Noise Suppression and Unpacking Accuracy Improvement Based on Generative Adversarial Networks for Interferometric Phase Maps

Bohang Zhong[1], Huaian Yi[1], Fuqing Miao[2]
[1]College of Mechanical and Control Engineering, Guilin, Guangxi, China
[2]Key Laboratory of Advanced Manufacturing and Automation Technology, Guilin, Guangxi, China
Corresponding author: Fuqing Miao (e-mail: 15247602709@163.com).

ABSTRACT Multiplicative speckle noise and additive Gaussian noise in the interferometric phase map introduce spurious phase jumps and residual points, disrupting phase continuity and reducing the reliability of the phase unpacking algorithm. To address this issue, this paper proposes a conditional generative adversarial network. The generator uses a multi-scale dilated convolutional U-Net to output a complex phase residual map. The denoised result is obtained by subtracting the residual from the noisy phase. A dual discriminator structure in the spatial and frequency domains, combined with weighted structural similarity loss, phase gradient consistency loss, and least-squares adversarial loss, jointly constrains the network training. Training data is generated through electromagnetic scattering simulation. The proposed method achieves a 97.3% phase unpacking success rate at a signal-to-noise ratio of 0–3 dB. Processing a 2048 $\times$ 2048 pixel image takes less than 0.25 seconds. The proposed method effectively suppresses non-stationary noise and improves the accuracy of phase unpacking.

INDEX TERMS Interferometric Phase Image Denoising; Generative Adversarial Network; Phase Unpacking; Complex Residual Learning

I. INTRODUCTION

INTERFEROMETRIC synthetic aperture radar (InSAR) and optical interferometry techniques can achieve high-precision measurements of surface deformation, digital elevation models, and optical surface topography by extracting phase information from interferometric phase maps. However, the actual acquired interferometric phase maps are simultaneously contaminated by multiplicative speckle noise and additive Gaussian noise. Multiplicative speckle noise originates from random interference between scatterer echoes and is closely related to the coherence coefficient; additive Gaussian noise comes from electronic components such as system thermal noise and quantization errors. The superposition of these two types of noise introduces a large number of spurious phase jumps and residual points into the phase map, severely disrupting phase continuity and causing subsequent phase unpacking algorithms to accumulate errors or even fail completely. Therefore, effectively suppressing noise in the interferometric phase map before unpacking, while preserving as much fringe edge and detailed structure as possible, is a core problem that urgently needs to be solved in the field of interferometry.

In recent years, deep learning techniques, especially convolutional neural networks (CNNs) and generative adversarial networks (GANs), have opened up new avenues for phase map denoising. Most existing deep learning methods directly learn the end-to-end mapping from noisy phase maps to clean phase maps, but they often treat denoising as a simple image restoration problem, ignoring two key characteristics of interferometric phase maps: first, the phase values are periodic (wrapping characteristics), and direct regression of phase values easily produces discontinuous artifacts; second, the phase gradient (i.e., the direction and density of fringes) carries the main information of terrain or deformation, and the geometric consistency of the gradient field should be maintained during denoising. Furthermore, existing methods typically only focus on image quality in the spatial domain, lacking effective constraints on the residual noise in the frequency domain (especially in high-frequency regions).

To address the aforementioned shortcomings, this paper proposes a method for noise suppression and unpacking accuracy improvement of interferometric phase maps based on conditional generative adversarial networks (GANs). The network outputs a complex phase residual map (real and imaginary parts), and the denoised result is obtained by subtracting the residual from the noisy phase. The residual signal is sparsely distributed in comparison to the direct generation of a clean phase with the advantage of maintaining the backbone structure of the original phase and achieving effective generation blur reduction. The parallel dilated convolutional modules with dilated rate 1, 2 and 3 are added at the bottom of the encoder to expand the receptive field without downsampling, and to capture large-scale fringe trends and local details. After each downsampling layer, a phase gradient attention module is added to it, and the gradient magnitude map is used to calculate the weights of spatial attention, highlighting fringe boundaries and regions with significant changes in direction. The spatial domain discriminator uses PatchGAN to judge the correctness of the spatial structure of the denoised phase, and the frequency domain discriminator applies short-time Fourier transform to the phase map to find higher-frequency noise residue on the time-frequency map. When the two discriminators are trained simultaneously, the generator has to conform to the statistical properties of the real clean phase map in the frequency domain, yielding suppression of residual high-frequency noise. This paper introduces the combination of weighted structural similarity loss (weighting the structural similarity loss based on the gradient magnitude of multi-scale image patches), the phase gradient consistency loss (penalizing the difference between the denoised phase and the ground truth phase in the gradient direction), and least-squares adversarial loss to overcome the problem of artifacts along the stripe boundaries while retaining the texture details.

The paper is organized as follows: it introduces research work on interferometric phase map denoising and related research areas; it introduces the theory of the proposed method, network architecture, loss function, and construction of training data; it validates the effectiveness of the proposed method by multiple sets of comparative experiments, ablation experiments, and efficiency analysis; it summarizes and outlines future research directions for the whole paper.

