


Article

Wavefront Reconstruction Using Two-Frame Random Interferometry Based on Swin-Unet

Xindong Shu ^{1,2}, Baopeng Li ^{1,*}  and Zhen Ma ¹

¹ Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; shuxindong21@mailsucas.ac.cn (X.S.); mazn@opt.ac.cn (Z.M.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: libaopeng@opt.ac.cn

Abstract: Due to its high precision, phase-shifting interferometry (PSI) is a commonly used optical component detection method in interferometers. However, traditional PSI, which is susceptible to environmental factors, is costly, with piezoelectric ceramic transducer (PZT) being a major contributor to the high cost of interferometers. In contrast, two-frame random interferometry does not require precise multiple phase shifts, which only needs one random phase shift, reducing control costs and time requirements, as well as mitigating the impact of environmental factors (mechanical vibrations and air turbulence) when acquiring multiple interferograms. A novel method for wavefront reconstruction using two-frame random interferometry based on Swin-Unet is proposed. Besides, improvements have been made on the basis of the established algorithm to develop a new wavefront reconstruction method named Phase U-Net plus (PUN+). According to training the Swin-Unet and PUN+ with a large amount of simulated data generated by physical models, both of the methods accurately compute the wrapped phase from two frames of interferograms with an unknown phase step (except for multiples of π). The superior performance of both methods is effectively showcased by reconstructing phases from both simulated and real interferograms, in comprehensive comparisons with several classical algorithms. The proposed Swin-Unet outperforms PUN+ in reconstructing the wrapped phase and unwrapped phase.

Keywords: wavefront reconstruction; deep learning; phase-shifting interferometry



Citation: Shu, X.; Li, B.; Ma, Z. Wavefront Reconstruction Using Two-Frame Random Interferometry Based on Swin-Unet. *Photonics* **2024**, *11*, 122. <https://doi.org/10.3390/photonics11020122>

Received: 2 December 2023

Revised: 15 January 2024

Accepted: 23 January 2024

Published: 28 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

PSI is one of the most popular techniques in optical metrology [1], known for its high robustness and accuracy. Traditional PSI require multiple interferograms with fixed and known phase shifts [2–4], but acquiring interferograms is time-consuming and susceptible to adverse effects from mechanical vibrations, environmental turbulence, and temperature variations. Therefore, it is desirable to minimize the number of interferograms. However, single-frame interferometry for wavefront reconstruction requires additional prior information for phase ambiguity. Takeda et al. [5] proposed a Fourier transform-based method that introduces a large spatial carrier frequency by adding significant tilt to the testing object or reference surface, which allows separating the phase from other information in the frequency domain. However, the method is not suitable for interferograms with closed fringes and suffers from fringe densification due to high tilt. Ge et al. [6] specified the concavity and convexity of the phase and were able to recover a monotonic phase from a single interferogram with closed fringes. The concavity or convexity of the phase is difficult to determine. Different from single-frame interferometry, two-frame random interferometry effectively solves the phase ambiguity problem and provides better reconstruction accuracy, which introduces an unknown phase step between the two frames of interferograms, significantly reducing detection costs and shortening capture time compared to traditional phase-shifting interferometry. In comparison, two-frame random interferometry achieves

a good balance between capturing time and reconstruction accuracy, which has received extensive research attention.

Kreis et al. [7] proposed a Fourier transform-based two-frame interferometric phase-shifting wavefront reconstruction, known as the Kreis method, which demodulated the phase from two frames of interferograms without introducing sign ambiguity. The original Kreis method does not consider the pre-filtering process, making it sensitive to noise in practical applications. Vargas et al. [8] proposed a two-frame reconstruction method based on regularized optical flow (OF), which calculated the motion direction of fringes from the two frames of interferograms, applying the spiral phase transform to one interferogram to obtain the wrapped phase and combing it with the fringe motion direction map to eliminate sign ambiguity. But OF requires subtracting the direct current (DC) component of interferograms in advance. Vargas et al. [9] proposed another self-tuning (ST) two-frame phase-shifting method that did not need to know the phase-step between interferograms, by which the quadrature filter was tuned sequentially at a predetermined discrete set of frequencies within $[0, 2\pi]$, and the reconstructed wrapped phase was obtained. ST method requires subtracting the DC component in interferograms before wavefront reconstruction. Besides, Vargas et al. [10] proposed a two-frame reconstruction method based on Gram–Schmidt (GS) orthogonalization. GS method demodulates the wrapped phase by treating the interferograms as independent vectors, which has the advantages of high efficiency and accuracy but also requires subtracting the DC component in advance.

In recent years, with the development of artificial neural networks, the typical U-shaped convolutional neural network U-Net [11] has been proposed and applied to biomedical image segmentation, consisting of a symmetric encoder–decoder with skip connections. The convolution in U-Net has local restrictions and does not effectively utilize global information until many layers of convolution are performed. Inspired by the tremendous success of Transformers in natural language processing (NLP), researchers have attempted to introduce Transformers into the field of computer vision (CV), Liu et al. [12] proposed Swin Transformer for image recognition tasks, whose self-attention mechanism has a natural advantage in extracting global information. Inspired by Swin Transformer, Cao et al. [13] proposed Swin-Unet for medical image segmentation, combining U-Net with Swin Transformer and achieving higher image segmentation accuracy.

The remarkable achievements of deep learning in CV activated researchers to explore its application in optical metrology. Different from traditional “physical based” approaches, deep learning-supported optical metrology is based on “data-driven” principles. The field of optical metrology has been developed inspired by the achievement of deep learning, such as enhancement [14], denoising [15–17], and phase unwrapping [18–21]. Li et al. [22] proposed a two-frame reconstruction method based on the Phase U-Net (PUN) accurately estimating the wrapped phase from two frames of interferograms. PUN requires normalization of the interferograms, offering higher accuracy compared to other two-frame reconstruction methods. To reach the higher precision requirements of reconstructing the wrapped phase in two-frame random interferometry and further improving the reconstruction accuracy, we propose a new two-frame reconstruction method inspired by PUN [22] and Swin-Unet [12], which just require normalization of the interferograms in advance. Concretely, our contributions can be summarized as follows: (1) Swin-Unet has been constructed for wavefront reconstruction from two-frame phase-shifting interferograms, which only needs one random phase shift, reducing control costs and time requirements, as well as mitigating the impact of environmental factors; (2) PUN+, based on original PUN, has been proposed, which includes the development of a bilinear interpolation operation for up-sampling, eliminating the need for transpose convolution, also ReLU has been applied in final convolution layer without Softmax or ELU. Experimental results have confirmed the effectiveness of these changes; (3) Simulations and experiments indicate that both of the two proposed methods have superior performance than other traditional methods and deep learning method (PUN) in wavefront reconstruction.

2. Methods

2.1. The Process of Proposed Method

As shown in the blue part of Figure 1a, the proposed method reconstructs the wrapped phase from two randomly shifted interferograms. During network training, as shown in the green part, it involves solving the mean squared error (MSE) loss between the predicted results and the ground truth, and computing gradients for backpropagation. The model parameters are then adjusted and optimized through the adaptive moment estimation (ADAM) optimizer. During network testing, only the blue part is needed, and the green part is not used. Figure 1b illustrates the unwrapped phase being recovered from the wrapped phase by the unwrapping algorithm (“unwrap” function in the MATLAB R2018b [22]).

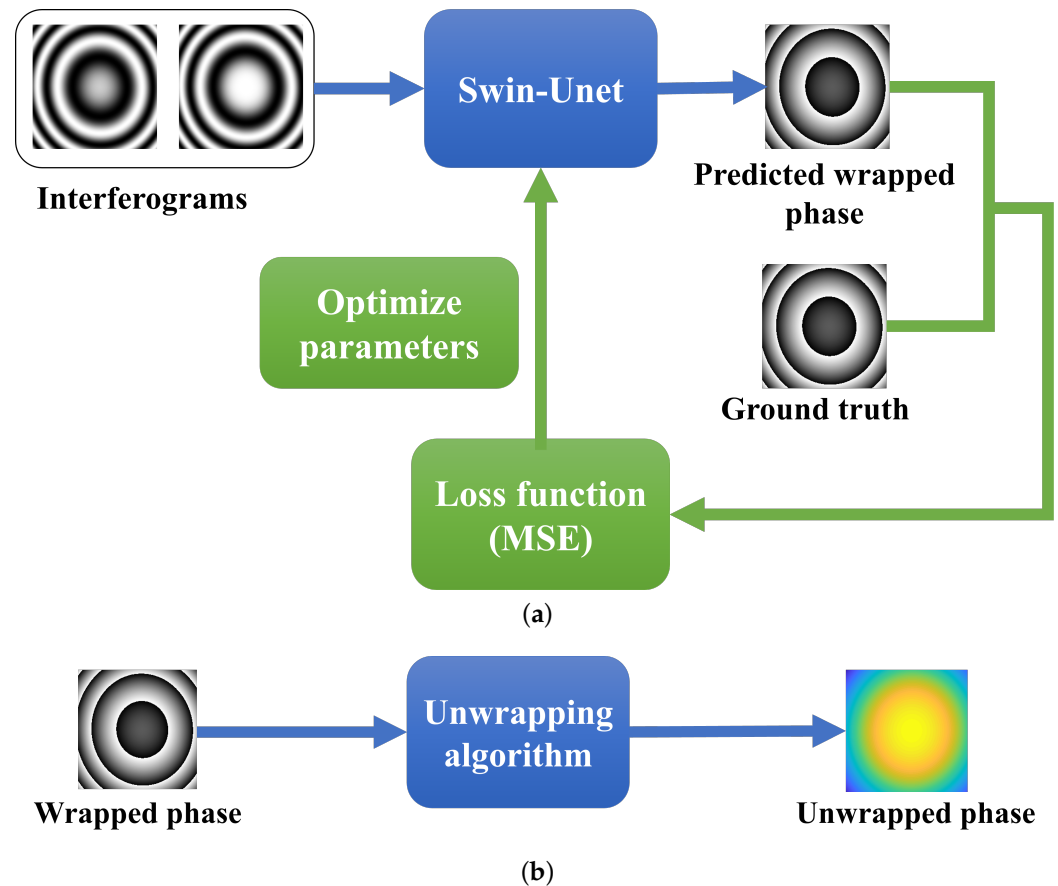


Figure 1. The workflow of two-frame PSI based on Swin-Unet; (a) the blue part represents wrapped phase prediction stage, whose input is two frame of interferograms with random step, the green part is Swin-Unet parameters update stage, where Swin-Unet weights are updated by computing MSE between the predicted wrapped phase and the ground truth; (b) the blue part is unwrapping stage, the unwrapping algorithm is used to recover the unwrapped phase.

