMSE value indicates closer similarity to the reference and thus higher fidelity. The formula is:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{4}$$

(4) ENL is used to quantify the smoothness of homogeneous regions in a denoised image. A higher ENL value indicates a smoother surface and a lower noise level. The ENL is computed as follows:

$$\mathrm{ENL} = \frac{\mu^2}{\sigma^2} \tag{5}$$

Table 1 gives the mean values of the four evaluation metrics for different methods on the test set. As can be shown, compared with NLM, BM3D, CNN and ResU-Net, the proposed method achieves the best performance across all four metrics. Specifically, it improves PSNR by 26.5%, 15.5%, 22.2%, and 11.9%, respectively; SSIM by 55.1%, 9.1%, 7.2%, and 4.9%, respectively; ENL

by 8.7%, 2.8%, 3.5%, and2.3%, respectively; and reduces RMSE by 80.6%, 64.9%, 60.0%, and 44.1%, respectively. These results demonstrate the effectiveness of the proposed approach.

Table 1. Mean values of four evaluation metrics for different methods on the test set.

	PSNR	SSIM	RMSE	ENL
NLM	18.9040	0.6172	0.3668	8.4225
BM3D	20.7283	0.8775	0.2026	8.9079
CNN	19.5880	0.8928	0.1777	8.8521
ResU-Net	21.3835	0.9124	0.1269	8.9496
Proposed method	**23.9324**	**0.9574**	**0.0710**	**9.1590**

3.4 Phase fidelity evaluation after phase unwrapping

To further evaluate the ability of different denoising methods to preserve true phase information, a quantitative phase fidelity evaluation is performed using simulated data for which the absolute-phase ground truth is available. Specifically, the denoised wrapped phases obtained by NLM, BM3D, CNN, ResU-Net, and the proposed method are unwrapped using the discrete cosine transform least-squares (DCT-LS) method [31] and then compared with the corresponding absolute-phase ground truth.

Because least-squares phase unwrapping yields a result with an arbitrary global additive constant, global phase alignment is first performed before error calculation by compensating for the mean phase offset between the unwrapped phase and the absolute-phase ground truth. This removes the influence of the global constant on the evaluation. The RMSE and mean absolute error (MAE) are then used to quantify the recovery accuracy of the absolute phase after unwrapping. Smaller RMSE and MAE values indicate a smaller deviation between the recovered phase and the absolute-phase ground truth and, consequently, higher phase fidelity. The results are presented in Table 2.

Table 2. Phase-fidelity comparison of different methods after DCT-LS phase unwrapping.

	RMSE	MAE
NLM	0.2802	0.2079
BM3D	0.0840	0.0623
CNN	0.0998	0.0752
ResU-Net	0.0777	0.0557
Proposed method	0.0581	0.0396

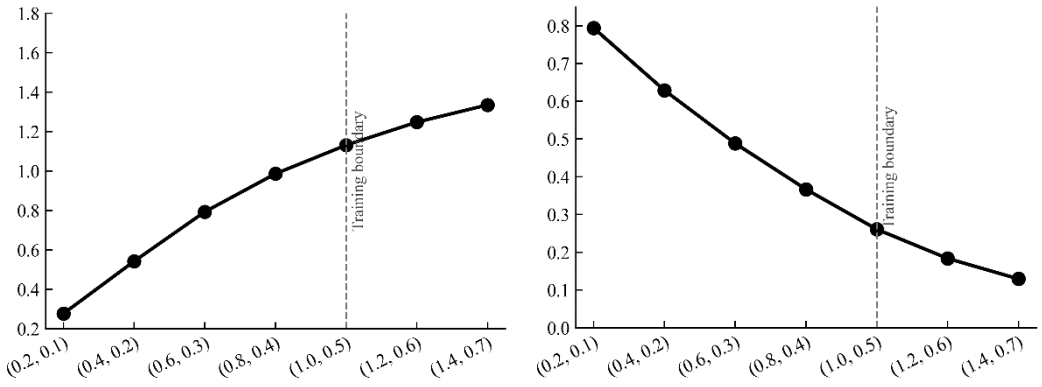
As can be shown, the proposed method achieves the lowest RMSE and MAE among all comparison methods, with values of 0.0581 rad and 0.0396 rad, respectively. Compared with NLM, BM3D, CNN, and ResU-Net, the RMSE is reduced by approximately 79.3%, 30.8%, 41.8%, and 25.3%, respectively, while the MAE is reduced by approximately 80.9%, 36.4%, 47.3%, and 28.8%. Under the same phase unwrapping method, the unwrapped phase obtained using the proposed method therefore exhibits a smaller numerical deviation from the absolute-phase ground truth. This result indicates that the performance improvement is reflected not only in reduced noise and local phase fluctuations but also in a lower absolute-phase recovery error after phase unwrapping, further demonstrating that the proposed method preserves true phase information while suppressing noise and improves phase fidelity after unwrapping.

3.5 Robustness analysis under mixed-noise conditions

To evaluate the adaptability of the proposed method to variations in noise strength and composition, robustness experiments are conducted with Gaussian and speckle noise applied

simultaneously. The network architecture and trained weights are kept unchanged, and the model is not retrained for individual test conditions. Mean RMSE and SSIM values in the wrapped-phase domain are used for quantitative evaluation.

Seven progressively stronger mixed-noise conditions are defined and labeled A through G in ascending order of noise strength. The standard deviation of the Gaussian noise increases from 0.2 to 1.4 in steps of 0.2, while the speckle intensity factor increases from 0.1 to 0.7 in steps of 0.1. Conditions A-E fall within the training noise parameter range, whereas F and G exceed that range and are used to further assess robustness to noise outside the training conditions. The trends of the two metrics under different noise levels are shown in Fig. 8.