II. RELATED WODR

The non-stationary noise and phase discontinuity in the interferometric phase map seriously reduce the reliability of the phase unpacking algorithm and the final measurement accuracy, which has become a key problem to be solved in the field of interferometry. Luo et al. [1] proposed a denoising method based on deep convolutional neural networks and constructed a deep network structure suitable for the characteristics of continuous light spatial heterodyne interferograms. The interferometric synthetic aperture radar (InSAR) interferogram is seriously affected by phase noise. Tao et al. [2] used multi-scale feature extraction and residual learning strategies to directly learn the nonlinear mapping from the noisy phase map to the clean phase map and realize adaptive filtering. Wang et al. [3] proposed a phase map denoising method based on the improved particle swarm algorithm (PSO), which transformed the phase denoising problem into an optimization problem. The improved PSO was used to adaptively search for the optimal filtering parameters (such as the filtering window size and weighting coefficients), and the search process was guided by the phase gradient information to realize adaptive denoising. Yadalam et al. [4] used the quantum interference principle to design a selective excitation scheme to isolate the target interference path and suppress the interference of the unwanted path by adjusting the phase matching condition; at the same time, the time-resolved detection technology was combined to distinguish the contribution of different paths. Gordienko et al. [5] designed an optical fiber parametric amplifier based on an interference structure. By precisely controlling the optical path difference and phase relationship between the pump and the signal, the unwanted four-wave mixing products are caused to undergo destructive interference and thus suppressed.

Fabry-Perot interferometers are widely used in fiber optic sensing, but traditional intensity demodulation is susceptible to light source fluctuations and transmission losses. Jie et al. [6] combined microwave photonics technology to map the interference phase information to the microwave frequency band and used microwave phase demodulation to replace direct optical phase detection, thereby improving anti-interference capability. Natale et al. [7] established a complete model that has been experimentally verified, including optical path interference effects, noise sources and electronic front-end characteristics. Florio et al. [8] used multi-baseline phase difference measurement combined with clustering algorithm and deblurring technology to separate and match interference phase data generated by different sources. Rasouli and Khoshkhatti [9] used spatial carrier frequency separation to generate two interferograms with an inherent phase shift of 90° at the same time, and extracted the phase information of the two images through spatial Fourier analysis to achieve phase demodulation under single exposure. Sánchez-Aguilar et al. [10] recorded two orthogonal interferograms with a fixed phase difference at the same time through polarization beam splitting, and directly calculated the phase using the intensity ratio of the two images without mechanical phase shift or wavelength scanning. Most existing deep denoising models directly regress the encapsulated phase value, ignoring the periodicity and complex representation characteristics of the phase itself. This leads to non-physical blurring or artifacts at the fringe boundaries in the denoising results. This paper introduces a conditional generative adversarial network (GAN) to replace direct phase mapping with complex residual learning, combines spatial and frequency domain discriminators, and designs a phase gradient consistency loss function. The aim is to achieve structure-preserving denoising and improve the unpacking accuracy of interferometric phase maps in high-noise environments.

III. METHOD

A. PROBLEM MODELING AND OVERALL FRAMEWORK

The interferometric phase map has multiplicative speckle noise and additive Gaussian noise. Multiplicative noise is caused by random interference from echoes of the scattering elements, and is related to the coherence coefficient, while additive noise is caused by thermal noise in the system. These overlapped noises result in the appearance of false phase jumps and/or residual points on the wrapped phase map, which compromise phase continuity and cause errors to accumulate or failure in subsequent unpacking algorithms. The proposed method treats the problem of noise suppression as a conditional generation problem between a noisy wrapped phase map and a clean phase map. The overall framework is shown in Fig. 1, using a conditional generative adversarial network.

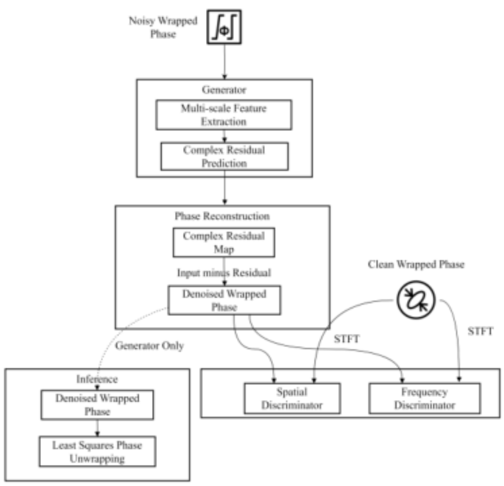

FIGURE 1. Denoising Framework

The generator accepts a noised wrapped phase map and returns a complex phase residual map that is the difference between noisy and clean phases in the complex domain. The output minus the residual is the denoised phase. During the training phase, the generator engages in adversarial training with two discriminators: a spatial domain discriminator takes the concatenation of the denoised phase map and the true clean phase map as input and outputs the true/false probability; a frequency domain discriminator performs a short-time Fourier transform on the phase map and identifies high-frequency noise residues on the time-frequency plot. In the inference stage, only the generator is used. After outputting the denoised phase map, it is input into the standard least squares unpacking algorithm. The reason for the generator outputting complex residuals is that the residual signal is sparsely distributed, which preserves the original phase backbone structure and reduces generation ambiguity [11].

The network is implemented in PyTorch. The optimizers for the generator and discriminator use Adam with an initial learning rate of 0.0002, $\beta_1 = 0.5$, $\beta_2 = 0.999$, and linear decay per epoch for a total of 200 epochs. The batch size is 4, and the input image size is cropped to 256×256 pixels.