2.2. Theoretical Background

Two-frame random phase-shifting interferometry obtains two-frame interferograms by altering the optical path difference between the reference wavefront and the testing wavefront [23]. The expressions for the intensity of the two obtained interference patterns are as follows:

$$I_1(x, y) = a(x, y) + b(x, y) \cos[\phi(x, y)] \tag{1}$$

$$I_2(x, y) = a(x, y) + b(x, y) \cos[\phi(x, y) + \delta] \tag{2}$$

where, $I_1(x, y)$ and $I_2(x, y)$ represent the intensity of the original and phase shifting interferograms, respectively, at the coordinate point (x, y) , $a(x, y)$, and $b(x, y)$ represent DC component and modulation component, respectively. $\phi(x, y)$ represents the phase of the testing wavefront at the coordinate point (x, y) , and δ represents the random phase step (except for 0 and π rad). Accurately generating the original phase $\phi(x, y)$ is crucial in simulations, and Zernike polynomials [24] play a key role in this process.

Zernike polynomials are a sequence of orthogonal and linearly independent polynomials defined on the unit circle. The orthogonality of the Zernike polynomials, enabling them to represent any square-integrable function within the unit disk, allows the coefficients of different polynomials to be independent of each other, which is advantageous in eliminating interference from accidental factors. In addition, Zernike polynomials and Seidel aberration coefficients can be easily correlated. Thus, any continuous arbitrary shape of the wavefront can be represented by a linear combination of Zernike polynomials [25], whose coefficients can be calculated using methods such as least squares fitting [26], Gram–Schmidt [27] and cubic B-spline [28]. The expression for the original phase $\phi(x, y)$ generated by Zernike polynomials is as follows:

$$\phi(\rho, \theta) = \sum_{r=0}^L Z_r U_r(\rho, \theta) \tag{3}$$

where L is the coefficient of the highest-order term, Z_r represents the Zernike coefficients for each term, and U_r represents the Zernike polynomials, whose expression is as follows:

$$U_n^{n-2m}(\rho, \theta) = R_n^{n-2m}(\rho) \begin{bmatrix} \sin \\ \cos \end{bmatrix} (n - 2m)\theta \tag{4}$$

where ρ represents the vector length between the coordinate point (x, y) and the origin point, and θ represents the angle between the vector and projection in the x -axis. When $(n - 2m) > 0$, $\sin[(n - 2m)\theta]$ is used, and when $(n - 2m) < 0$, the $\cos[(n - 2m)\theta]$ is used. Polynomial $R_n^{n-2m}(\rho)$ is as follows:

$$R_n^{n-2m}(\rho) = \sum_{s=0}^m (-1)^s \frac{(n - s)!}{s!(m - s)!(n - m - s)!} \rho^{n-2s}. \tag{5}$$

The expression for the continuous orthogonality of Zernike polynomials is as follows:

$$\int_0^1 \int_0^{2\pi} U_n^l(\rho, \theta) \cdot U_m^k(\rho, \theta) \rho d\theta d\rho = \begin{cases} 0 & (n = m, l = k) \\ \frac{\pi}{2(n+1)} \delta_{nm} \delta_{lk} & (n \neq m, l \neq k) \end{cases} \tag{6}$$

where $U_n^l(\rho, \theta)$ and $U_m^k(\rho, \theta)$ represent Zernike polynomials.

To visualize that Zernike polynomials can accurately represent continuous wavefront shapes, we measure and analyze the plane mirror by ZYGO GPI-XP/D4 laser interferometer and software Metro Pro[®] Version 8.3.5 [29] to obtain the real phase map. Figure 2a,b shows the real phase map and phase map generated by Zernike polynomials, respectively.

After the original phase is generated, the wrapped phase is computed as the corresponding ground truth for the network. The wrapped phase ϕ_w represents the phase angle of the original phase, which can be calculated by the MATLAB R2018b function "angle" [18], and then can be mapped directly to the target interval $[-\pi/2, \pi/2]$.

$$\phi_w = \text{angle}[\exp(j\phi)] \tag{7}$$

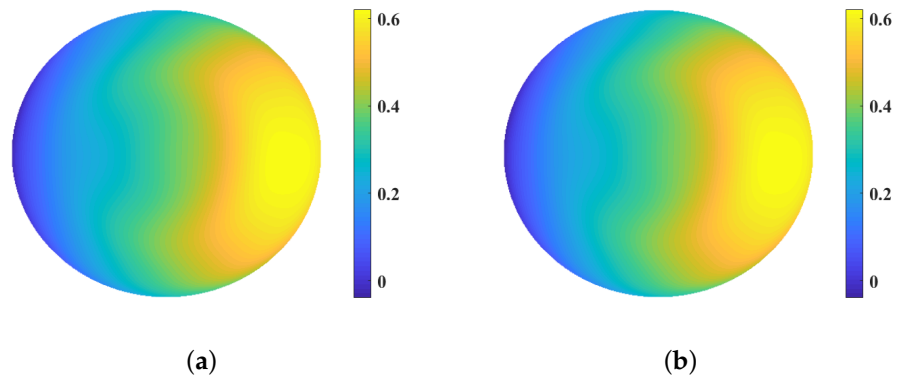


Figure 2. (a) Real phase map, (b) phase map generated by Zernike polynomials.

2.3. The Architecture of Neural Networks

2.3.1. PUN+

Based on the inspiration from PUN and the original U-Net, we have developed a PUN+ framework depicted in Figure 3, which comprises three main components: the left-side feature extraction network, the right-side feature fusion network, and the bridging network in the middle.

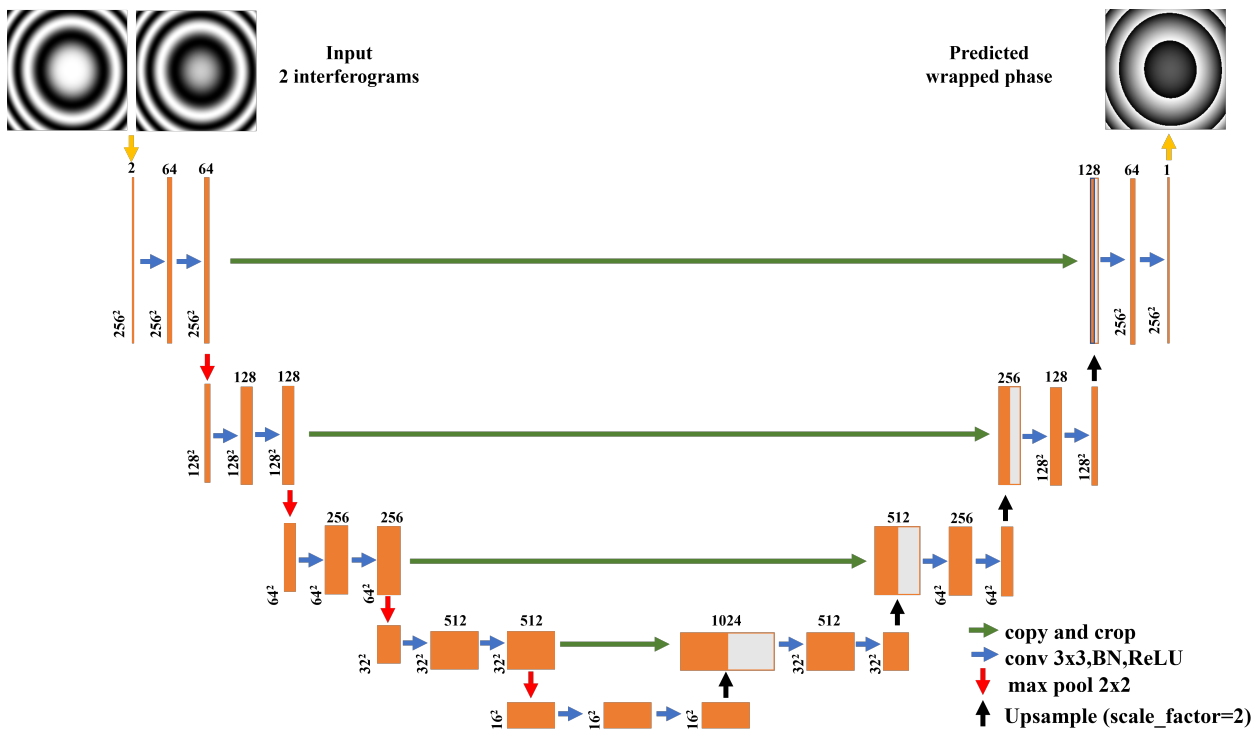


Figure 3. The architecture of PUN+, whose orange boxes correspond to multi-channel feature maps, gray-dotted boxes represent copied feature maps. The number of channels is on top of the box and the feature map size is denoted at the lower left edge of the box.

The left-side feature extraction network consists of four convolution blocks, each containing two sets of 3×3 convolutions, batch normalization, and rectified linear unit (ReLU) activation. Additionally, a downsampling layer using max pooling is included. Batch normalization accelerates network convergence and mitigates overfitting.

The right-side feature fusion network consists of four convolution blocks and an upsampling layer. The upsampled feature maps are combined with multi-scale feature maps from the left-side network through skip connections. The purpose of the feature

fusion is to compensate for the loss of spatial information during downsampling. Different from PUN and U-Net, we utilize the bilinear interpolation for upsampling. Estimating the pixel value at the target coordinates based on the surrounding pixel values and relative position enables image upscaling, and bilinear interpolation maintains the smoothness and details of the image to some extent which achieves a good balance between computational cost and scaling accuracy, compared to other interpolation algorithms.

The bridging network in the middle comprises a single convolution block that connects the left-side and right-side networks. Different from the original U-Net and PUN employing softmax and ELU activations in the output layer, respectively, PUN+ proposed employs ReLU activation to predict the wrapped phase. The output pixel values of the network image are expected to fall within the range $[0, 1]$, allowing for direct mapping to the target range $[-\pi/2, \pi/2]$. Since ReLU activation produces only positive values and has no negative output, we can ensure the network's output remains within the desired range by applying a suitable threshold.

For training, PUN+ employs the MSE loss function which measures the difference between the predicted phase and the ground truth.

2.3.2. Swin-Unet

The Swin-UNet architecture, shown in Figure 4, is a Transformer-based network inspired by U-Net, which consists of an encoder, bottleneck, decoder, and skip connections, and comprises 12 Swin Transformer blocks.