(a) (b)

Fig. 8. Robustness evaluation of DARU-Net under different mixed-noise parameter combinations: (a) RMSE; (b) SSIM. The vertical dashed line denotes the upper bound of the training noise parameter range; test conditions to its right lie outside the training range.

As the mixed noise strength increases, the RMSE of the recovered phase gradually increases, while the SSIM gradually decreases. Across the seven tested noise levels, both metrics change progressively, with no abrupt performance drop observed near the boundary of the training range. The results for the two conditions outside the training range continue the same trend. This behavior reflects the progressive degradation of the model under the tested variations in noise strength and provides experimental evidence of its noise robustness. The results are listed in Table 3.

Table 3. Denoising performance under mixed noise of different strengths.

	A	B	C	D	E	F	G
RMSE	0.2764	0.5418	0.7926	0.9857	1.1308	1.2473	1.3346
SSIM	0.7942	0.6291	0.4884	0.3667	0.2608	0.1836	0.1294

To further evaluate robustness to changes in mixed noise composition, two cross-composition conditions are tested: relatively strong Gaussian noise combined with weak speckle noise (1.0, 0.1), and relatively weak Gaussian noise combined with strong speckle noise (0.2, 0.5). The results are listed in Table 4.

Table 4. Denoising performance under different mixed noise compositions.

	(1.0, 0.1)	(0.2, 0.5)
RMSE	0.7056	0.8387
SSIM	0.5371	0.5136

As can be observed, under the two cross-composition conditions, the RMSE values remain both below 0.84 and the SSIM values remain above 0.51, providing further evidence of the phase recovery capability of the same model under different noise compositions. Together with the

progressive metric changes observed as noise strength increases, these experiments demonstrate the robustness of the proposed method to the tested variations in mixed noise strength and composition.

3.6 Computational efficiency analysis

To further evaluate computational efficiency, the parameter count, floating-point operations, peak GPU memory consumption, and inference time of CNN, ResU-Net, and the proposed method were compared. All models were evaluated under identical test conditions, and the results are presented in Table 5.

Table 5. Computational efficiency comparison of different deep-learning methods.

	Params (M)	FLOPs (G)	GPU Memory (MB)	Inference Time (ms)
CNN	31.03	27.27	166.69	1.10
ResU-Net	63.85	50.35	290.55	2.55
Proposed method	87.82	52.93	399.97	3.62

As can be seen, the proposed method has a higher parameter count, computational cost, and memory consumption than CNN and ResU-Net, mainly due to the introduced multi-scale dilated convolutions and attention mechanism. Despite this additional computational cost, its average inference time is 3.62ms, retaining millisecond-level inference speed and providing a favorable trade-off between denoising performance and computational complexity.

3.7 Ablation study

To quantitatively analyze the contribution of each component of the proposed method to wrapped-phase denoising, five network variants are designed: U-Net, U-Net-RL, U-Net-RL-DCB, U-Net-RL-HAB, and the DARU-Net. The U-Net variant directly predicts the clean wrapped phase; U-Net-RL employs residual learning to predict the periodic noise residual in the noisy wrapped phase; U-Net-RL-DCB and U-Net-RL-HAB add the dilated convolutional block and the HAB, respectively, to the residual-learning baseline; and the DARU-Net incorporates residual learning, DCB, and HAB.

Except for the prediction target used to evaluate residual learning and the network modules indicated by the model names, all models use identical training, validation, and test splits, as well as the same loss function, optimizer, learning rate and decay strategy, number of training epochs, batch size, best-model selection criterion, and evaluation procedure, thereby ensuring a fair comparison among the network variants.

Table 6. Ablation results for the components of DARU-Net.

	PSNR	SSIM	RMSE	ENL
U-Net	19.1834	0.6972	0.1051	6.6423
U-Net-RL	19.6215	0.7136	0.1009	6.7546
U-Net-RL-DCB	22.4389	0.8817	0.0806	8.5873
U-Net-RL-HAB	22.8731	0.8921	0.0773	8.8462
DARU-Net	23.9876	0.9579	0.0708	9.1874

Table 6 shows that each component improves phase denoising performance. As can be shown, relative to the U-Net, residual learning reduces RMSE by 4.0%. Adding DCB and HAB separately to U-Net-RL further reduces RMSE by 20.1% and 23.4%, respectively. The DARU-Net reduces RMSE by 29.8% relative to U-Net-RL and by 32.6% relative to the U-Net, while achieving the best PSNR, SSIM, and ENL values. These results indicate that DCB and HAB provide complementary benefits by enhancing multi-scale feature extraction and the representation of important features.

4 Experimental verifications

4.1 Experimental setup

To further validate the effectiveness of the proposed method, a transmission digital holographic optical path imaging system is constructed, as illustrated in Fig. 9. A laser with a wavelength of 632.8nm is used as the illumination source, the image sensor has a pixel pitch of 7.4μm, the recording distance is 800mm, and the image resolution used for phase reconstruction and denoising is 256×256 pixels. A USAF-1951 resolution target and a micro-hole specimen with an aperture of 800μm, shown in Fig. 10, are used as the test specimens. Holograms of the regions marked by the red boxes are acquired for experimental validation, and the denoising results are compared with those obtained using NLM, BM3D, CNN, and ResU-Net.