B. GENERATOR

The generator uses a multi-scale dilated convolutional U-Net. The encoder contains four downsampling layers, each consisting of two convolutional blocks and a 2×2 max pooling operation. The first convolutional block uses 64 3×3 convolutional kernels with a stride of 1 and padding of 1, followed by batch normalization and LeakyReLU (negative slope 0.2). The second convolutional block doubles the number of channels. The output channels of each layer are 64, 128, 256, and 512, respectively. Three parallel dilated convolutional branches are placed at the bottom of the encoder, with dilation rates of 1, 2, and 4. Each branch has 256 3×3 convolutional kernels. The output is concatenated along the channels and then downsampled back to 256 channels by a 1×1 convolution. This module expands the receptive field and captures both large-scale stripe trends and local details. A phase gradient attention module is embedded after each downsampling layer: the gradient magnitude map G is obtained by calculating the horizontal and vertical differences of the input feature map F:

$$G = \sqrt{(\nabla_x F)^2 + (\nabla_y F)^2}. \tag{1}$$

∇_x and ∇_y are the difference operators in the horizontal and vertical directions, respectively. G is convolved by 3×3 (output channel 1, sigmoid) to generate a spatial attention weight map A:

$$A = \sigma(\mathrm{Conv}_{3\times3}(G)). \tag{2}$$

σ is the sigmoid function. Multiplying A and F channel by channel yields a weighted feature map, highlighting stripe boundaries and areas of drastic orientation changes. The decoder contains four upsampling layers. Each layer first uses a 2×2 transposed convolution to double its size and halve its channels, then concatenates it with the corresponding encoder layer output (attention-weighted) along the channel dimension, followed by two 3×3 convolutional blocks. The final layer outputs 64 channels, which are reduced to 2 channels by a 3×3 convolution, corresponding to the real and imaginary parts of the complex residual. The generator has approximately 7.8 million parameters, and all convolutions use reflection padding. The input noisy phase map is mapped to two channels (real and imaginary parts) via sin/cos mapping. After obtaining the residual from the generator, it is subtracted from the input and then passed through atan2 to obtain the denoised wrapped phase map:

$$\phi_{\text{denoised}} = \mathrm{atan2}\left(\sin\phi_{\text{noisy}} - \hat{r}_{\text{imag}}, \cos\phi_{\text{noisy}} - \hat{r}_{\text{real}}\right). \tag{3}$$

ϕ_{denoised} is the denoised wrapped phase map; ϕ_{noisy} is the noisy wrapped phase map; $\hat{r}_{\text{real}}$ and $\hat{r}_{\text{imag}}$ are the real and imaginary parts of the complex residual output by the generator, respectively.

C. DISCRIMINATOR

The dual discriminator architecture includes spatial and frequency domain discriminators, both based on PatchGAN. Both output $N \times N$ feature maps, with each point corresponding to the authenticity of a local region, and the final average is taken. The spatial domain discriminator has a 6-channel input: the generator output denoised phase map (2 channels after sin/cos) is concatenated with the true clean phase map (2 channels) along the channel dimension, and an additional noisy input phase map (2 channels) is introduced as a condition. The discriminator has 5 convolutional layers: the first layer has 64 4×4 convolutional kernels, stride 2, padding 1, LeakyReLU; the second layer has 128 4×4 kernels, stride 2, padding 1, batch normalization, LeakyReLU; the third layer has 256 4×4 kernels, stride 2, padding 1, batch normalization, LeakyReLU; the fourth layer has 512 4×4 kernels, stride 1, padding 1, batch normalization, LeakyReLU; the fifth layer has 1 4×4 kernel, stride 1, padding 1, and outputs the discriminant map. All convolutional reflections are filled. The frequency domain discriminator first performs short-time Fourier transform on the denoised phase map and the real clean phase map (single-channel wrapped phase value). The STFT uses a Hanning window with a window length of 32, an overlap stride of 16, and 64 Fourier points, and takes the amplitude. For the 256×256 phase map, the time-frequency map after STFT has a spatial dimension of 16×16 and a frequency dimension of 33 [12]. STFT transformation is:

$$\mathrm{STFT}\{x\}(m, \omega) = \sum_{n=0}^{N-1} x[n]h[n-m]e^{-j\omega n}. \tag{4}$$

x is the input signal; n is the discrete-time index; N is the window function length; h is the Hanning window function; m is the time frame index; ω is the angular frequency; j is the imaginary unit. The denoised spectrum and the true spectrum are concatenated in the channel dimension (2 channels) and input into the frequency domain discriminator. The discriminator's convolution kernel is changed to 3×3, with a stride of 2 and 4 layers, resulting in a smaller output discriminant map. The weighted average of the outputs of the two discriminators (0.6 for the spatial domain and 0.4 for the frequency domain) yields the final adversarial loss:

$$D_{\mathrm{final}}(x) = 0.6D_{\mathrm{spatial}}(x) + 0.4D_{\mathrm{freq}}(x). \tag{5}$$

D_{spatial} is the output of the spatial domain discriminator; D_{freq} is the output of the frequency domain discriminator. During training, the generator is updated twice for every update of the discriminator to ensure that the generator obtains sufficient gradients.