The encoder consists of a linear embedding layer, three sets of two consecutive Swin Transformer blocks, and a patch merging layer. The input image is divided into non-overlapping patches of size 4×4 , with each patch having a feature dimension of 32. The linear embedding layer projects the patch features to a specified dimension, generating patch tokens. The tokenized patches, with a resolution of $\frac{H}{4} \times \frac{W}{4}$, undergo feature representation learning by two consecutive Swin Transformer blocks. The Swin Transformer blocks maintain the feature dimension and resolution. Simultaneously, the patch merging layer downsamples the patches by a factor of $2\times$ and reduces the resolution to $\frac{H}{8} \times \frac{W}{8}$. The process is repeated three times.

The decoder comprises Swin Transformer blocks and a patch-expanding layer. The patch-expanding layer performs $2\times$ upsampling, expanding the feature map from a resolution of $\frac{H}{32} \times \frac{W}{32}$ to $\frac{H}{16} \times \frac{W}{16}$ while halving its dimension. Similar to U-Net, the output features from the patch expanding layer are fused with multi-scale features from the encoder through skip connections, mitigating spatial information loss caused by downsampling. The final patch expanding layer uses $4\times$ upsampling to restore the feature map resolution to the original ($W \times H$) while maintaining the same dimension. A linear projection layer is applied to generate pixel-level predictions.

The encoder, bottleneck, and decoder are interconnected, with the bottleneck consisting of two consecutive Swin Transformer blocks that preserve the feature map's dimension and resolution. Figure 4 illustrates the details of two consecutive Swin Transformer blocks. Each block includes a layer normalization (LN) layer, a window-based multi-head self-attention (W-MSA) model, a shifted window-based multi-head self-attention (SW-MSA) model, residual connections, and a multi-layer perceptron (MLP). W-MSA incorporates window partitioning into the conventional multi-head self-attention (MSA), while SW-MSA additionally incorporates window shifting operations. Using this window partitioning mechanism, two consecutive Swin Transformer blocks can be represented as follows:

$$\hat{z}^l = W\text{-MSA}\left(LN\left(z^{l-1}\right)\right) + z^{l-1}, \quad (8)$$

$$z^l = MLP\left(LN\left(\hat{z}^l\right)\right) + \hat{z}^l, \quad (9)$$

$$\hat{z}^{l+1} = SW\text{-MSA}\left(LN\left(z^l\right)\right) + z^l, \quad (10)$$

$$z^{l+1} = MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \quad (11)$$

where z^{l+1} and z^l are outputs of MLP, and \hat{z}^l and \hat{z}^{l+1} are outputs of W -MSA and SW -MSA, respectively.

Refer to reference [12], the expression for the self-attention mechanism is as follows:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (12)$$

where $Q, K, V \in \mathbb{R}^{M^2 \times d}$, Q, K, V represent query, key, and value matrices, respectively. M^2 represents the number of patches by splitting windows, and d represents the dimension of the query or key.

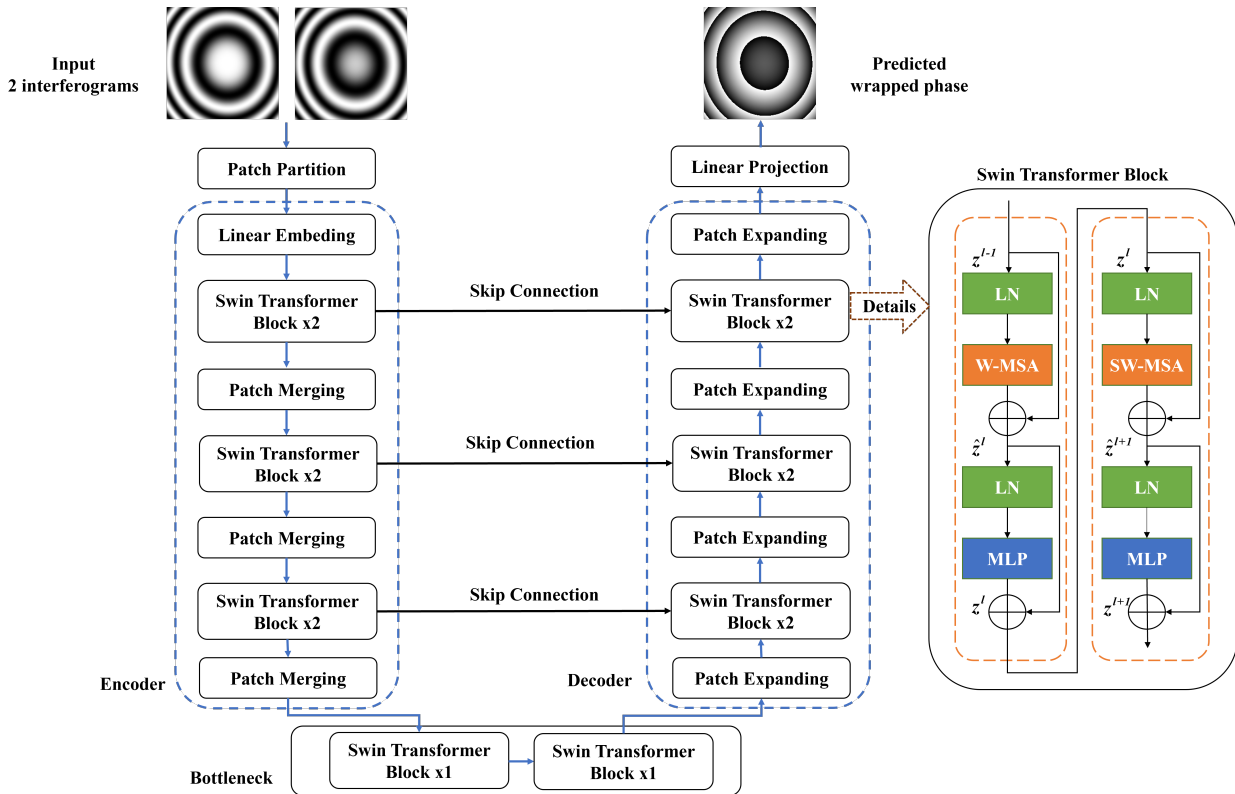


Figure 4. The architecture of Swin-Unet, which consists of four main components: encoder, bottleneck, decoder, and skip connections, constructed by Swin Transformer blocks.

2.4. Network Training

PUN+ and Swin-Unet are implemented using Python 3.9 and PyTorch 1.11.1. Training and testing are conducted on a PC with an NVIDIA GeForce RTX 3090 GPU, Xeon Platinum 8260C CPU. Both of the two training processes include 300 epochs on a dataset of 24,000 pairs of images, and the batch size of the dataloader is 32.

3. Results and Analysis

3.1. Simulation Dataset

Based on Section 2.2, we generate the simulated training data based on ZYGO’s Zernike polynomials, employing a linear combination of nine Zernike polynomials to generate the initial phase. To avoid generating excessively dense fringes, we choose Zernike coefficients from the 2nd to the 10th order (excluding Piston) and randomly assign amplitudes ranging from -9 to $+9$. This approach enables us to generate a variety of original phase patterns with different types of aberrations while maintaining manageable fringe densities in the interferograms.

To ensure that the trained network has a strong generalization ability and performs well in various scenarios, we utilize a large amount of simulated phase and its correspond-

ing interferograms generated from physical models for network training, respectively. By incorporating diverse variations in the training data, we aim to enhance the network's ability to handle different situations and improve its overall performance. Based on Equations (1), (2) and (7), we generate a total of 24,000 pairs of interferograms and their corresponding wrapped phase. These data are used as inputs and targets for the network after undergoing normalization preprocessing. The phase step for each pair of interferograms is a random value between 0 and π rad (excluding 0 and π rad). The dataset is divided into a training set (90%) and a testing set (10%). These interferograms are normalized between 0 and 1 to be the inputs of neural networks, while the predicted wrapped phase is the output. Figure 5a shows an example of an original phase generated by the Zernike polynomials, Figure 5b,c show the corresponding two-frame interferograms with random phase step; Figure 5d shows the corresponding wrapped phase.

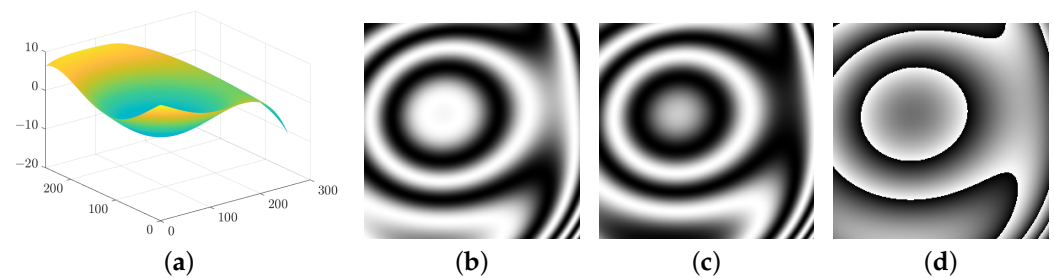


Figure 5. (a) An original phase generated by the Zernike polynomials method, (b,c) two-frame interferograms with random phase step generated by original phase, (d) wrapped phase generated by original phase.

To better simulate real-world conditions, random Gaussian white noise is added to the interferograms. Based on local means and local variances method (LMLVM) [30], we compute the signal-to-noise ratio (SNR) of testing interferograms with different noise levels. Figure 6 shows simulated interferograms with different noise levels. In addition, we set the DC component and modulation component of the interferograms to follow Gaussian distributions. Figure 7 shows simulated interferograms with different background intensities and modulations.

Our proposed method is trained on simulated data generated based on a physical model. Therefore, the predictive performance of the proposed method heavily relies on the content of the interferograms in the dataset, including but not limited to the settings of random phase shifts, background light intensity, modulation, and noise levels. To achieve better predictive results, it is essential to construct a large dataset for training. The training process poses a challenge in terms of computational resources and time, requiring the use of high-performance GPUs or TPUs, and it involves a lengthy training duration.

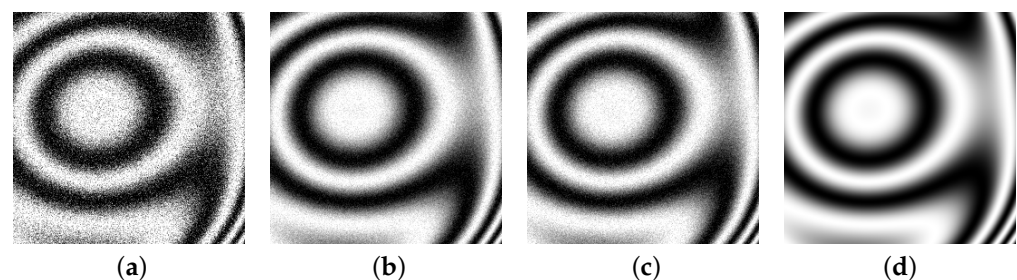


Figure 6. The interferograms with different noise levels, (a) 13.8 dB, (b) 23.9 dB, (c) 26.9 dB, (d) 38.9 dB.