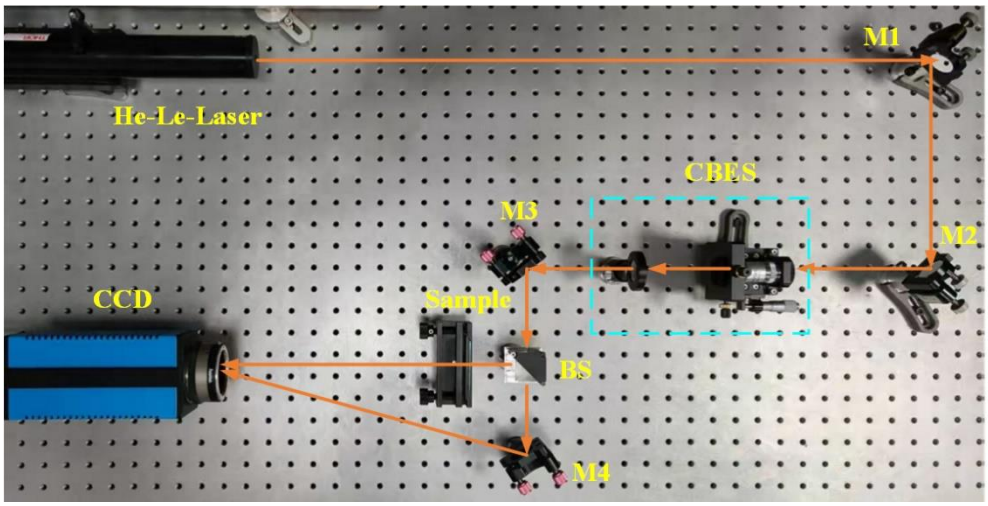
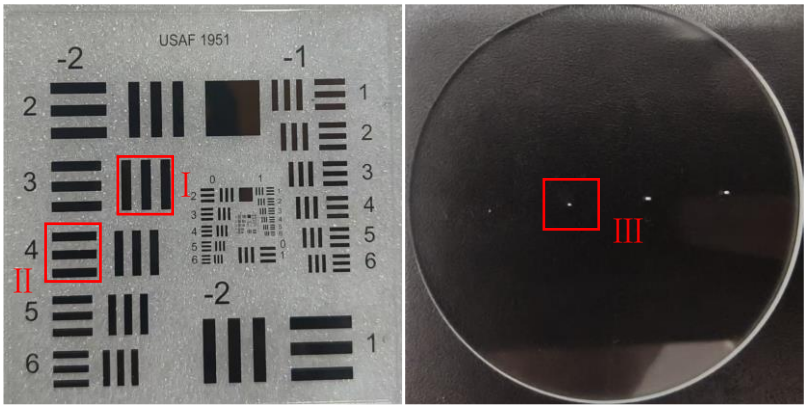


Fig. 9. Experimental setup.



(a) (b)

Fig. 10. Test specimens: (a) USAF-1951 resolution target; (b) micro-hole specimen.

4.2 Experimental results and analysis

To quantitatively assess the quality of the denoised wrapped phase, the discrete cosine transform least-squares (DCT-LS) method is employed to perform phase unwrapping. Three metrics, such as horizontal gradient standard deviation (HGSD), vertical gradient standard deviation (VGSD), and local variance mean (LVM) are selected to evaluate the model performance [3]. HGSD and VGSD are used to calculate the standard deviations of phase gradients in the horizontal and vertical directions, respectively, to assess the stability of phase variations; smaller values indicate

more uniform phase variations.

First, the phase gradients in the horizontal and vertical directions are computed as:

$$G_x(i,j) = U(i,j+1) - U(i,j) \tag{6}$$

$$G_y(i,j) = U(i+1,j) - U(i,j) \tag{7}$$

where $U(i,j)$ represents the unwrapped phase value at coordinate (i,j).

HGSD and VGSD are then calculated respectively by:

$$\mathrm{HGSD} = \sqrt{\frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[G_x(i,j) - \bar{G}_x]^2} \tag{8}$$

$$\mathrm{VGSD} = \sqrt{\frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[G_y(i,j) - \bar{G}_y]^2} \tag{9}$$

where M and N are the height and width of the phase map, respectively; $\bar{G}_x$ and $\bar{G}_y$ are the mean values of the horizontal and vertical gradients, respectively.

LVM is used to measure the average local variance in a phase image, reflecting the smoothness of these regions. A smaller LVM value indicates smoother regions and lower noise levels. Where (i, j) denotes the index of a local region, and (x, y) denotes a pixel position within in that region. Its formula is:

$$\mathrm{LVM} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{1}{\left|R_{i,j}\right|}\sum_{(x,y)\in R_{i,j}}\left(U(x,y) - \overline{U_{ij}}\right)^2 \tag{10}$$

where $R_{i,j} = \{U(x,y) \mid x \in [i, i+S-1], y \in [j, j+S-1]\}$, S represents the window size, $\left|R_{i,j}\right|$ represents the number of elements in the local region $R_{i,j}$, and $\overline{U_{ij}}$ represents the mean phase.

Fig. 11 shows the phase denoising results for the region I hologram obtained using different methods, together with the corresponding unwrapped-phase profiles at the positions indicated by the red dashed lines. As can be observed, the wrapped-phase maps obtained using NLM and BM3D still contain obvious residual noise, with some phase edges appearing blurred and exhibiting unnatural transitions; the corresponding unwrapped-phase profiles display pronounced fluctuations and distortion. CNN and ResU-Net still exhibit slight fluctuations in regions with large phase gradient variations, indicating limited adaptability to complex phase changes. In contrast, the proposed method achieves the best overall performance, effectively suppressing speckle noise while accurately preserving fine phase edge details. The cross-sectional profile remains smooth, even in regions with drastic phase gradient changes.