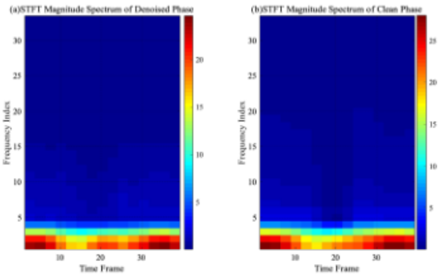

FIGURE 2. Time-Frequency Amplitude Spectrum

Fig. 2 shows the phase map obtained after denoising the same noisy wrapped phase map using the proposed method, and the corresponding real clean phase map, respectively, after short-time Fourier transform and time-frequency amplitude spectra. The left side is the time spectrum of the denoised phase map, and the right side is the time spectrum of the real clean phase map. In the time spectrum of the denoised phase map, the energy distribution in the high-frequency region (frequency index greater than 20) is highly consistent with the real spectrum, and no abnormal energy peaks caused by residual speckle noise or artifacts appear. The amplitude fluctuations along the time frame direction in the time-frequency spectrum are smooth, indicating that the denoising process does not introduce time domain discontinuities. The contours of the two time spectra in the low-frequency principal components (frequency indices 5 to 15) are similar, verifying the generator's ability to preserve the main interference fringe structure. The frequency domain discriminator makes a true/false judgment based on such spectral feature differences, forcing the generator output to approximate the statistical characteristics of the real clean phase map in the frequency domain [13], [14].

D. LOSS FUNCTION

The generator's total loss consists of three components: weighted structural similarity loss, phase gradient consistency loss, and adversarial loss, all with fixed weights. The weighted structural similarity loss acts between the denoised phase image and the true clean phase image, replacing the L1/L2 loss to preserve texture details. First, the phase image is converted to a two-channel format using sin/cosine, and the structural similarity index is calculated on each channel and averaged. A spatial weight map based on the phase gradient magnitude is introduced: regions with larger gradient magnitudes (stripe edges) are assigned higher weights, with the weight coefficient being the normalized gradient magnitude plus 1. This loss is calculated on five image patches of different scales (patch sizes from 11×11 to 27×27, with the stride halved), and the final value is obtained after weighted averaging. Weighted Structural Similarity Loss:

$$\mathcal{L}_{\text{WSSIM}} = \sum_{k=1}^{5} \lambda_k \cdot E_{p_k} \left[w(p_k) \cdot (1 - \text{SSIM}(\phi_{\text{denoised}}(p_k), \phi_{\text{clean}}(p_k))) \right]. \tag{6}$$

$\mathcal{L}_{\text{WSSIM}}$ is the weighted structural similarity loss; k is the scale index; λ_k is the weight coefficient of the k-th scale; p_k is the image patch at the k-th scale; $w(p_k)$ is the spatial weight map within the patch based on the phase gradient magnitude; ϕ_{clean} is the true clean wrapping phase map.

Phase gradient consistency loss suppresses artifacts across stripe boundaries. The difference between the denoised phase map and the true clean phase map in the horizontal and vertical directions is calculated to obtain the gradient field. Then, the cosine similarity between the two gradient fields (the cosine of the angle between the gradient vectors at corresponding positions) is calculated. The loss is defined as 1 minus the average value of this cosine similarity across the entire image. This loss penalizes directional deviation rather than magnitude deviation, because directional deviation almost entirely comes from artifacts. The phase gradient consistency loss is:

$$\mathcal{L}_{\text{PGC}} = 1 - \frac{1}{|\Omega|} \sum_{(p,q)\in\Omega} \frac{\Delta_x \phi_{\text{denoised}} \Delta_x \phi_{\text{clean}} + \Delta_y \phi_{\text{denoised}} \Delta_y \phi_{\text{clean}}}{\sqrt{(\Delta_x \phi_{\text{denoised}})^2 + (\Delta_y \phi_{\text{denoised}})^2} \sqrt{(\Delta_x \phi_{\text{clean}})^2 + (\Delta_y \phi_{\text{clean}})^2}}. \tag{7}$$

$\mathcal{L}_{\text{PGC}}$ represents the phase gradient consistency loss; Ω represents the pixel spatial domain; p, q represent pixel coordinates; Δ_x and Δ_y represent the gradient operators in the horizontal and vertical directions, respectively. The adversarial loss adopts the least squares GAN form to provide a smoother gradient. The outputs of the spatial domain discriminator and the frequency domain discriminator are used to calculate the least squares loss: the generator wants the discriminator output to be close to 1, the discriminator wants the output of the real image to be close to 1, and the output of the generated image to be close to 0. The adversarial loss is the sum of the adversarial losses of the two discriminators. The least squares generator adversarial loss is:

$$\mathcal{L}_{\text{Adv}} = \frac{1}{2} E\left[(D_{\text{spatial}}(\phi_{\text{denoised}}) - 1)^2 \right] + \frac{1}{2} E\left[(D_{\text{freq}}(\phi_{\text{denoised}}) - 1)^2 \right]. \tag{8}$$

The weights of the three losses are: weighted structural similarity loss 1.0, phase gradient consistency loss 0.5, and adversarial loss 0.05. These weights are determined by parameter tuning: a large adversarial loss weight leads to unrealistic textures, while a small weight results in insufficient guidance from the discriminator; a phase gradient consistency loss of 0.5 is sufficient to suppress artifacts without interfering with structural preservation. The generator's total loss is:

$$\mathcal{L}_{\text{total}} = 1.0 \cdot \mathcal{L}_{\text{WSSIM}} + 0.5 \cdot \mathcal{L}_{\text{PGC}} + 0.05 \cdot \mathcal{L}_{\text{Adv}}. \tag{9}$$

$\mathcal{L}_{\text{total}}$ is the generator's total loss.

