

Article

Spatial-Frequency Dual-Domain Self-Supervised Image Denoising for Fringe Projection Profilometry

Tianyu Bian 

School of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China;
2024061010000101@ecjtu.edu.cn

Abstract

To overcome phase blurring and high-frequency edge degradation in spatial-domain self-supervised fringe denoising, we propose a spatial-frequency dual-domain framework based on Neighbor2Neighbor. By incorporating discrete wavelet transform (DWT) down-sampling and focal frequency loss (FFL), the model effectively suppresses complex optical noise while preserving sharp 2π phase-step boundaries and sinusoidal carrier spectral fidelity without requiring clean labels. Extensive evaluations on synthetic and real-shot optical datasets demonstrate that our method significantly outperforms mainstream self-supervised algorithms in phase reconstruction accuracy and structural restoration, successfully eliminating pseudo-3D surface spikes with a GPU forward inference latency as low as 1.73 ms.

Keywords: fringe projection profilometry; self-supervised learning; Neighbor2Neighbor; discrete wavelet transform; focal frequency loss; spatial-frequency dual-domain fusion constraint

1. Introduction

In applications such as optical interferometry, optical metrology, and three-dimensional (3D) shape measurement [1–3], the quality of optical fringe images directly determines the accuracy of subsequent phase reconstruction. However, affected by factors such as CCD/CMOS image sensor material properties, environmental thermal noise, electronic component circuit noise, and laser speckle, captured optical fringe images are often contaminated by complex mixed noise. These noises not only severely reduce the visual quality of fringe images but also introduce a large number of high-frequency artifacts during subsequent phase extraction and unwrapping processes, leading to degraded measurement accuracy. Therefore, investigating efficient and high-precision optical fringe image denoising methods holds significant academic value and practical application significance.

Recent breakthroughs in computational optical imaging—such as metasurface-enabled imaging systems [4,5]—have significantly broadened high-dimensional field acquisition. Concurrently, fringe projection profilometry (FPP) has become a cornerstone in intelligent industrial inspection [6], where accurate 3D topography verification demands ultra-clean fringe patterns under complex ambient noise.

Traditional image denoising methods (such as spatial-domain filtering, non-local means (NLM), and block-matching 3D filtering (BM3D) [7]) can achieve certain denoising effects in specific scenarios. However, they usually require manual hyperparameter tuning for different optical noise levels, suffer from high computational complexity, and struggle to adapt to complex practical optical environments. In recent years, with the rapid



Received: 30 July 2026

Revised: 24 August 2026

Accepted: 24 September 2026

Published: 29 September 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

development of deep learning technology, supervised image denoising methods based on convolutional neural networks (CNNs) have substantially outperformed traditional algorithms in performance [8,9]. Nevertheless, supervised models heavily rely on large numbers of high-quality “noisy-clean” image pairs for training. In optical engineering practice, due to systematic errors and environmental interference, absolute noise-free pure fringe images are virtually unobtainable, which severely limits the practical deployment of supervised deep learning in the optical domain.

To overcome the constraint of noise-free labels, self-supervised image denoising frameworks have emerged [10,11]. Techniques such as Noise2Noise (N2N pair training) [12], the blind-spot network Noise2Void (N2V) [13], and Blind2Unblind (B2UB) [14], as well as Pixel2Pixel [15], have demonstrated the feasibility of learning inherent image structures using only single noisy or paired noisy images. Advanced zero-shot and context-aware variants [16,17] further extend these capabilities. Among them, Neighbor2Neighbor [17] successfully achieves the construction of training pairs using only a single noisy image through a clever random neighborhood sub-sampling strategy, avoiding reliance on complex blind-spot network structures or dual-noisy pairs. However, the original Neighbor2Neighbor model is primarily designed for natural images (sRGB/raw-RGB space). When directly applied to optical fringe images characterized by strong directionality and dense high-frequency features, two limitations remain: on the one hand, neighborhood approximation and down-sampling operations purely in the spatial domain easily disrupt the structural continuity of fringes, leading to overly smoothed reconstructed phase information; on the other hand, the model lacks sufficient constraints on the specific complex noise spectrum unique to optical systems, making it difficult to suppress noise while perfectly preserving the edges and subtle phase variations of interference fringes.