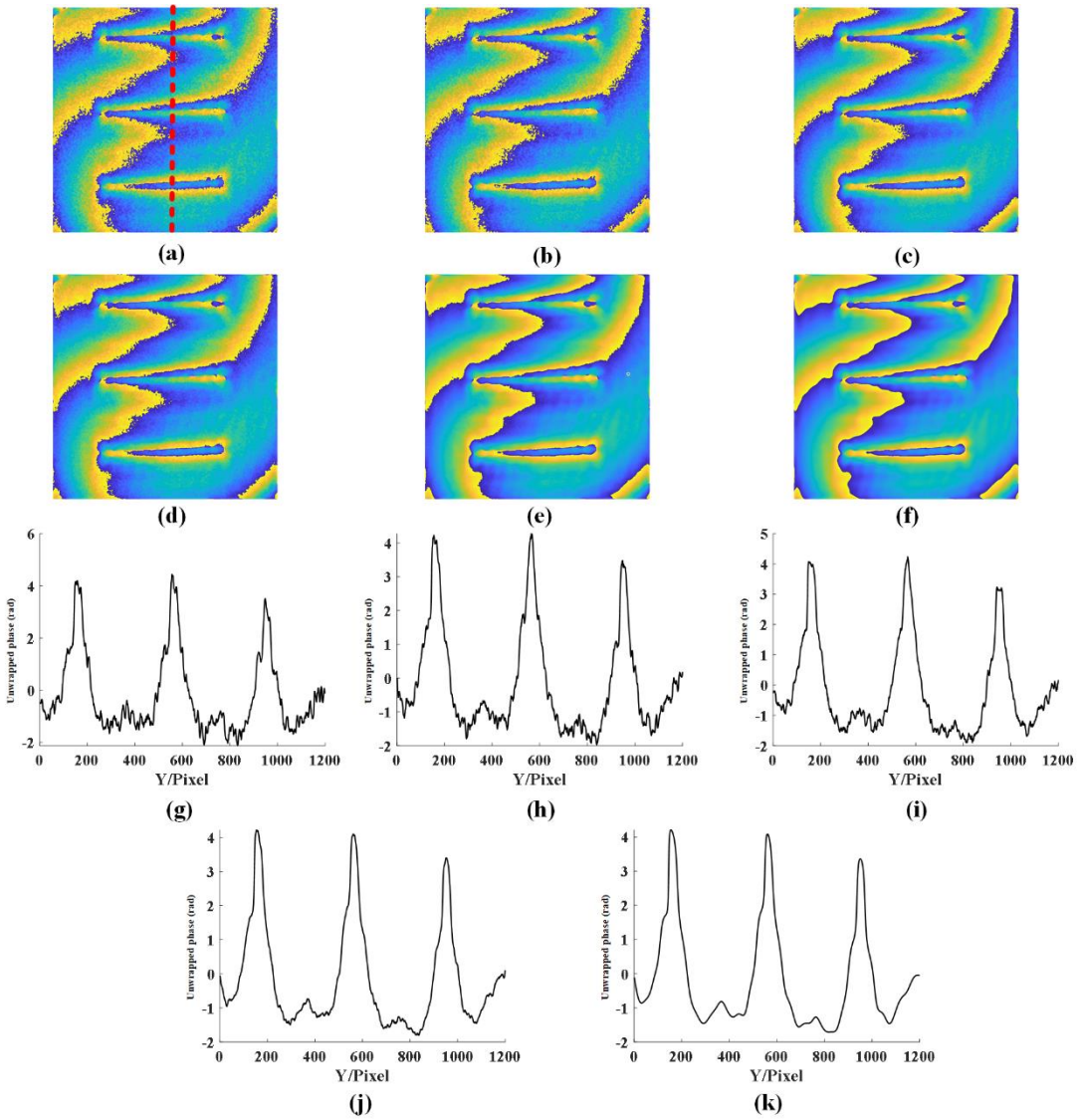


Fig. 11. Comparison of phase denoising results for region I hologram using different methods: (a) noisy phase; (b-f) denoised results obtained by NLM, BM3D, CNN, ResU-Net and the proposed method; (g-k) corresponding unwrapped phase cross-sectional profiles along the red dashed lines.

Table 7 presents the three evaluation metrics for phase denoising of the region I hologram using different methods. As can be seen, NLM yields the highest HGSD, VGSD and LVM values, indicating significant gradient fluctuations and high local variance in the denoised phase, with noticeable residual noise and edge blurring. BM3D, CNN and ResU-Net show some improvements, but none match the proposed method. In contrast, the proposed method achieves optimal values of 0.0194, 0.0427 and 0.0039 for the three metrics, respectively, demonstrating its superiority in noise suppression, gradient stabilization, and local uniformity enhancement.

Table 7. Evaluation metrics for phase denoising results of the region I hologram using different methods.

	NLM	BM3D	CNN	ResU-Net	Proposed method
HGSD	0.0839	0.0589	0.0507	0.0264	**0.0194**
VGSD	0.0924	0.0704	0.0635	0.0465	**0.0427**
LVM	0.0182	0.0106	0.0086	0.0046	**0.0039**

Fig. 12 shows the phase denoising results for the region II hologram obtained by various

methods, and Table 8 gives the corresponding the values of the three evaluation metrics. As can be observed, NLM and BM3D exhibit limited denoising capability, leaving large areas of speckle noise in the wrapped phase maps. CNN and ResU-Net improve the phase quality and better preserve fringe details, but residual noise remains along the edges, and the cross-sectional profiles remain insufficient smooth. In contrast, the proposed method outperforms all other methods across all three metrics, demonstrating superior phase information integrity and clarity. This confirms its significant advantages in noise robustness and phase fidelity.