E. TRAINING DATASET CONSTRUCTION

The training dataset was generated through electromagnetic scattering simulation to avoid the problem of real InSAR data lacking a true clean phase map. The simulation was based on four terrain units: slope, ridge, depression, and step. The terrain height map size was 512×512 pixels with a resolution of 1.5 meters. A backscattering model was used to generate complex scattering coefficients: the amplitude followed a Rayleigh distribution, the phase followed a uniform distribution, and the spatial correlation between adjacent pixels was introduced by the terrain slope. The interferometric phase was generated through two imaging geometry steps: the antenna position of the main image was fixed, and the vertical baseline of the auxiliary image was randomly set from 50 to 200 meters with an incident angle of 30 to 50°. The true interferometric phase of each point was calculated and wrapped to obtain a clean label. Noise was added using a composite model simulating multiplicative speckle noise and additive Gaussian noise.

Multiplicative noise: the complex scattering coefficient of each pixel was multiplied by a complex circular Gaussian noise term, the variance of which was related to the coherence coefficient, which was randomly set from 0.3 to 0.9. Additive noise: Complex Gaussian noise with zero mean and variable variance is added to the complex image domain to achieve a uniform SNR distribution between 0 and 15 dB. Each sample simultaneously records the clean wrapped phase, the noisy wrapped phase, and the true noise residual. The dataset is divided into training/validation/test sets in an 8:1:1 ratio, with the test set having two subsets: SNR=3 dB and 1 dB. All phase images are stored as 16-bit floating-point NumPy arrays, normalized to the range $[-\pi, \pi]$. Training data augmentation: random horizontal flipping, vertical flipping, 90° rotation, and cropping to 256×256 from the center of a 512×512 dataset. Noise parameters are not augmented on the validation set. Noise parameters are resampled randomly from the range for each epoch to improve generalization ability. The multiplicative noise coherence coefficient and additive noise variance are calculated for different SNR intervals. SNR is uniformly divided into 0–15 dB increments, with 200 samples per increment. The mean and standard deviation of the coherence coefficient and noise variance are calculated. The results are shown in Table 1.

TABLE 1. Statistics of Noise Parameters in the Training Dataset

SNR range (dB)	Number of samples	Mean coherence coefficient	Mean additive noise variance (rad^2)
0–3	400	0.34	0.182
3–6	400	0.52	0.095
6–9	400	0.68	0.041
9–12	400	0.79	0.018
12–15	400	0.88	0.007

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The comparison method used classical filtering and advanced denoising algorithms: Goldstein filtering used a 32×32 sliding window with filtering parameter $\alpha = 0.7$

and adaptive adjustment of the frequency domain mask intensity; InSAR-BM3D was based on 3D block matching collaborative filtering with a block size of 8×8, a search window of 15×15, a hard threshold $\lambda = 0.3$, and a Wiener filtering threshold of 0.5. Both transformations used two-dimensional wavelets plus one-dimensional wavelets. In the evaluation metrics, the residual point reduction rate was determined by detecting and counting 2π jump points using four-neighbor difference; the phase gradient error (PGE) was defined as the root mean square of the gradient difference between the denoised phase and the true wrapped phase in the horizontal/vertical direction, in radians; the root mean square error of unpacking (RMSE) was calculated using the pixel-by-pixel error between the unpacked and true absolute phases using least squares; the unpacking success rate was the percentage of pixels with RMSE $< \pi/2$. The random seed is fixed at 42 in both Python and MATLAB. The input phase is normalized to the $[-\pi, \pi]$ interval and stored as a float32 array.

B. NOISE SUPPRESSION PERFORMANCE EVALUATION

Goldstein filtering (window size 32×32, $\alpha = 0.7$, frequency domain mask adaptive), InSAR BM3D (block size 8×8, search window 15×15, hard threshold $\lambda = 0.3$, Wiener filter threshold 0.5), and the proposed method (using only the trained generator, excluding the discriminator) are applied to the noisy wrapped phase map of each sample. The phase gradient error (PGE) is calculated according to equation (7), taking the average of 200 samples within each interval. The obtained data are shown in Table 2.