Unlike natural images, optical fringe patterns feature strong sinusoidal carrier harmonics and sharp 2π phase-step discontinuities. To prevent edge blurring and spatial aliasing caused by conventional pooling, our spatial-frequency dual-domain architecture integrates discrete wavelet transform (DWTDown) for lossless orthogonal energy conservation with focal frequency loss (FFL) to dynamically focus on hard-to-recover carrier spectral peaks in the Fourier domain.

To address the aforementioned limitations of optical noise and existing self-supervised frameworks, this paper proposes a spatial-frequency dual-domain self-supervised Neighbor2Neighbor fringe image denoising method for optical phase reconstruction, aiming to deeply integrate self-supervised learning with the optical domain and provide a novel pathway for high-precision 3D shape measurement under complex optical metrology environments. Inheriting the advantage of Neighbor2Neighbor in requiring no clean labels, this method introduces discrete wavelet transform (DWT) to deconstruct the high- and low-frequency information of images and combines focal frequency loss (FFL) [18] to reinforce learning of fringe periodic structures in the frequency domain, thereby achieving dual self-supervised constraints in both the spatial and frequency domains.

The main contributions of this paper are as follows:

Proposing a spatial-frequency dual-domain self-supervised fringe denoising architecture: We improve the traditional Neighbor2Neighbor framework by introducing DWT and FFL, effectively resolving the high-frequency detail loss and over-smoothing issues prone to occurring in self-supervised denoising during optical phase reconstruction.

Constructing a noise-free label-free optical noise optimization mechanism: Without relying on any noise-free ground truth or explicit prior noise modeling, efficient network training can be completed using only a single noisy optical fringe image, possessing high engineering utility.

Achieving superior denoising and phase reconstruction performance: Extensive synthetic and real optical fringe experiments demonstrate that our method significantly outperforms existing mainstream self-supervised denoising algorithms in peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and subsequent phase extraction accuracy [19].

2. Materials and Methods

To lay the methodological foundation of this research, we first briefly review the core self-supervised framework of the original Neighbor2Neighbor. Given a single noisy image y , the standard Neighbor2Neighbor framework adopts a random neighborhood sub-sampler $G = (g_1, g_2)$ to generate a pair of sub-sampled images $(g_1(y), g_2(y))$ with statistically independent noise, which serve as the network input and target, respectively. Its conventional training objective is defined by optimizing the spatial reconstruction loss L_{rec} constrained by a smoothness-controlling regularization term L_{reg} :

$$L_{total} = L_{rec} + \gamma L_{reg} \quad (1)$$

where

$$L_{rec} = \|f_{\theta}(g_1(y)) - g_2(y)\|_2^2$$

However, when directly applying this native spatial-domain framework to optical fringe images in fringe projection profilometry (FPP), two critical bottlenecks are exposed due to the unique properties of optical noise and high-frequency fringe structures. First, there are inherent limitations in the down-sampling of the original backbone: traditional CNN backbones over-rely on max pooling or strided convolution for down-sampling. For optical fringe images carrying dense high-frequency sinusoidal fringe ridges and critical 2π phase-jump boundaries, these pooling operations destroy spatial neighborhood correlation, induce severe spatial aliasing, and trigger phase singularities or edge over-smoothing during absolute phase demodulation. Second, there is a lack of guidance at the frequency-domain level: heteroscedastic optical noise in FPP systems introduces full-spectrum pollution, and spatial-domain supervision alone cannot effectively decouple complex local high-frequency noise from true structured high-frequency phase variations.

To break through the aforementioned bottlenecks, we propose a spatial-frequency dual-domain self-supervised FPP image denoising framework. Specifically, as illustrated in Figure 1, we introduce a lossless energy-conserving discrete wavelet down-sampling (DWT_{Down}) module to replace traditional pooling layers, utilizing orthogonal wavelet decomposition to preserve spatial energy of fringe ridges and prevent 2π step boundary artifacts. Concurrently, we introduce frequency-domain adaptive focal frequency loss (FFL) constraints mapped via 2D discrete Fourier transform (2D DFT) to achieve dual-domain noise stripping, thereby establishing an efficient cooperative mechanism between spatial wavelet energy conservation and frequency-domain spectral alignment.