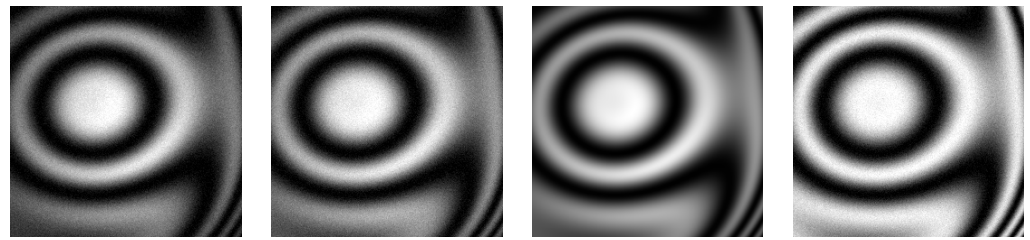


Figure 7. The interferograms with different background intensities and modulations.

3.2. Accuracy Test

To begin with, we conduct thorough validation to assess the feasibility and accuracy of the proposed method. The trained networks are subjected to test using the testing dataset. In Figure 8, the demodulation results of Kreis [7], OF [8], ST [9], GS [10], PUN [22], PUN+, and Swin-Unet are presented for a testing image with SNR = 43.5 dB and step = 1 rad. In the magnified detail images, it can be observed that the details reconstructed by PUN+ and Swin-Unet are closer to the ground truth. For further distinction and comparison, Table 1 provides the corresponding root mean square errors (RMSEs) comparing the reconstructed results with the ground truth. The RMSEs between Kreis, OF, ST, GS, and PUN, and the ground truths are 0.5255 rad, 0.5234 rad, 0.2534 rad, 0.2792 rad, and 0.1418 rad, respectively.

Table 1. Performance comparison of the reconstruction of different methods while step = 1 rad; flops and parameters of deep learning methods.

Methods	Flops (G)	Parameters (M)	Time (s)	13.9 dB	28.7 dB	43.5 dB
				RMSE (rad)	RMSE (rad)	RMSE (rad)
Kreis	—	—	0.0463	0.8843	0.5894	0.5255
OF	—	—	0.1655	0.7085	0.5464	0.5234
ST	—	—	0.1094	0.7652	0.3282	0.2534
GS	—	—	0.0079	0.7651	0.3498	0.2792
PUN	18.02	58.61	0.0117	0.2068	0.1813	0.1418
PUN+	40.15	17.26	0.0158	0.1840	0.1106	0.0921
Swin-Unet	7.72	27.14	0.0433	0.1647	0.1081	0.0719

Our proposed PUN+ and Swin-Unet whose reconstruction errors are 0.0921 rad and 0.0719 rad, respectively, outperform the other five existing reconstruction methods. While PUN is currently the most accurate method available, the reconstruction RMSEs of PUN+ and Swin-Unet are approximately 65% and 50% lower, than that of PUN, respectively.

Furthermore, we investigate the accuracy of the proposed method under phase steps ranging from 0 to π rad (excluding 0 and π rad). We compute the RMSEs between the wrapped phase obtained by seven dual-frame reconstruction methods and the ground truth, with interferograms that are devoid of additional noise. The results are plotted in Figure 9. Under noise-free conditions, the fluctuation range of PUN+ (brown line) and Swin-Unet (light blue line) does not exceed 0.055 rad and 0.045 rad, respectively, as the phase shift step changes from 0 to π rad, showing overall stable performance. Additionally, PUN+ and Swin-Unet consistently exhibit RMSEs below 0.1205 rad and 0.1124 rad, respectively, as the phase shift step changes from 0 to π rad. Compared to PUN (purple line), the proposed methods consistently outperform with significantly lower RMSEs for different phase shift steps. Moreover, they also demonstrate better performance near the singular phase π . Unlike OF (red line) and ST (blue line) which experience a rapid decline in reconstruction accuracy near the singular phase, resulting in a jump in RMSE, our proposed methods maintain better performance. These findings demonstrate that the trained Swin-Unet consistently outperformed other methods, which affirms the effectiveness and precision of the proposed method.

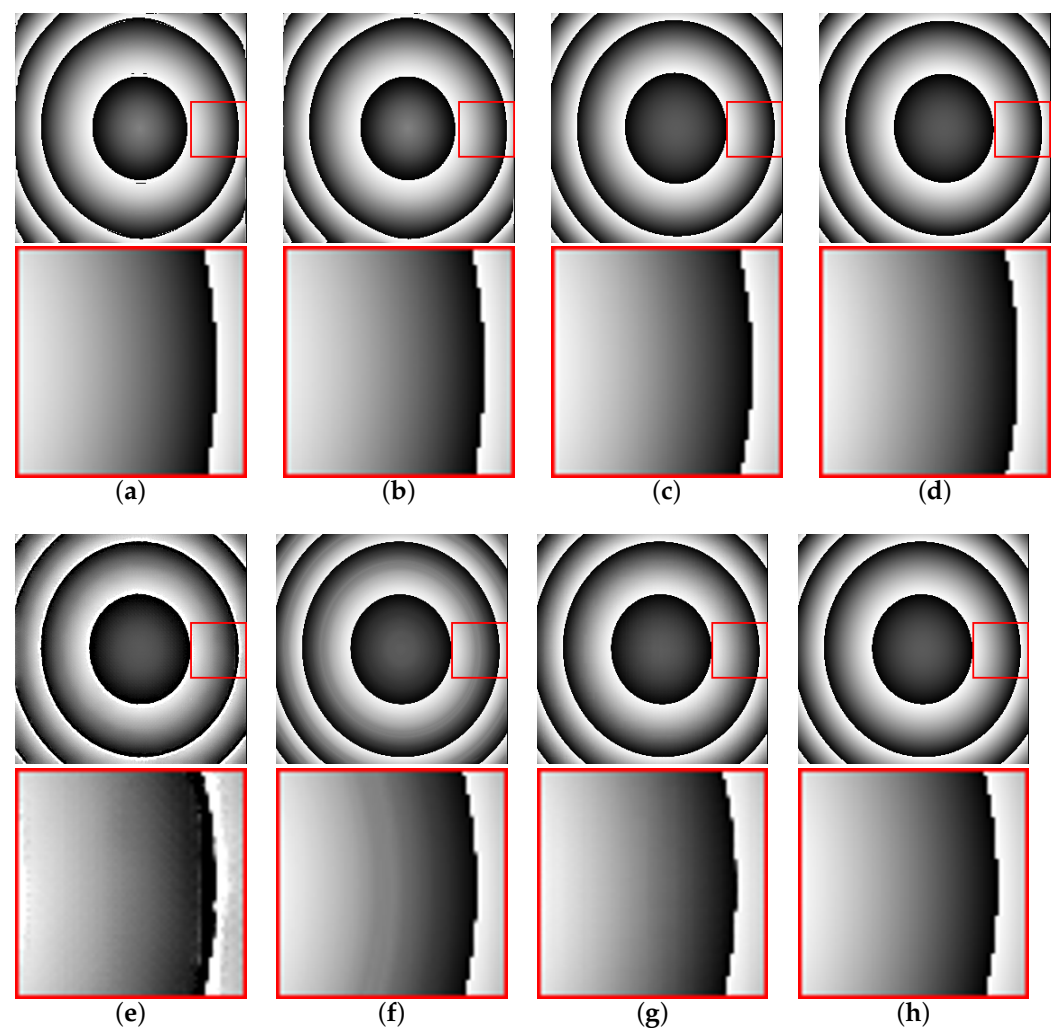


Figure 8. The wrapped phase obtained by different reconstruction methods on the simulation testing image with SNR = 43.5 dB, step = 1 rad, and 256×256 pixels, (a) Kreis, (b) OF, (c) ST, (d) GS, (e) PUN, (f) PUN+, (g) Swin-Unet, (h) ground truth.

Cautiously, when the phase step between the two-frame interferograms is greater than π radians, our proposed method can still accurately predict the wrapped phase, as the result of that the interferogram obtained by adding $(\pi + \Delta\delta)$ to the original phase is equivalent to adding $(\pi - \Delta\delta)$ to the original phase ($0 < \Delta\delta < \pi$). Moreover, based on Equations (1) and (2), it is evident that the intensity expression is a periodic function with a period of 2π . As a result, the proposed method can be applied to two-frame interferograms with phase shift of any random value greater than 0 (except integer multiples of π). To address scenarios with negative phase shift, predefined PZT movement directions are commonly used as a preventive approach.

3.3. Anti-Noise Performance

In real measurement environments, noise is present in interferograms and cannot be avoided, making it necessary to test the proposed method's noise resistance. Gaussian white noise is added to the randomly selected two-frame interferograms in the testing set, resulting in their SNR varying from 13.9 dB to 43.5 dB. For the convenience of analysis, the DC component and modulation component are both set to 1. Figure 10 shows testing two-frame interferograms with SNR = 13.9 dB. As shown in Figure 11, we perform wavefront reconstruction on the low SNR interferograms by Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet. It is clear that in high-noise conditions, PUN+ and Swin-Unet have better recon-

struction performance, which is not achievable by traditional existing methods. Figure 12 depicts the RMSEs between the wrapped phase obtained by different methods and the ground truth as SNR varying from 13.9 dB to 43.5 dB. Table 1 provides the RMSEs for Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet under 13.9 dB, 28.7 dB, and 43.5 dB SNR levels. Compared to PUN and PUN+, Swin-Unet has the acceptable network parameters [31] and the fewest flops, with 27.14 M and 7.72 G, respectively.