TABLE 2. Phase Gradient Error of Each Method Under Different SNR Intervals

SNR range (dB)	Goldstein PGE (rad)	InSAR-BM3D PGE (rad)	Proposed PGE (rad)
0–3	0.184	0.093	0.052
3–6	0.119	0.065	0.031
6–9	0.077	0.042	0.019
9–12	0.049	0.026	0.012
12–15	0.032	0.017	0.008

The PGE of all three methods decreased with increasing SNR, but the proposed method achieved the lowest value across all ranges. Under the worst noise conditions of 0–3 dB, the PGE of the proposed method was 0.052 rad, which was 71.7% and 44.1% lower than that of Goldstein filtering (0.184 rad) and InSAR-BM3D (0.093 rad), respectively. As the SNR increased to 12–15 dB, the PGE of the proposed method decreased to 0.008 rad, approaching the level of a true clean phase gradient, and the rate of decrease was significantly faster than that of the comparative methods, indicating that it can effectively maintain the phase gradient structure even in high-noise environments. To further quantify the ability of the proposed method to suppress spurious phase jumps, four-neighbor differential detection of 2π jump points was used to calculate the number of residual points per 10^4 pixels, and the residual point reduction rate compared to the two comparative methods was obtained (reduction rate = (number of residual points in comparative methods − number of residual points in proposed method) / number of residual points in comparative methods $\times 100\%$). The average statistical results of 200 samples in each SNR interval are shown in Table 3.

TABLE 3. Number of Residual Points and Reduction Rate of the Proposed Method in Different SNR Intervals

SNR range (dB)	Proposed residual points ($\times 10^{-4}$)	Reduction vs. Goldstein (%)	Reduction vs. InSAR-BM3D (%)
0–3	132.1	86.2	71.5
3–6	79.5	85.1	69.8
6–9	43.8	84.3	68.2
9–12	22.4	83.9	67.5
12–15	10.9	82.7	66.1

The proposed method achieves a residual number of 132.1×10^{-4} at low SNR (0–3 dB), representing an 86.2% reduction compared to Goldstein filtering and a 71.5% reduction compared to InSAR-BM3D. As SNR increases, the residual number rapidly decreases to 10.9×10^{-4}, while the reduction rate remains stable between 82.7%–86.2% and 66.1%–71.5%, showing a slight downward trend. This is mainly because the residual number of the proposed method is already low at high SNR, limiting the potential for further improvement. These data validate the advantages of the proposed method in suppressing non-stationary noise and spurious phase jumps in interferometric phase maps.

C. VERIFICATION OF PHASE UNPACKING ACCURACY IMPROVEMENT

The least squares unpacking algorithm (with discrete cosine transform preconditioning acceleration and 1000 iterations) was used for each SNR interval in the test set. The unpacking error was calculated pixel by pixel and compared with the true absolute phase. The percentage of pixels with an error less than $\pi/2$ was taken as the success rate of that sample. Finally, the average value was taken. The results are shown in Fig. 3.

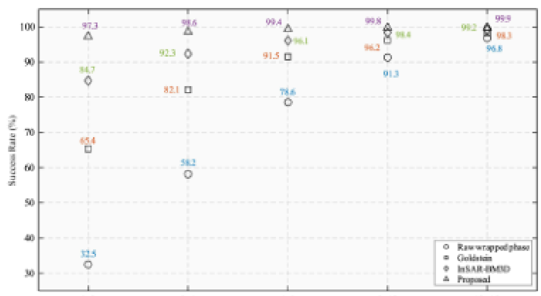

FIGURE 3. Unpacking success rate of each method in different SNR ranges (unit: %)

The proposed method achieves the highest unpacking success rate in all SNR ranges. In the noisiest 0–3 dB range, the proposed method achieves a success rate of 97.3%, an improvement of 31.9 percentage points compared to Goldstein filtering (65.4%) and 12.6 percentage points compared to InSAR-BM3D (84.7%), while direct unpacking with noisy conditions only achieves 32.5%, almost failing. As the SNR increases to 3–6 dB, the success rate of the proposed method exceeds 98%, and stabilizes above 99% after

6–9 dB. Among the comparative methods, InSAR-BM3D outperforms Goldstein filtering. These results demonstrate that the proposed denoising method effectively preserves phase continuity, enabling the least squares unpacking algorithm to reliably recover the absolute phase in high-noise regions.

D. ABLATION EXPERIMENTS

To verify the effectiveness of each core module, pre-trained weights for the full model and three ablation variants were loaded onto the test set: a phase gradient-free attention module (w/o PGAM, no attention weights embedded after encoder downsampling), a frequency domain discriminator (w/o Freq-D, only the spatial domain discriminator is retained and the adversarial loss weights remain unchanged), and a phase gradient consistency loss (w/o PGC). For 200 samples within each SNR interval, the number of residual points per 10^{-4} pixels was calculated using the same four-neighbor difference method as in Section 4.2, and the phase gradient error PGE was calculated for each configuration. All variant training hyperparameters were consistent with the full model (generator learning rate 0.0002, generator updated twice for every discriminator update). The average results are shown in Table 4.