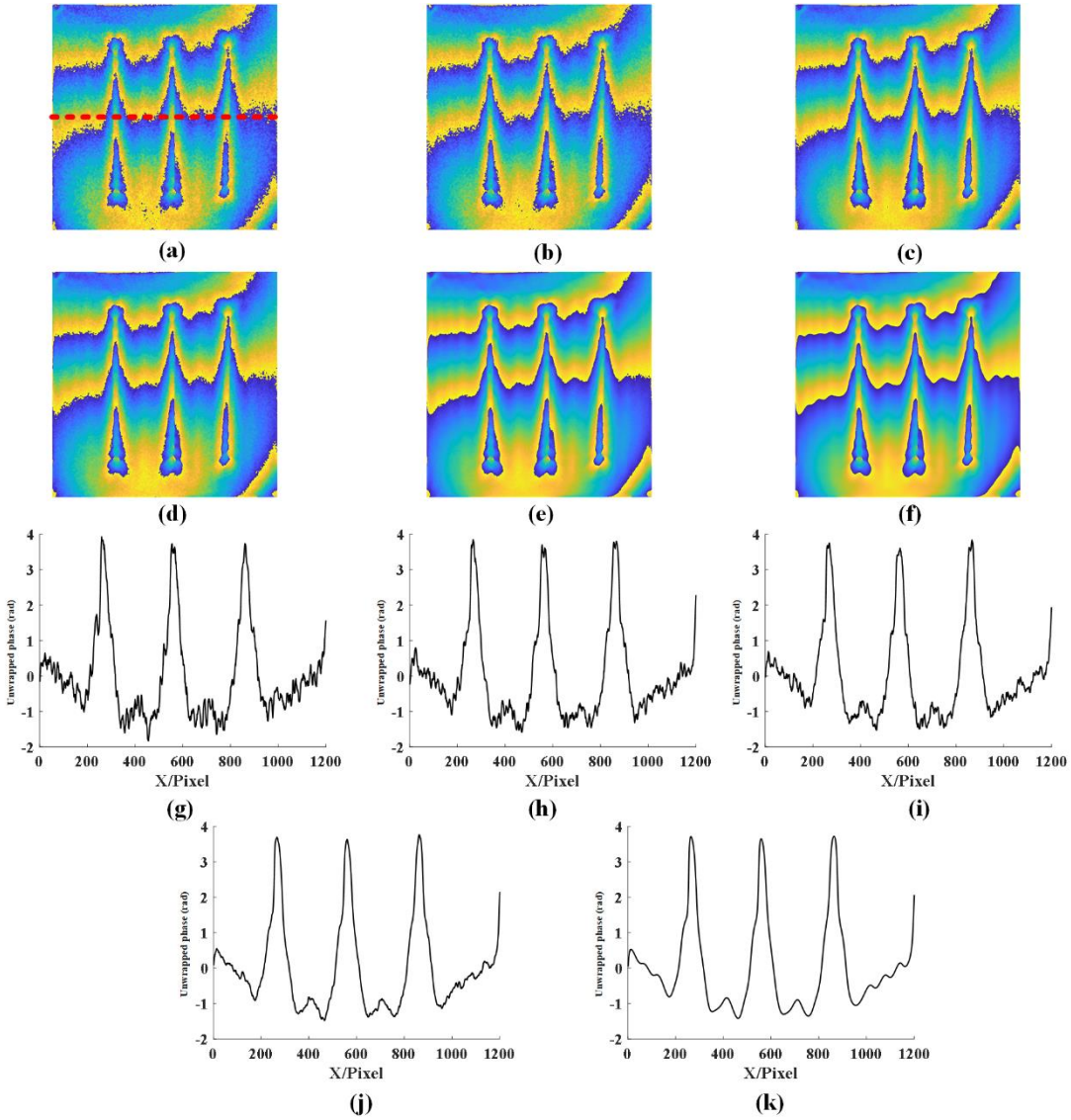


Fig. 12. Comparison of phase denoising results for region II hologram using different methods: (a) noisy phase; (b-f) denoised results obtained by NLM, BM3D, CNN, ResU-Net and the proposed method; (g-k) corresponding unwrapped phase cross-sectional profiles along the red dashed lines.

Table 8. Evaluation metrics for phase denoising results of the region II hologram using different methods.

	NLM	BM3D	CNN	ResU-Net	Proposed method
HGSD	0.0935	0.0717	0.0653	0.0489	**0.0253**
VGSD	0.0839	0.0588	0.0505	0.0263	**0.0191**
LVM	0.0187	0.0112	0.0091	0.0052	**0.0045**

Fig. 13 presents the phase denoising results for the region III hologram obtained by different

methods, and Table 9 summarizes the corresponding three evaluation metrics. As can be shown, NLM and BM3D cause distortion of some fine fringe structures and localized phase jumps, and produce cross-sectional profiles with large fluctuations, resulting in poor overall accuracy. Although CNN and ResU-Net preserve the general phase contours, residual noise remains around the micro-hole and in the edge regions. The proposed method achieves the best overall performance. The denoised wrapped phase exhibits clear fringe information and optimal micro-hole morphology recovery, without obvious jumps or distortions and smooth, stable cross-sectional profiles. In terms of the evaluation metrics, the proposed method obtains HGSD, VGSD and LVM values of 0.0128, 0.0126 and 0.0004, respectively, confirming its capability for distortion-free denoising and high-precision phase morphology reconstruction under complex noise conditions.