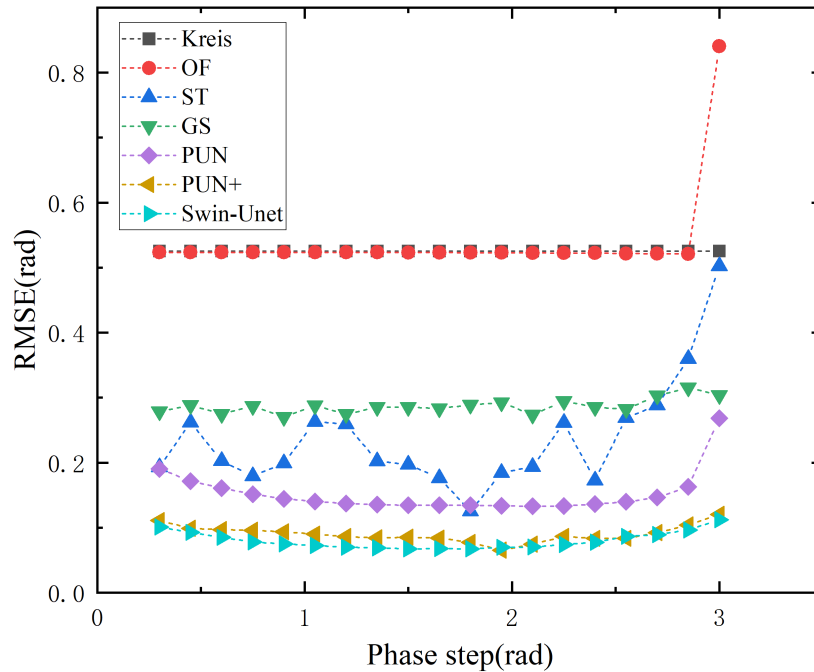


Figure 9. RMSEs of the Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet when phase step changes from 0 to π rad.

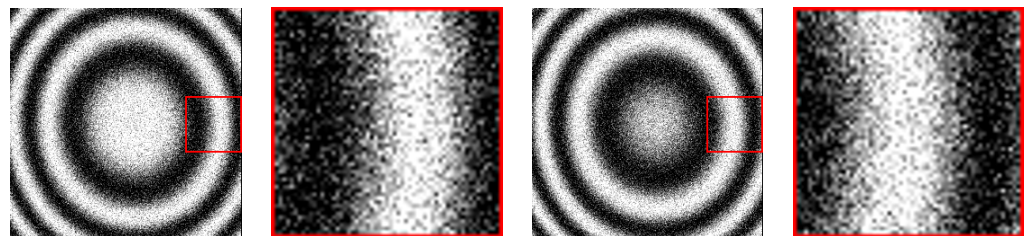


Figure 10. The testing two-frame interferograms with SNR = 13.9 dB and 256×256 pixels.

While SNR varies from 28.7 dB to 43.5 dB, the proposed PUN+ (brown line) and Swin-Unet (light blue line) consistently outperform other methods, with average RMSEs of 0.1031 rad and 0.09617 rad, respectively. Conversely, when the interferograms contain significant noise (13.9 dB), the reconstruction accuracy of nearly all methods drops. However, PUN+ (brown line) and Swin-Unet (light blue line) still exhibit RMSEs of 0.1840 rad and 0.1647 rad, respectively, which remain lower than other methods. This simulation confirms the superior performance of PUN+ and Swin-Unet across the entire tested SNR range. It should be noted that PUN+ corresponds to higher RMSEs than that of Swin-Unet, indicating that the noise resistance performance of PUN+ is slightly weaker than that of Swin-Unet. In practical measurements, to achieve optimal reconstruction accuracy, pre-denoising of interferograms is necessary.

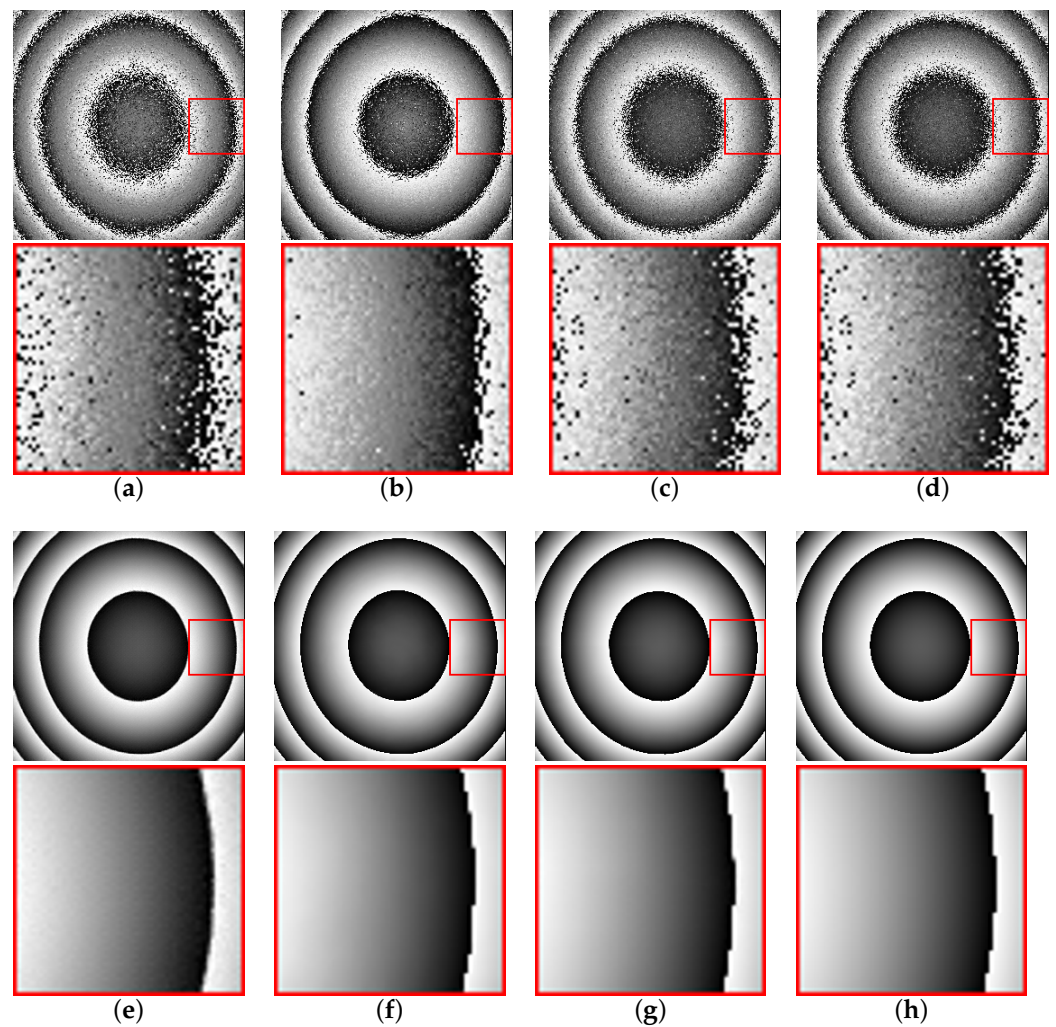


Figure 11. The wrapped phase obtained by different reconstruction methods on the testing image with low SNR (13.9 dB), (a) Kreis, (b) OF, (c) ST, (d) GS, (e) PUN, (f) PUN+, (g) Swin-Unet, (h) ground truth.

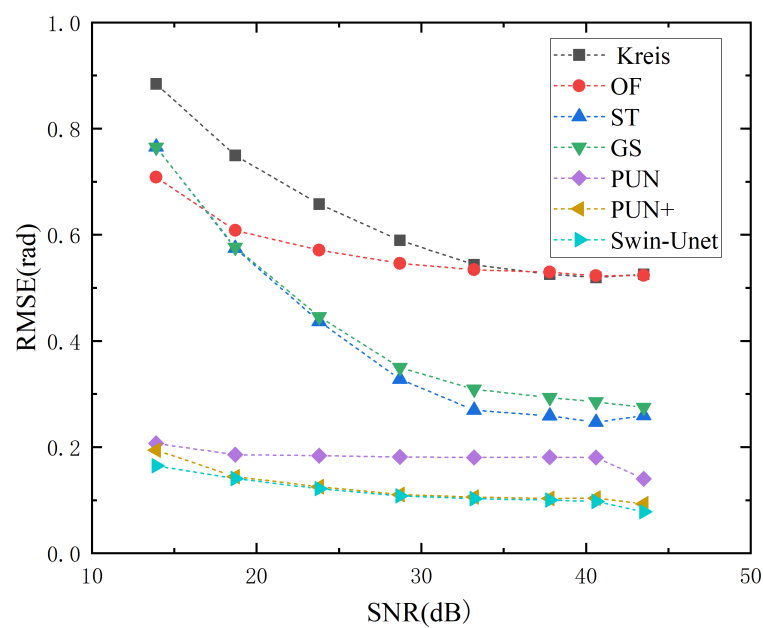


Figure 12. RMSEs of the Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet when SNR varies from 13.9 dB to 43.5 dB.

3.4. Low Modulation Test

In the actual measurement process, due to the influence of stray light and uneven illumination, interferograms often exhibit a visual effect of a bright central region and a dark surrounding area. For the two-frame phase-shifting interferometry, we assume that the background intensity and modulation of the two interferograms follow spatial inconsistency and temporal consistency. Spatial inconsistency means that the background intensity and modulation follow different Gaussian distributions, not constant in spatial distribution. Temporal consistency means that the PZT only needs to move once between the two interferograms, with a very short time duration, and the impact on the two interferograms is extremely similar. The background intensity and modulation are not functions of time.

As shown in Figure 13, we generated random phase-shifted dual-frame interferograms with different background intensities, modulations, and a moderate amount of Gaussian white noise. The residual maps between the reconstruction results obtained from different methods and the ground truth are illustrated in Figure 14. Table 2 provides the corresponding specific RMSEs. It can be clearly seen that under low modulation (29.7 dB), compared to other methods, the proposed PUN+ and Swin-Unet perform better, preserving fewer details lost, with RMSEs of 0.1329 rad and 0.1166 rad, respectively. As the modulation of the interferograms increases, the reconstruction errors of all methods show a decreasing trend. However, the reconstruction errors of PUN+ and Swin-Unet are consistently lower than other methods, with Swin-Unet's reconstruction accuracy slightly higher than PUN+.