TABLE 4. Number of Residual Points ($\times 10^{-4}$) and Phase Gradient Error PGE (rad) for Each Ablation Configuration in Different SNR Intervals

SNR range	Metric	Complete model	w/o PGAM	w/o Freq-D	w/o PGC
0–3	Points	132.1	189.4	168.7	201.2
	PGE	0.052	0.071	0.064	0.083
3–6	Points	79.5	112.3	98.1	119.6
	PGE	0.031	0.044	0.039	0.052
6–9	Points	43.8	61.9	54.6	66.7
	PGE	0.019	0.027	0.024	0.033
9–12	Points	22.4	32.7	28.5	35.9
	PGE	0.012	0.018	0.015	0.022
12–15	Points	10.9	16.8	14.2	18.7
	PGE	0.008	0.012	0.010	0.016

Table 4 shows that the complete model has the smallest number of residual points and PGE for all SNR intervals, which reveals the importance of the three modules to work together. Removing phase gradient attention module (w/o PGAM) leads to a larger number of residual points (132.1×10^{-4} to 189.4×10^{-4}) in the 0–3 dB range, since the generator's response to the fringe boundaries and steep gradient regions becomes weaker without the attention mechanism, leading to inadequate suppression of spurious phase jumps. Removing the frequency domain discriminator (w/o Freq-D) results in an increase of 21.7% and 18.75% in the number of residual points and PGE, respectively, and the performance of the discriminator is more affected in the high-frequency noise-sensitive range, which confirms the discriminator's capability to detect residual high-frequency noise. Removing the phase gradient consistency loss (w/o PGC) has the largest effect on the performance: residual points are increased by an average of 34.3% and PGE is increased by 37.3%. This is because the generator can give inconsistent directions of the gradient at fringe boundaries without the PGC loss, thereby creating many artifacts and extra residual points. This bias cannot be addressed with adversarial and structural similarity losses. Based on the above results, it is confirmed that the fringe boundary has been well preserved by the phase gradient attention module, the frequency domain discriminator can reduce the high-frequency noise, and the phase gradient consistency loss is a key constraint to suppress cross-fringe artifacts and reduce the number of residual points.

E. COMPUTATIONAL EFFICIENCY

The processing efficiency in practical interferometric synthetic aperture radar (InSAR) processing workflow is directly linked to the feasibility of the engineering applications, particularly in the processing of large-size interferometric phase maps (e.g., 2048×2048 pixels). The average processing time for the proposed method (generator inference stage) is compared with that of Goldstein filtering (frequency domain mask) and InSAR-BM3D (CPU implementation) in the same hardware platform. The average value was taken for each image size, with each being run 30 times. The input sizes included popular resolutions (256×256 to 2048×2048) and were all converted to megapixels (MP) for easy comparison. The results are shown in Fig. 4.

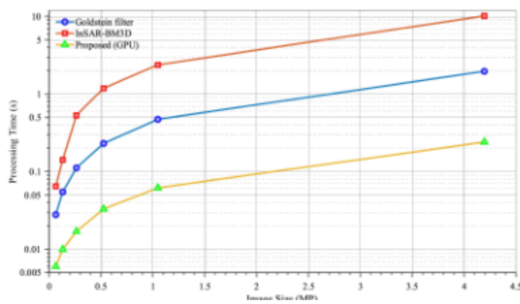

FIGURE 4. Processing Time at Different Image Sizes

Specifically, for a 1024×1024 image (1.048 MP), Goldstein filtering takes 0.474 s, InSAR-BM3D takes 2.385 s, while the proposed method only takes 0.062 s, achieving speedups of 7.6 times and 38.5 times, respectively. In the case of large images of 2048×2048, the proposed method is still able to keep the processing time below 0.25 s, while InSAR-BM3D is more than 10 s, which is difficult to meet real-time or near real-time processing requirements. The above results show that the designed network structure of the proposed generator (downsampling to lower spatial resolution and dilated convolution to keep the receptive field in the network) and the acceleration of GPU can significantly improve the inference speed compared to the traditional filtering algorithm based on CPU, and it can be applied in practice.

F. CASES: DENOISING AND UNPACKING RESULTS FOR TYPICAL LOW SNR SAMPLES

In order to illustrate the capacity of the proposed method to process in extreme noise, this section takes a sample representative of the test set for detailed analysis. The sample is from the SNR range 0–3 dB and has an actual SNR of 2.1 dB, a coherence coefficient of 0.38, and an additive noise variance of 0.176 rad^2 and is thus one of the samples with the highest noise intensity. In the image of the wrapped phase the multiplicative speckle noise and the additive Gaussian noise are much greater, and it is not easy to see the direction of the interference fringes from the image visually. After processing with the method presented in this paper, the phase gradient error (PGE) of this sample is 0.049 rad, a decrease of 88.7% compared to the noisy phase image. The number of residual points is suddenly cut down from 962 points in every 1000 to 128 points in every 1000 points. All the remaining residual points show a random distribution, no residual clump of points or no residual cluster. The resulting outputs of the generator's complex residuals are sparse, noisy and thus beneficial for this characteristic; on the other hand, the loss of phase gradient consistency directly penalizes deviations in the direction of the gradient, and avoids the generation of spurious residuals at the source. After employing the least squares unpacking algorithm (with discrete cosine transform preconditions and 1000 iterations), the unpacking success rate of this sample reached 96.5%, with unpacking errors for all pixels less than 0.8 rad (far below the failure threshold of $\pi/2$), complete phase recovery, and terrain undulation features highly consistent with the true label. In contrast, the success rate of direct unpacking of noisy phase maps was only 28.3%, almost a complete failure. This case study, from a single-sample perspective, confirms that the proposed method, through the synergistic effect of complex residual learning, phase gradient attention, and phase gradient consistency loss, reduces the number of residuals in high-noise environments by an order of magnitude, providing the least squares unpacking algorithm with ideal input conditions of "sparse and isolated residuals," thus enabling reliable recovery of the absolute phase even under extreme noise conditions.