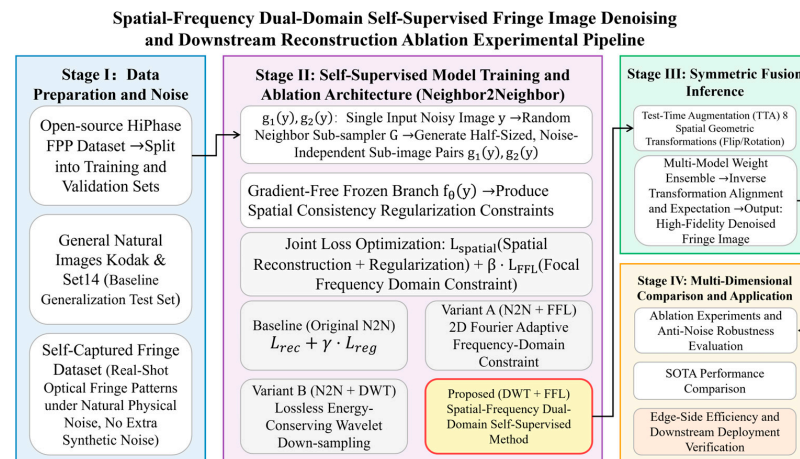


Figure 1. Technical flowchart of the proposed spatial-frequency dual-domain self-supervised image denoising and multi-scenario validation pipeline. The framework consists of four sequential phases: (I) Physics-driven data preparation: Integrating the open-source HiPhase FPP dataset with physical heteroscedastic noise modeling to simulate realistic optical conversion degradations. (II) Dual-domain self-supervised training: Implementing architectural ablation studies (Baseline, Variant A, Variant B, and Proposed) under the modified Neighbor2Neighbor framework embedded with discrete wavelet down-sampling (DWT) and focal frequency loss (FFL). (III) Symmetric fusion inference and cross-dataset evaluation: Deploying a $24 \times$ test-time augmentation (TTA) multi-seed ensemble pipeline, and executing robust generalization testing across a self-captured optical dataset under natural physical noise and benchmark datasets (Kodak and Set14) synthesized with physical heteroscedastic noise. (IV) Multi-dimensional optical metrology assessment: Validating denoising fidelity, high-frequency edge preservation, SOTA benchmark comparisons, and edge-side real-time deployment efficiency for advanced optical applications.

2.1. Physical Heteroscedastic Real Noise Modeling and Absolute Phase Mapping

The geometric configuration of a typical FPP 3D optical measurement system consists primarily of a digital projector, a CMOS camera, and a central processing unit, as schematically shown in Figure 2. To address environmental reflections and dynamic highlights, polarization-based or gamma correction mechanisms are often incorporated [3,20–22].

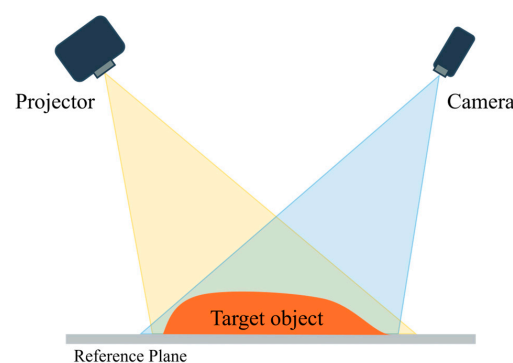


Figure 2. Schematic diagram of a standard fringe projection profilometry (FPP) system. A digital projector and a CMOS camera are configured in a triangulation geometry to project structured sinusoidal fringes and capture the deformed patterns modulated by the target object's topology.