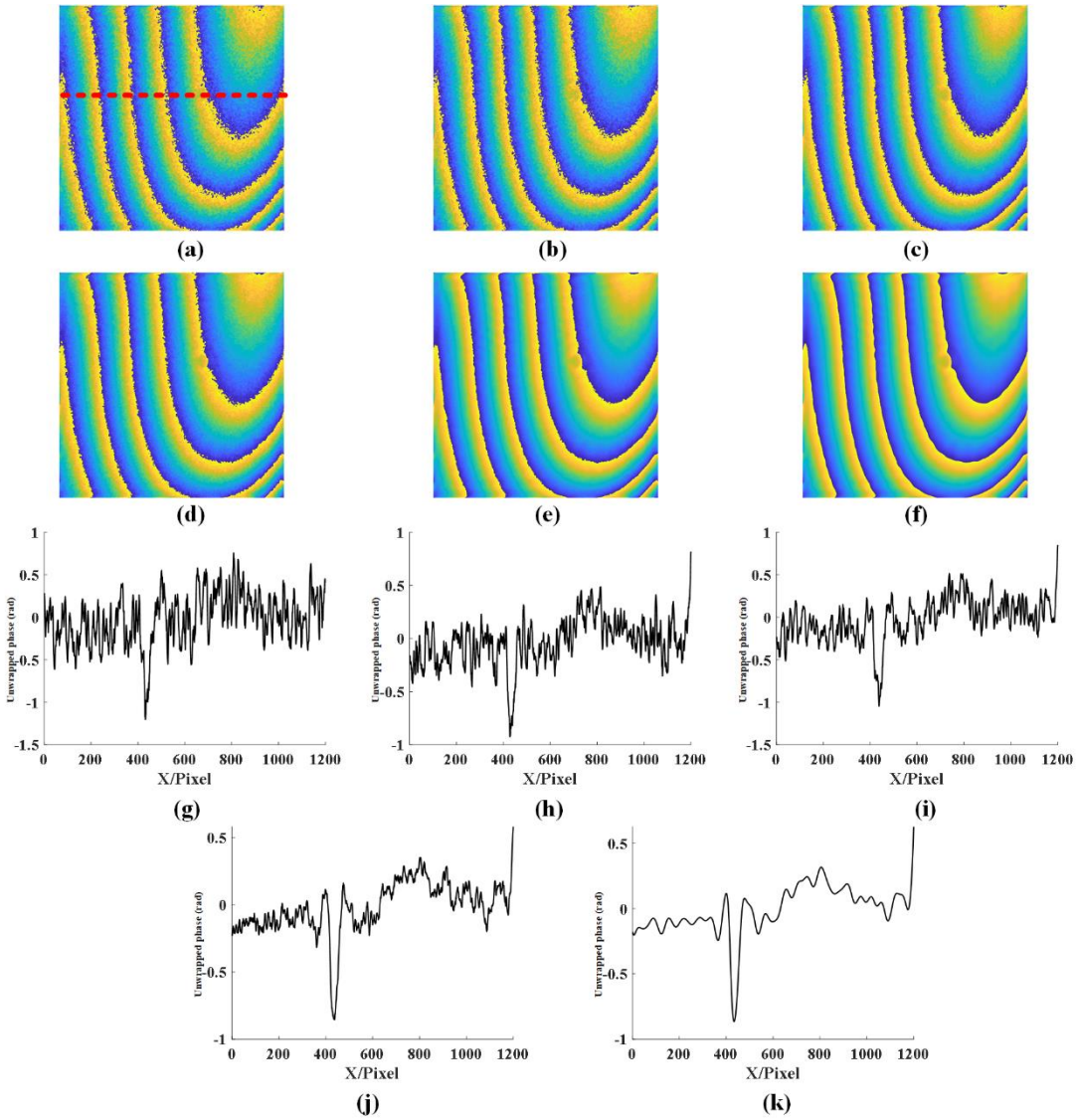


Fig. 13. Comparison of phase denoising results for region III hologram using different methods: (a) noisy phase; (b-f) denoised results obtained by NLM, BM3D, CNN, ResU-Net and the proposed method; (g-k) corresponding unwrapped phase cross-sectional profiles along the red dashed lines.