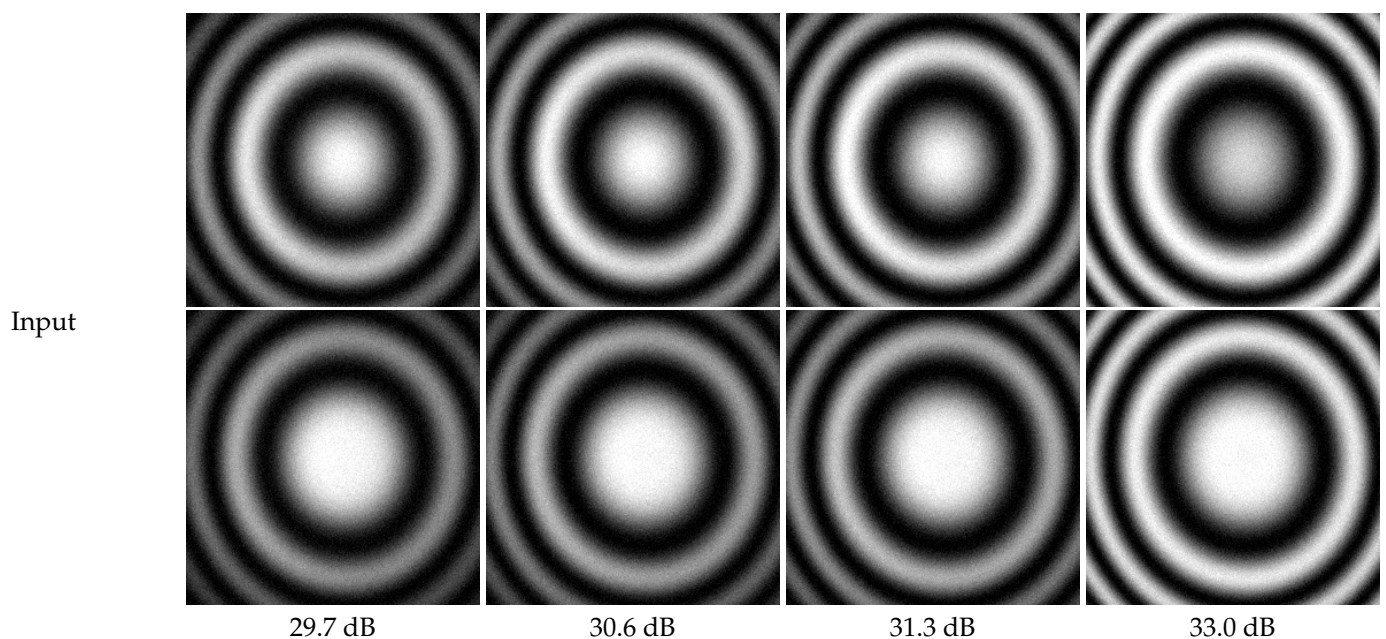


Figure 13. Two-frame interferograms with different background intensities, modulations, and noise levels. Background intensities, modulations, and noise levels of interferograms are identical in each column. The interferograms in different columns have different background intensities, modulations, and various noise levels (from left to right, SNR varies from 29.7 dB to 33.0 dB, and background intensity and modulation increase from low to high). Among them, the interferograms (29.7 dB) have the lowest background intensity and modulation, while the interferograms (33.0 dB) have the highest.

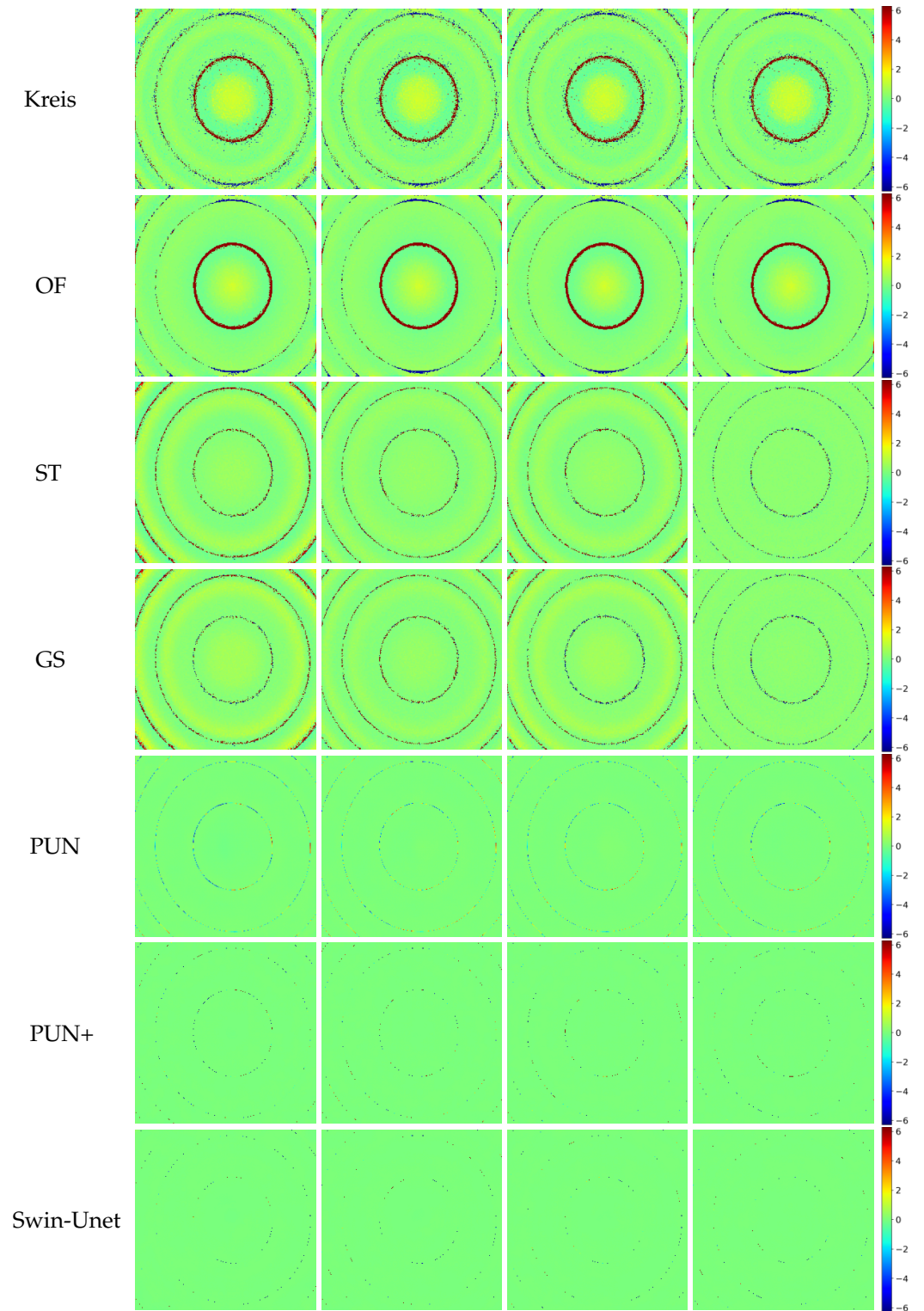


Figure 14. Residual maps between the predicted results reconstructed by different methods and the ground truth. The residual maps between the predicted results and the ground truth for Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet are sequentially presented from top to bottom.

Table 2. Performance comparison of the reconstruction of different methods under different background intensities and modulations.

Methods	29.7 dB	30.6 dB	31.3 dB	33.0 dB
	RMSE (rad)	RMSE (rad)	RMSE (rad)	RMSE (rad)
Kreis	0.6276	0.6077	0.6267	0.5847
OF	0.5668	0.5571	0.5706	0.5534
ST	0.5251	0.4141	0.4605	0.3310
GS	0.5010	0.4258	0.4396	0.3391
PUN	0.1636	0.1596	0.1567	0.1486
PUN+	0.1329	0.1249	0.1133	0.1110
Swin-Unet	0.1166	0.1074	0.1013	0.0940

3.5. Experiment

To further evaluate the effectiveness of the proposed methods, we conduct experiments by interferograms with random phase step in the experimental setup shown in Figure 15. The phase step between the two-frame interferograms is an unknown constant. Therefore, the linear phase shift error induced by the linear drift of the PZT has been effectively eliminated in the two-frame interferometry. Non-linear phase shift errors caused by vibration exist in the form of harmonics. To address this issue, we use a vibration isolation platform and a high-precision PZT to ensure that the phase shift between the two-frame interferograms is an unknown constant as much as possible, minimizing harmonic errors at this point. In actual measurement, by using ZYGO's GPI-XP/D4 laser interferometer, we capture multiple phase-shifting interferograms, Figures 16a and 17a show two real interferograms with random phase step in first set, and calculate the corresponding wrapped phase which served as the ground truth by the traditional thirteen-steps PSI [32], as shown in Figure 16i and Figure 17i, respectively. Then, in Figure 16, the wrapped phases are calculated from the first set of real interferograms using Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet methods, and both PUN and the proposed methods seem to perform well. In the locally magnified image, PUN is not smooth along the edges and contours, while PUN+ and Swin-Unet capture contour details that are closer to the ground truth. For ease of analysis, we compute RMSEs (compared to traditional PSI) for the seven reconstruction methods, which are 0.4235 rad, 0.4193 rad, 0.3505 rad, 0.3214 rad, 0.2650 rad, 0.25190 rad, and 0.2304 rad, respectively. The performance of the proposed PUN+ and Swin-Unet is better than PUN, reducing the RMSEs by 0.01 rad and 0.03 rad, respectively.

Figure 17 shows the wrapped phases are calculated from the second set of real interferograms using Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet methods. PUN, PUN+, and Swin-Unet perform better than others. In the locally enlarged image, the wrapped phase map reconstructed by PUN exhibits overall blurriness in the high-frequency signals, with a significant loss of fine details. In contrast, the wrapped phase maps reconstructed by PUN+ and Swin-Unet show clear and well-preserved fine details. Particularly, Swin-Unet's reconstruction results are remarkably close to the ground truth. After calculation and analysis, RMSEs between the wrapped phase obtained by the seven reconstruction algorithms (Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet) and the ground truth are 0.4383 rad, 0.4372 rad, 0.3788 rad, 0.3459 rad, 0.2804 rad, 0.2619 rad, and 0.2543 rad, respectively. Although PSI cannot be considered the actual ground truth, it is still reliable regardless of the noise present in the interferograms.

Furthermore, to verify the reconstruction accuracy of the demodulated results, we use the classic and simple "unwrap" function in MATLAB R2018b to unwrap the wrapped phase and obtain the unwrapped phase, as shown in Figures 18 and 19. Also, we compare the unwrapped results obtained from the seven reconstruction methods with the ground truth. In the first experimental data, RMSEs of Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet methods are 0.5413 rad, 0.5804 rad, 0.2801 rad, 0.2929 rad, 0.1510 rad, 0.0682 rad, and 0.0546 rad, respectively. In the second experimental data, RMSEs of Kreis, OF, ST, GS, PUN, PUN+, and Swin-Unet methods are 0.7562 rad, 0.7066 rad, 0.2025 rad, 0.1382 rad, 0.1176 rad,

0.0602 rad, and 0.0449 rad, respectively. The wrapped phase of PUN exhibits noticeable distortions in edge regions and high-frequency bands, leading to misalignment during unwrapping and further amplifying the RMSE. However, using PUN+ and Swin-UNET can effectively mitigate the distortion in high-frequency regions.