V. CONCLUSION

This paper's method replaces direct phase mapping with complex residual learning, combining a multi-scale dilated convolutional U-Net with a phase gradient attention module, and introducing dual discriminators in the spatial and frequency domains, along with a phase gradient consistency loss function, to achieve both preservation of fringe structures and effective suppression of high-frequency noise. Although the proposed method achieves significant performance improvements on simulation data, its generalization ability to complex real-world InSAR scenarios involving atmospheric disturbances and registration errors still needs further verification, and the joint parameter tuning of multiple losses is quite sensitive. Future research can explore incorporating physical priors such as coherence coefficients as conditional inputs, developing unsupervised learning frameworks to reduce dependence on clean labels, and attempting end-to-end joint optimization of the denoising network and unpacking algorithm, thereby promoting the application of this method in practical interferometric engineering.

REFERENCES

[1] W. Luo, S. Ye, W. Xiong, Z. Zhang, X. Wang, S. Li, and F. Wang, "A noise reduction method for continuous optical spatial heterodyne interferograms based on deep convolutional neural networks," Infrared and Laser Engineering, vol. 54, no. 3, pp. 352–365, 2025, doi: https://doi.org/10.3788/IRLA20240392.

[2] L. Tao, G. Huang, S. Yang, T. Wang, H. Sheng, and H. Fan, "An interferometric image denoising method using convolutional neural networks," Journal of Wuhan University (Information Science Edition), vol. 48, no. 4, pp. 559–567, 2023, doi: https://doi.org/10.13203/j.whugis20200589.

[3] S. Wang, J. Liu, J. Wu, G. Liu, X. Ling, and L. Yu, "A Denoising Method for Sheared Speckle Phase Map Based on Improved Particle Swarm Optimization Algorithm," Measurement & Control Technology, vol. 42, no. 6, pp. 78–83, 2023, doi: https://doi.org/10.19708/j.ckjs.2023.06.009.

[4] H. K. Yadalam, M. Kizmann, J. R. Rouxel, et al., "Quantum interferometric pathway selectivity in difference-frequency-generation spectroscopy," The Journal of Physical Chemistry Letters, vol. 14, no. 48, pp. 10803–10809, 2023, doi: https://doi.org/10.1021/acs.jpclett.3c02341.

[5] V. Gordienko, F. M. Ferreira, V. Ribeiro, et al., "Design of an interferometric fiber optic parametric amplifier for the rejection of unwanted four-wave mixing products," Optics Express, vol. 31, no. 5, pp. 8226–8239, 2023, doi: https://doi.org/10.1364/OE.476884.

[6] R. Jie, H. Zheng, O. Alsalman, et al., "High-sensitivity Fabry-Perot interferometric sensor based on microwave photonics with phase demodulation," Journal of Lightwave Technology, vol. 42, no. 19, pp. 6958–6966, 2024, doi: https://doi.org/10.1109/JLT.2024.3409536.

[7] T. Natale, P. B. Núñez, L. Dindelli, et al., "Modeling interferometric optical gyroscopes: An experimentally validated approach," IEEE Sensors Journal, vol. 25, no. 6, pp. 9560–9566, 2025, doi: https://doi.org/10.1109/JSEN.2025.3530638.

[8] A. Florio, G. Avitabile, and G. Coviello, "Multiple source angle of arrival estimation through phase interferometry," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 69, no. 3, pp. 674–678, 2022, doi: https://doi.org/10.1109/TCSII.2022.3141247.

[9] S. Rasouli and S. Khoshkhatti, "Single-frame double-grating interferometry with two inherent-phase-shifted interferograms," Optics Express, vol. 32, no. 13, pp. 22429–22447, 2024, doi: https://doi.org/10.1364/OE.521731.

[10] J. R. Sánchez-Aguilar, A. K. Reyes, L. García-Lechuga, et al., "Polarizing white light interferometry for phase measurements using two simultaneous interferograms," Applied Optics, vol. 62, no. 27, pp. 7280–7287, 2023, doi: https://doi.org/10.1364/AO.496606.

[11] S. Kim, J. Jeon, Y. Kim, et al., "Design and assessment of phase-shifting algorithms in optical interferometer," International Journal of Precision Engineering and Manufacturing-Green Technology, vol. 10, no. 2, pp. 611–634, 2023, doi: https://doi.org/10.1007/s40684-022-00495-z.

[12] F. Wang, X. Li, S. Ye, S. Li, and X. Wang, "Error analysis method for spatial heterodyne interferograms using wavelet transform," Spectroscopy and Spectral Analysis, vol. 45, no. 1, pp. 8–14, 2025, doi: https://doi.org/10.3964/j.issn.1000-0593(2025)01-0008-07.

[13] P. J. de Groot, L. L. Deck, R. Su, et al., "Contributions of holography to the advancement of interferometric measurements of surface topography," Light: Advanced Manufacturing, vol. 3, no. 2, pp. 258–277, 2022, doi: https://doi.org/10.37188/lam.2022.007.

[14] P. J. Stas, Y. C. Wei, M. Sirotin, et al., "Entanglement-assisted non-local optical interferometry in a quantum network," Nature, vol. 651, no. 8105, pp. 326–332, 2026.