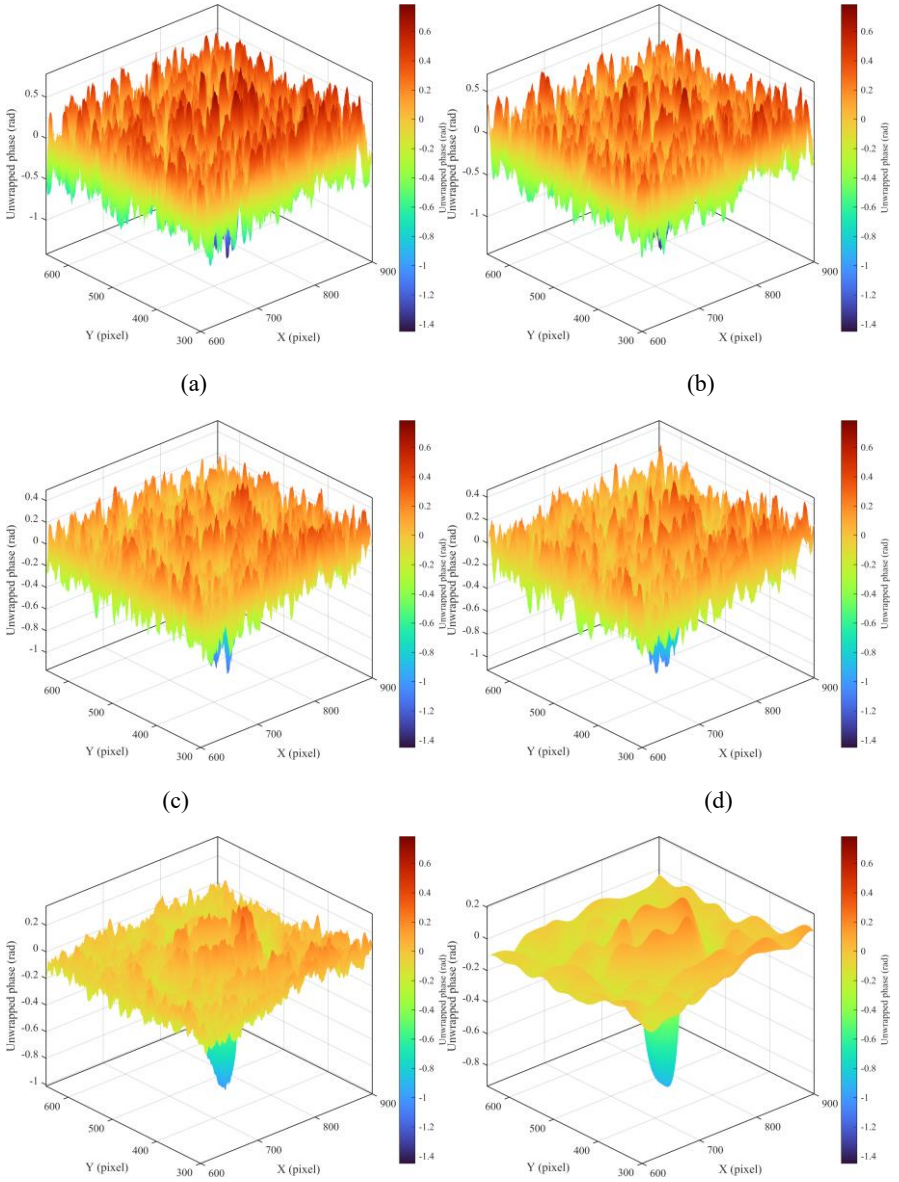
Table 9. Evaluation metrics for phase denoising results of the region III hologram using different methods.

	NLM	BM3D	CNN	ResU-Net	Proposed method
HGSD	0.0826	0.0571	0.0483	0.0223	**0.0128**

VGSD	0.0827	0.0570	0.0485	0.0222	**0.0126**
LVM	0.0146	0.0071	0.0052	0.0011	**0.0004**

Across all three experimental cases, traditional NLM and BM3D methods are significantly outperformed by deep learning-based methods in terms of noise suppression, fringe detail preservation and phase unwrapping accuracy, and are insufficient for high-precision phase processing tasks. While CNN and ResU-Net demonstrate improved performance and can handle phase processing in conventional tasks, they remain suboptimal in detail recovery and achieving smooth transitions at complex phase edges. In contrast, the proposed method achieves the best results across all three experiments. It effectively suppresses speckle noise while precisely preserving fringe details and phase edge features. The phase unwrapping results are free from jumps or distortion, and the phase cross-sectional profiles remain smooth and stable, fully validating its significant advantages in noise robustness and phase fidelity.

To further compare the recovery of phase structures by different denoising methods, Fig. 14 presents the three-dimensional phase distributions obtained after DCT-LS unwrapping for the Region III micro-hole structure.



(e) (f)

Fig. 14. Three-dimensional unwrapped phase results for the Region III micro-hole structure obtained using different methods: (a) noisy phase; (b)-(f) NLM, BM3D, CNN, ResU-Net, and the proposed method, respectively.

As can be observed, the results obtained using NLM and BM3D still exhibit pronounced phase fluctuations, and the micro-hole contour remains affected by noise. Although CNN and ResU-Net suppress part of the noise, local undulations persist on the phase surface. In contrast, the three-dimensional phase surface obtained using the proposed method is smoother, while the depression morphology and edge contour of the micro-hole are more completely preserved without obvious structural distortion. The proposed method therefore achieves the best overall result. These observations demonstrate that it effectively suppresses phase noise while preserving the phase morphology and structural details of the micro-hole.

5 Conclusions

To address the issue of incomplete phase information caused by speckle noise, we propose a DARU-Net-based method for digital holographic phase denoising. This method adopts a residual learning strategy, in which the noisy phase is fed into the network as input, and the network is trained to predict the noise residual. The denoised phase is then recovered by subtracting the predicted residual from the input phase. The main conclusions are as follows:

(1) Based on the U-Net architecture, multiple DCBs with varying dilation rates are introduced to achieve multi-scale feature extraction by enlarging the receptive field, thereby enhancing the network's sensitivity to noise at different scales in phase maps.

(2) CAM and SAM are incorporated into the skip connections. Through adaptive learning of feature weights, key phase information is highlighted, effectively enhancing phase representational and edge structure preservation.

(3) The ablation results demonstrate that residual learning, DCB, and HAB each improve wrapped-phase denoising performance. The DARU-Net achieves the best PSNR, SSIM, RMSE, and ENL values, confirming the complementary effects of multi-scale contextual feature extraction and spatial-channel adaptive feature enhancement.

(4) A simulated dataset covering diverse fringe densities and noise levels is constructed using RME and GFS for model training and testing. Simulation results demonstrate that the proposed method outperforms the other methods across all four metrics: PSNR, SSIM, ENL and RMSE.