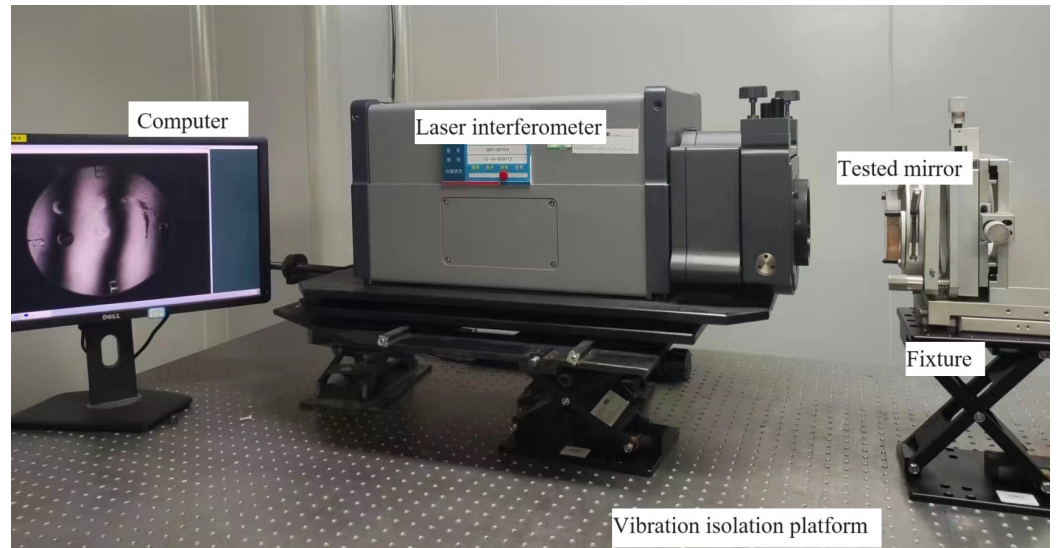


Figure 15. Environmental setup including computer, laser interferometer, tested mirror, fixture, and vibration isolation platform.

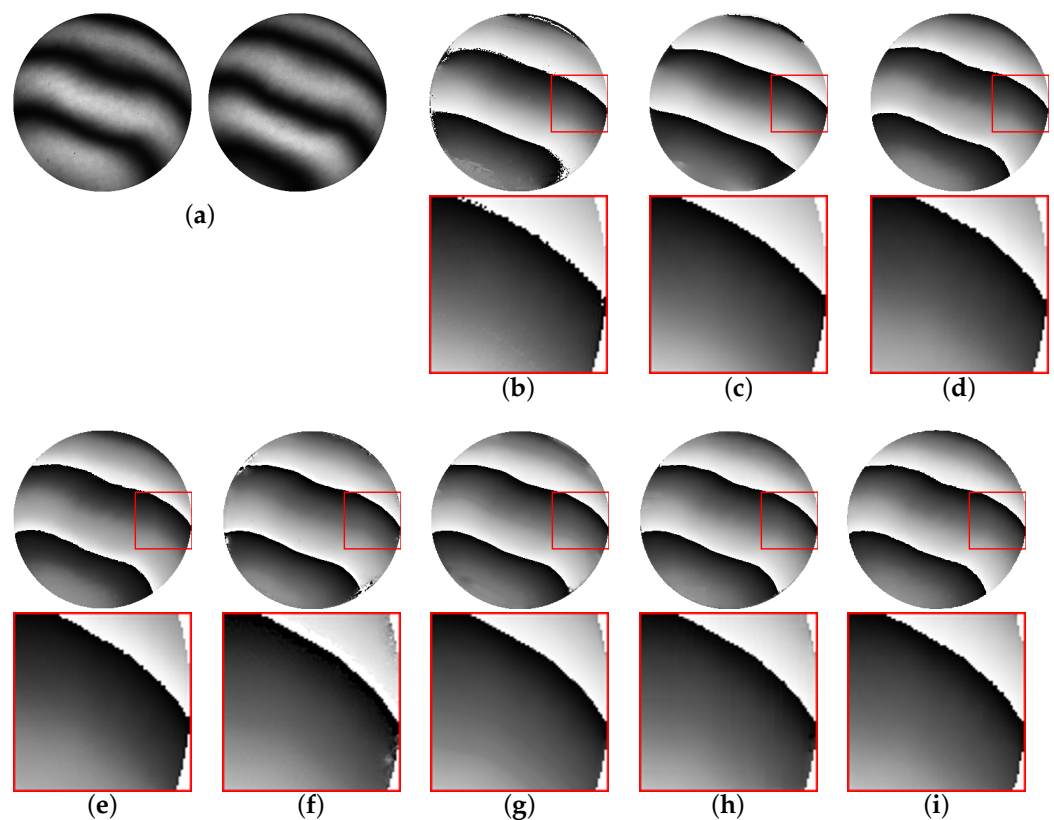


Figure 16. Evaluation of different two-frame methods with the first set of experimental data. (a) Real interferograms with random phase shift, the wrapped phase obtained by different reconstruction methods on the real interferograms, (b) Kreis, (c) OF, (d) ST, (e) GS, (f) PUN, (g) PUN+, (h) Swin-UNET, (i) ground truth.

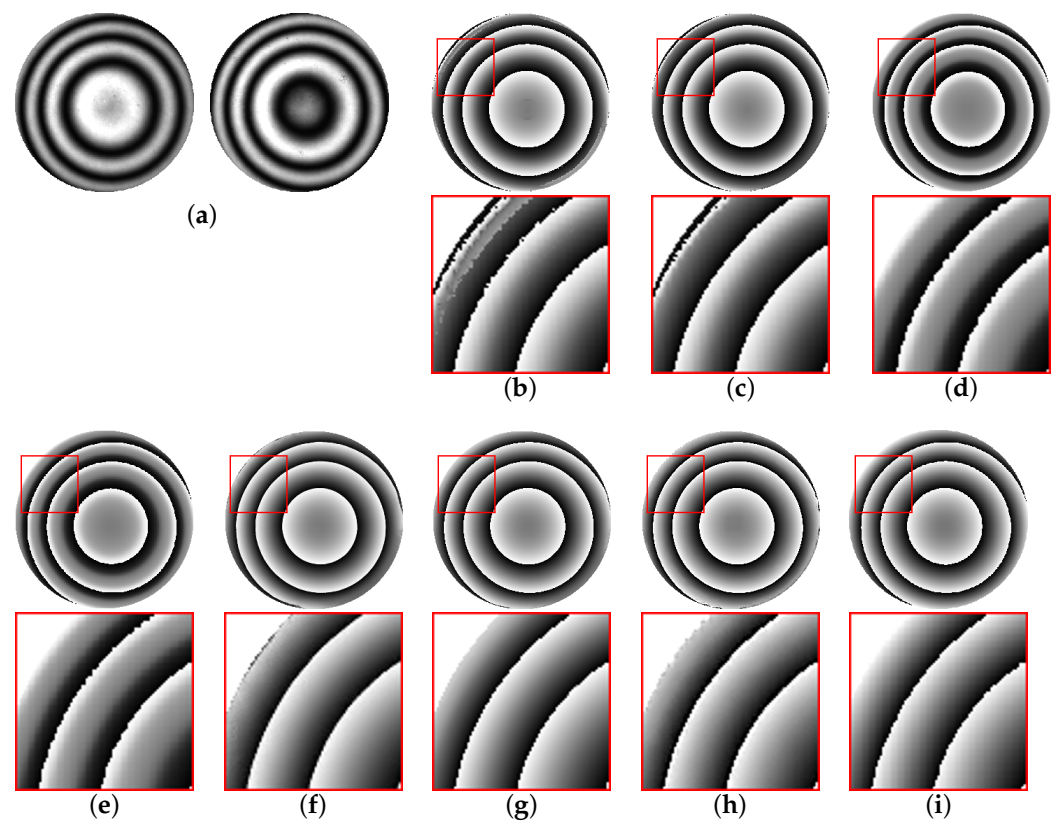


Figure 17. Evaluation of different two-frame methods with the second set of experimental data. (a) Real interferograms with random phase shift, the wrapped phase obtained by different reconstruction methods on the real interferograms, (b) Kreis, (c) OF, (d) ST, (e) GS, (f) PUN, (g) PUN+, (h) Swin-Unet, (i) ground truth.

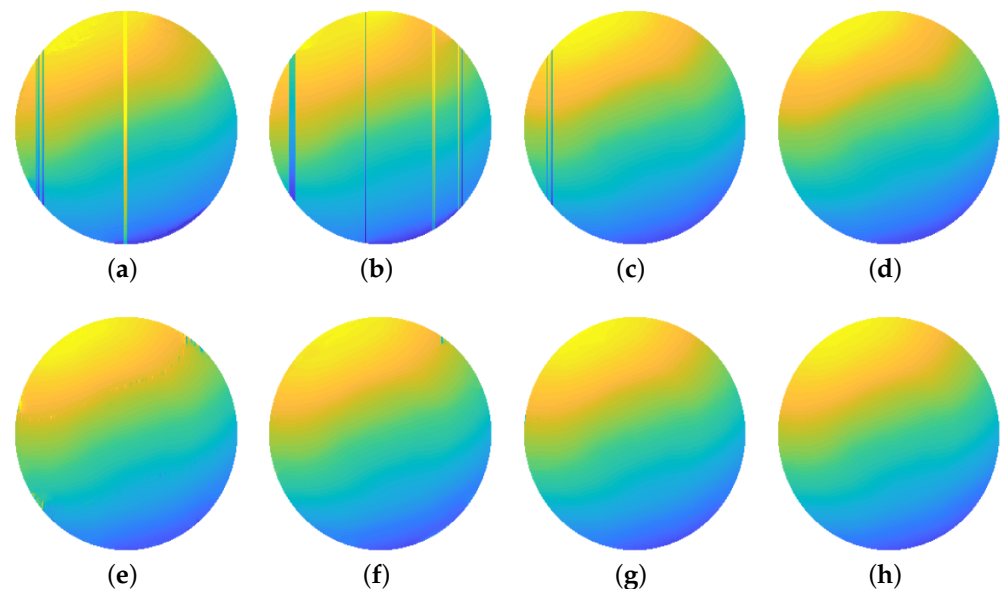


Figure 18. The reconstructed unwrapped phase obtained by different methods on the first set of real interferograms, (a) Kreis, (b) OF, (c) ST, (d) GS, (e) PUN, (f) PUN+, (g) Swin-Unet, (h) ground truth.