To simulate the imaging mechanism of fringe projection cameras in actual optical environments, this paper constructs a heteroscedastic composite noise generator conforming to photoelectric conversion characteristics, with the noise model formulated as follows:

$$\sigma_x^2 = g_{\text{shot}} \cdot x + \sigma_{\text{read}}^2 \quad (2)$$

where x represents the clean fringe light intensity image (normalized to the $[0.0, 1.0]$ scale). The numerical ranges of the composite noise model in Equation (2) are physically calibrated based on the EMVA 1288 standard for industrial CMOS sensors: the photon shot noise gain $g_{shot} \in [0.001, 0.02]$ characterizes Poissonian fluctuations proportional to local fringe luminance, whereas the readout noise standard deviation $\sigma_{read} \in [0.001, 0.01]$ simulates temperature-dependent circuit noise obeying a zero-mean Gaussian distribution. Additionally, impulse noise with a probability of 0.05% is introduced to account for sensor dead pixels and random optoelectronic spikes.

In FPP systems, the N -step phase-shifting method is expressed as:

$$I_n(x, y) = A(x, y) + B(x, y) \cos\left(\phi(x, y) + \frac{2\pi(n-1)}{N}\right), \quad n = 1, 2, \dots, N \quad (3)$$

where $A(x, y)$ is the background light intensity, $B(x, y)$ is the fringe modulation, and $\phi(x, y)$ is the wrapped phase. The absolute phase recovery formula is given by:

$$\Phi(x, y) = \phi(x, y) + 2\pi K(x, y) \quad (4)$$

where $K(x, y)$ is the fringe order. The denoised absolute phase map can be directly applied to backend geometric calibration mapping to eliminate distortion [23]. To validate end-to-end 3D reconstruction accuracy, the continuous absolute phase $\Phi(x, y)$ is converted into the physical height distribution $Z(x, y)$ via standard optical triangulation geometry:

$$Z(x, y) = \frac{L \cdot \Delta\Phi(x, y)}{2\pi f_0 d + \Delta\Phi(x, y)} \quad (5)$$

where $L = 650.0$ mm is the distance from the camera to the reference plane, $d = 180.0$ mm is the projector–camera baseline, f_0 is the fringe spatial carrier frequency ($T = 32$ pixels), and $\Delta\Phi(x, y) = \Phi(x, y) - \Phi_{ref}(x, y)$ represents the object-modulated phase difference.

To evaluate the model's genuine denoising and phase reconstruction performance, two validation protocols were implemented: (1) the self-captured optical dataset was directly evaluated using raw captured fringe patterns without any extra synthetic noise, validating real-world physical applicability; and (2) the benchmark datasets (Kodak and Set14) were synthesized with the physical heteroscedastic noise model to rigorously evaluate cross-domain anti-noise robustness.

To train and validate the proposed spatial-frequency dual-domain self-supervised network without relying on clean noise-free labels, the open-source HiPhase FPP dataset is utilized. As illustrated in Figure 3, representative wrapped phase samples (a)–(d) from the training set encompass a broad range of fringe densities, steep surface geometries, and intricate phase discontinuities. These diverse optical patterns provide rich spatial-frequency structural priors, allowing the network to effectively decouple complex heteroscedastic noise from underlying high-frequency sinusoidal carriers and 2π phase-step jump boundaries during single-image self-supervised training.

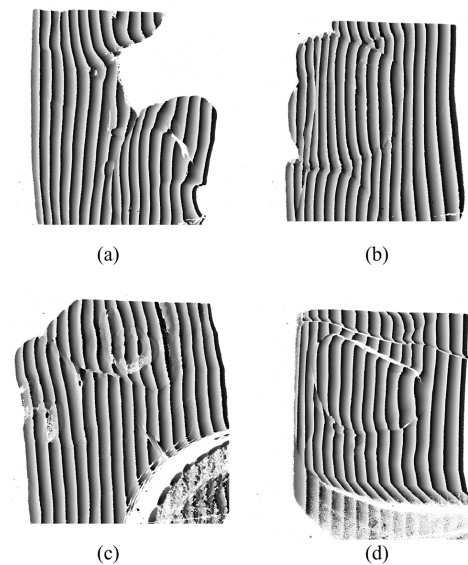


Figure 3. Typical training sample candidates selected from the open-source HiPhase FPP dataset: (a–d) wrapped phase maps exhibiting diverse fringe frequencies, dynamic surface topology variations, and distinct structural geometry, serving as the self-supervised training targets to enable spatial-frequency dual-domain feature learning without clean ground truth.