(5) A transmission digital holographic optical path imaging system is constructed to acquire the holograms of test specimens for experimental validation. Experimental results demonstrate that the proposed method achieves superior performance across all three metrics: HGSD, VGSD and LVM.

In summary, the proposed method effectively suppresses speckle noise, exhibits notable advantages in noise robustness and phase fidelity, and significantly improves phase unwrapping accuracy.

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] Wang Z, Ma H, Chen Y, et al. Autofocusing in digital holography based on an adaptive genetic algorithm[J]. Journal of the Optical Society of America A, 2024, 41(6): 976-987.

[2] Chen Y, Wang Q, Zhang G, et al. Phase unwrapping in digital holography based on SRDU-net[J]. Optics Communications, 2024, 573: 131055.

[3] Wang Z, Ma H, Chen Y, et al. Phase unwrapping based on self-attention dense residual network[J]. Optics and Laser Technology, 2025, 191: 113321.

[4] 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.

[5] Bianco V, Memmolo P, Leo M, et al. Strategies for reducing speckle noise in digital holography[J]. Light: Science & Applications, 2018, 7: 48.

[6] Pradeep S and Nirmaladevi P. A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2021, 1055(1): 012116.

[7] Dohet-Eraly J, Yourassowsky C, Mallahi A E, et al. Quantitative assessment of noise reduction with partial spatial coherence illumination in digital holographic microscopy[J]. Optics Letters, 2015, 41(1): 111-114.

[8] Restrepo J, Correa-Rojas N and Herrera-Ramirez J. Speckle noise reduction in digital holography using a DMD and multi-hologram resampling[J]. Applied Sciences, 2020, 10(22): 8277.

[9] Jiang H, Ma Y, Su Z, et al. Speckle-interferometric phase fringe patterns de-noising by using fringes' direction and curvature[J]. Optics and Lasers in Engineering, 2019, 119: 30-36.

[10] Zada S, Tounsi Y, Kumar M, et al. Contribution study of monogenic wavelets transform to reduce speckle noise in digital speckle pattern interferometry[J]. Optical Engineering, 2019, 58(3): 034109.

[11] Li J, Tang C, Xu M, et al. DBDNet for denoising in ESPI wrapped phase patterns with high density and high speckle noise[J]. Applied Optics, 2021, 60(32): 10070-10079.

[12] Hao F, Tang C, Min M, et al. Batch denoising of ESPI fringe patterns based on convolutional neural network[J]. Applied Optics, 2019, 58(13):3338-3346.

[13] Chen J, Liao H, Kong Y, et al. Speckle denoising based on Swin-UNet in digital holographic interferometry[J]. Optics Express, 2024, 32(19): 33465-33482.

[14] Fang Q, Li Q, Song Q, et al. Convolutional and Fourier neural networks for speckle denoising of wrapped phase in digital holographic interferometry[J]. Optics Communications, 2024, 550: 129955.

[15] Yan K, Yu Y, Sun T, et al. Wrapped phase denoising using convolutional neural networks[J]. Optics and Lasers in Engineering, 2020, 128: 105999.

[16] Fang Q, Xia H, Song Q, et al. Speckle denoising based on deep learning via a conditional generative adversarial network in digital holographic interferometry[J]. Optics Express, 2022, 30(12): 20666-20683.

[17] Montresor S, Tahon M, Laurent A, et al. Computational de-noising based on deep learning for phase data in digital holographic interferometry[J]. APL Photonics, 2020, 5(3): 030802.

[18] Gurrola-Ramos J, Dalmau O, Alarcón T. U-Net based neural network for fringe pattern denoising[J]. Optics and Lasers in Engineering, 2022, 149: 106829.

[19] Ram V S, Gannavarpu R. Robust interferogram processing using deep learning and signal subspace method for phase derivative estimation[J]. Optics and Laser Technology, 2025, 192: 113616.

[20] Ram V S, Gannavarpu R. Fringe pattern normalization using conditional generative adversarial networks[J]. Optik, 2024, 313: 171999.

[21] Chen Y, Li P, Ma H, et al. Phase unwrapping in digital holography based on the Dv3p-IRA[J]. Applied Optics, 2025, 64(21): 6042-6051.

[22] Wang K, Kemao Q, Di J, et al. Deep learning spatial phase unwrapping: a comparative review[J]. Advanced Photonics Nexus, 2022, 1(1): 014001.

[23] Qin Y, Wan S, Wan Y, et al. Direct and accurate phase unwrapping with deep neural network[J]. Applied Optics, 2020, 59(24): 7258-7267.

[24] Vengala K, Paluru N and Subrahmanyam Gorthi R. 3D deformation measurement in digital holographic interferometry using a multitask deep learning architecture[J]. Journal of the Optical Society of America A, 2021, 39(1): 167-176.

[25] Sumanth K, Ravi V and Gorthi R. A multi-task learning for 2D phase unwrapping in fringe projection[J]. IEEE Signal Processing Letters, 2022, 29: 797-801.

[26] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005, 2: 60-65.

[27] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.

[28] Lan Y and Zhang X. Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet[J]. IEEE Access, 2020, 8: 195327-195340.

[29] Wang Y, Hu Y, Li C, et al. Application of optimized 2D S-transform in digital speckle interferometry phase pattern denoising[J]. Optics Communications, 2023, 533: 129220.

[30] Yan L, Wang Q, Huang L, et al. Phase unwrapping of digital holographic microscopy using adaptive region segmentation and phase derivative calibration with respect to fringe density[J]. Optics and Lasers in Engineering, 2022, 148: 106780.

[31] Ghiglia D C, Romero L A. Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods[J]. Journal of the Optical Society of America A, 1994, 11(1): 107-117.