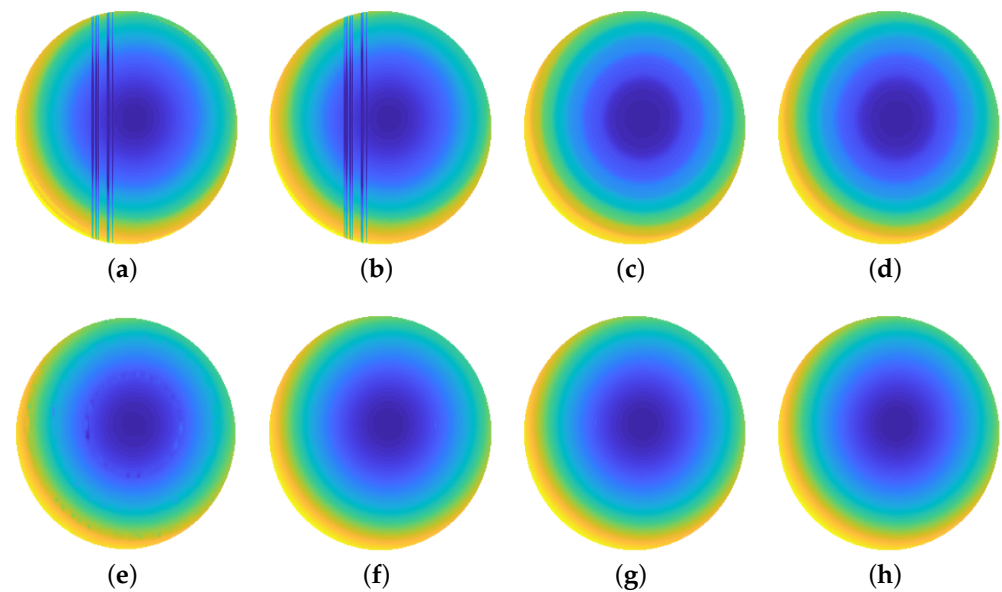


Figure 19. The reconstructed unwrapped phase obtained by different methods on the second set of real interferograms, (a) Kreis, (b) OF, (c) ST, (d) GS, (e) PUN, (f) PUN+, (g) Swin-Unet, (h) ground truth.

4. Conclusions

In conclusion, a novel approach is proposed for wavefront reconstruction based on Swin-Unet to accurately estimate the wrapped phase from two interferograms. Additionally, we have proposed PUN+, building upon the foundation of PUN [22], for wavefront reconstruction. By evaluating the methods on both simulated and real interferograms, and comparing their performance against the classical Kreis [7], OF [8], ST [9], GS [10], and PUN [22] methods. The accuracy of the proposed PUN+ and Swin-Unet have been verified through simulation and experimental results which demonstrate that our proposed methods compared to the above several methods exhibit superior performance in terms of demodulation results while being able to operate at a relatively acceptable time cost. Furthermore, by unwrapping the obtained wrapped phase in experiments, we further indicate that the original phase obtained from both methods still maintains higher accuracy. Overall, in the above processes, the comprehensive performance of the proposed Swin-Unet is superior to PUN+. The proposed Swin-Unet is a promising approach for wavefront reconstruction based on two-frame random interferometry.

Author Contributions: Conceptualization, Z.M. and B.L.; Methodology, X.S. and B.L.; Software, X.S.; Validation, X.S.; Formal analysis, X.S. and B.L.; Data curation, X.S.; Writing—original draft, X.S.; Writing—review and editing, B.L. and Z.M.; Supervision, B.L. and Z.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the Open Research Fund of State Key Laboratory of Transient Optics and Photonics (SKLST202114), and in part by the West Light Foundation of the Chinese Academy of Sciences (XAB2022YN08).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Malacara, D. *Optical Shop Testing*; John Wiley & Sons: Hoboken, NJ, USA, 2007; Volume 59.
2. Kinnstaetter, K.; Lohmann, A.W.; Schwider, J.; Streibl, N. Accuracy of phase shifting interferometry. *Appl. Opt.* **1988**, *27*, 5082–5089. [[CrossRef](#)] [[PubMed](#)]
3. Abdelsalam, D.; Yao, B.; Gao, P.; Min, J.; Guo, R. Single-shot parallel four-step phase shifting using on-axis Fizeau interferometry. *Appl. Opt.* **2012**, *51*, 4891–4895. [[CrossRef](#)] [[PubMed](#)]
4. Deng, J.; Wang, K.; Wu, D.; Lv, X.; Li, C.; Hao, J.; Qin, J.; Chen, W. Advanced principal component analysis method for phase reconstruction. *Opt. Express* **2015**, *23*, 12222–12231. [[CrossRef](#)] [[PubMed](#)]
5. Takeda, M.; Ina, H.; Kobayashi, S. Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.* **1982**, *72*, 156–160. [[CrossRef](#)]
6. Ge, Z.; Kobayashi, F.; Matsuda, S.; Takeda, M. Coordinate-transform technique for closed-fringe analysis by the Fourier-transform method. *Appl. Opt.* **2001**, *40*, 1649–1657. [[CrossRef](#)] [[PubMed](#)]
7. Kreis, T.M.; Jueptner, W.P. Fourier transform evaluation of interference patterns: Demodulation and sign ambiguity. In Proceedings of the Laser Interferometry IV: Computer-Aided Interferometry, San Diego, CA, USA, 1 January 1992; Volume 1553, pp. 263–273.
8. Vargas, J.; Quiroga, J.A.; Sorzano, C.; Estrada, J.; Carazo, J. Two-step interferometry by a regularized optical flow algorithm. *Opt. Lett.* **2011**, *36*, 3485–3487. [[CrossRef](#)] [[PubMed](#)]
9. Vargas, J.; Quiroga, J.A.; Belenguer, T.; Servín, M.; Estrada, J. Two-step self-tuning phase-shifting interferometry. *Opt. Express* **2011**, *19*, 638–648. [[CrossRef](#)]
10. Vargas, J.; Quiroga, J.A.; Sorzano, C.; Estrada, J.; Carazo, J. Two-step demodulation based on the Gram–Schmidt orthonormalization method. *Opt. Lett.* **2012**, *37*, 443–445. [[CrossRef](#)]
11. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
12. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
13. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 205–218.
14. Shi, J.; Zhu, X.; Wang, H.; Song, L.; Guo, Q. Label enhanced and patch based deep learning for phase retrieval from single frame fringe pattern in fringe projection 3D measurement. *Opt. Express* **2019**, *27*, 28929–28943. [[CrossRef](#)]
15. Yan, K.; Yu, Y.; Huang, C.; Sui, L.; Qian, K.; Asundi, A. Fringe pattern denoising based on deep learning. *Opt. Commun.* **2019**, *437*, 148–152. [[CrossRef](#)]
16. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [[CrossRef](#)]
17. Hao, F.; Tang, C.; Xu, M.; Lei, Z. Batch denoising of ESPI fringe patterns based on convolutional neural network. *Appl. Opt.* **2019**, *58*, 3338–3346. [[CrossRef](#)]
18. Wang, K.; Li, Y.; Kema, Q.; Di, J.; Zhao, J. One-step robust deep learning phase unwrapping. *Opt. Express* **2019**, *27*, 15100–15115. [[CrossRef](#)] [[PubMed](#)]
19. Spoorthi, G.; Gorthi, S.; Gorthi, R.K.S.S. PhaseNet: A deep convolutional neural network for two-dimensional phase unwrapping. *IEEE Signal Process. Lett.* **2018**, *26*, 54–58. [[CrossRef](#)]
20. Zhang, J.; Tian, X.; Shao, J.; Luo, H.; Liang, R. Phase unwrapping in optical metrology via denoised and convolutional segmentation networks. *Opt. Express* **2019**, *27*, 14903–14912. [[CrossRef](#)] [[PubMed](#)]
21. Yin, W.; Chen, Q.; Feng, S.; Tao, T.; Huang, L.; Trusiak, M.; Asundi, A.; Zuo, C. Temporal phase unwrapping using deep learning. *Sci. Rep.* **2019**, *9*, 20175. [[CrossRef](#)] [[PubMed](#)]
22. Li, Z.; Li, X.; Liang, R. Random two-frame interferometry based on deep learning. *Opt. Express* **2020**, *28*, 24747–24760. [[CrossRef](#)] [[PubMed](#)]
23. Cheng, Z.; Liu, D. Fast and accurate wavefront reconstruction in two-frame phase-shifting interferometry with unknown phase step. *Opt. Lett.* **2018**, *43*, 3033–3036. [[CrossRef](#)] [[PubMed](#)]
24. Zernike, F. Phase contrast. *Z. Tech. Physik.* **1935**, *16*, 454.
25. Malacara, Z.; Servín, M. *Interferogram Analysis for Optical Testing*; CRC Press: Boca Raton, FL, USA, 2018.
26. Wang, J.; Silva, D.E. Wave-front interpretation with Zernike polynomials. *Appl. Opt.* **1980**, *19*, 1510–1518. [[CrossRef](#)]
27. Cubalchini, R. Modal wave-front estimation from phase derivative measurements. *J. Opt. Soc. Am.* **1979**, *69*, 972–977. [[CrossRef](#)]
28. Ares, M.; Royo, S. Comparison of cubic B-spline and Zernike-fitting techniques in complex wavefront reconstruction. *Appl. Opt.* **2006**, *45*, 6954–6964. [[CrossRef](#)] [[PubMed](#)]
29. Kometer, R.; Hofbauer, E. Fast and reliable in-situ measurements of large and complex surfaces using a novel deflectometric device. In Proceedings of the Fifth European Seminar on Precision Optics Manufacturing, Teisnach, Germany, 10–11 April 2018; Volume 10829, pp. 70–76.

30. Van der Meer, F.D.; De Jong, S.M. *Imaging Spectrometry: Basic Principles and Prospective Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011; Volume 4.
31. Géron, A. Hands-on machine learning with Scikit-Learn. In *Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019; Volume 1.
32. Zhu, X.; Wu, Y.Q.; Liu, F. Noise suppression performance of typical phase shifting algorithms. In *Proceedings of the 8th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test, Measurement Technology, and Equipment*, Suzhou, China, 26–29 April 2016; Volume 9684, pp. 1008–1013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.