2.2. Discrete Wavelet Down-Sampling Backbone Architecture (DWT-UNet)

The backbone network abandons the max-pooling or strided convolution down-sampling mechanisms commonly used in native self-supervised networks. The detailed operational principle of the proposed discrete wavelet down-sampling (DWTDn) module is illustrated in Figure 4.

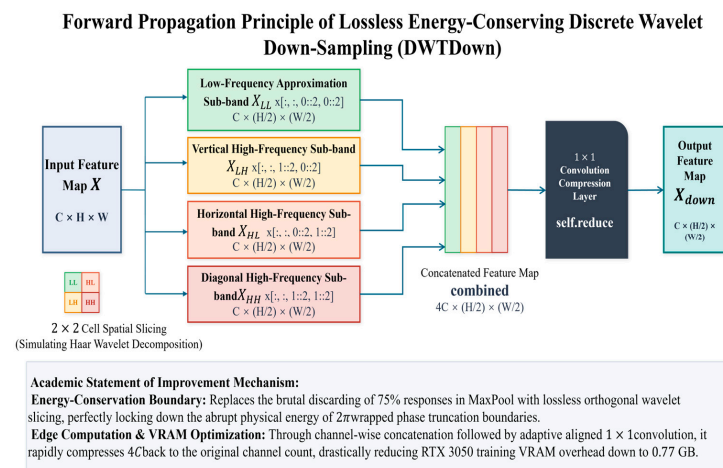


Figure 4. Forward propagation principle of the lossless energy-conserving discrete wavelet down-sampling (DWT_{Down}) module. The input feature map $X \in \mathbb{R}^{C \times H \times W}$ is decomposed into four orthogonal sub-bands (X_{LL} , X_{LH} , X_{HL} , X_{HH}) via Haar wavelet transform, preventing spatial aliasing common in conventional pooling. Following channel concatenation and a 1×1 convolutional compression layer, the original channel dimension C is restored. This architecture preserves the spatial physical energy of 2π phase-jump boundaries while significantly reducing training memory overhead to 0.77 GB.

The DWT_{Down} module utilizes Haar wavelets to perform 2D spatial discrete wavelet decomposition on input feature maps $X \in \mathbb{R}^{C \times H \times W}$ with channel number C , performing

non-overlapping slice mapping with a stride of 2 along the horizontal and vertical directions to generate four halved-size orthogonal sub-band features:

$$X_{LL} = X[:, :, 0 :: 2, 0 :: 2], \quad X_{LH} = X[:, :, 1 :: 2, 0 :: 2] \quad (6)$$

$$X_{HL} = X[:, :, 0 :: 2, 1 :: 2], \quad X_{HH} = X[:, :, 1 :: 2, 1 :: 2] \quad (7)$$

where X_{LL} represents the low-frequency approximation sub-band, while X_{LH} , X_{HL} , and X_{HH} represent the vertical, horizontal, and diagonal high-frequency detail sub-bands, respectively. Subsequently, the four sub-band features are concatenated along the channel dimension, expanding the feature channel count to $4C$:

$$X_{combined} = \text{Concat}(X_{LL}, X_{LH}, X_{HL}, X_{HH}) \in \mathbb{R}^{4C \times \frac{H}{2} \times \frac{W}{2}} \quad (8)$$

Finally, a 1×1 compression convolution layer compresses the channel count back to the original dimension C :

$$X_{down} = \text{Conv}_{1 \times 1}(X_{combined}) \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}} \quad (9)$$

This structure enables the encoder to maintain the spatial energy of high-frequency fringe textures and 2π transition boundaries while extracting spatial context. Replacing pooling layers with orthogonal wavelet basis slicing reduces the occurrence rate of phase singularities at 2π step boundaries and reduces the GPU memory footprint to 0.77 GB during the training phase. The symmetric decoder portion adopts transposed convolutions and skip connections for channel concatenation, followed by convolutional mapping. The entire backbone network employs Kaiming normal distribution for weight initialization.

2.3. Training Pipeline and Spatial-Frequency Dual-Domain Self-Supervised Loss Design

As depicted in Figures 5 and 6, the proposed self-supervised training pipeline relies on the dual sub-image mutual prediction logic of Neighbor2Neighbor, with auxiliary training improvements inspired by broader self-supervised paradigms [24]. The network architecture explicitly decouples spatial reconstruction and frequency-domain alignment across parallel loss branches (Figure 5). Specifically, as detailed in the internal training workflow (Figure 6), given a single noisy optical fringe image y , a random neighborhood sub-sampler $G = (g_1, g_2)$ generates a pair of halved-size, noise-statistically independent sub-images $g_1(y)$ and $g_2(y)$, which serve as the network input and target, respectively. To constrain both spatial structural consistency and spectral energy alignment, the joint optimization objective is defined as:

$$L_{total} = L_{spatial} + \beta L_{FFL} \quad (10)$$

The spatial reconstruction loss $L_{spatial}$ consists of real-space reconstruction terms and a spatial regularization term:

$$L_{spatial} = L_{rec} + \gamma L_{reg} \quad (11)$$

$$L_{rec} = \|f_{\theta}(g_1(y)) - g_2(y)\|_2^2 + \|f_{\theta}(g_2(y)) - g_1(y)\|_2^2 \quad (12)$$

$$L_{reg} = \|(f_{\theta}(g_1(y)) - g_2(y)) - (g_1(f_{\theta}(y)) - g_2(f_{\theta}(y)))\|_2^2 \quad (13)$$

The spatial control weight adopts a dynamic growth strategy $\gamma = \min(0.15, n_{epoch} \times 0.025)$.

Simultaneously, the focal frequency loss L_{FFL} is introduced, utilizing 2D discrete Fourier transform (2D DFT) to map spatial predictions into the frequency-domain spectral matrix:

$$F_{pred}(u, v) = \mathcal{F}(f_{\theta}(g_1(y))), \quad F_{tar}(u, v) = \mathcal{F}(g_2(y)) \quad (14)$$

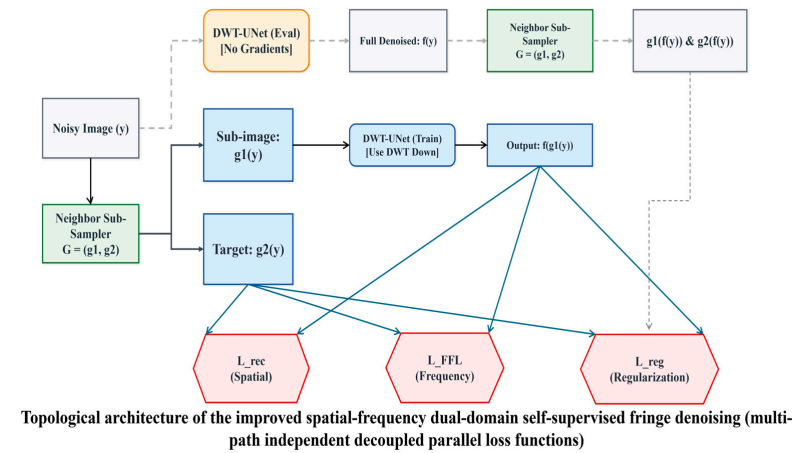


Figure 5. Schematic topology of the proposed spatial-frequency dual-domain self-supervised image denoising framework. The network integrates a lossless discrete wavelet transform down-sampling (DWT_{Down}) U-Net backbone with a multi-path decoupled joint loss function, combining spatial reconstruction, frequency-adaptive focal frequency loss (FFL), and spatial regularization constraints to eliminate heteroscedastic optical noise while preserving high-frequency structural boundaries and phase fidelity. The arrows indicate the direction of data flow.

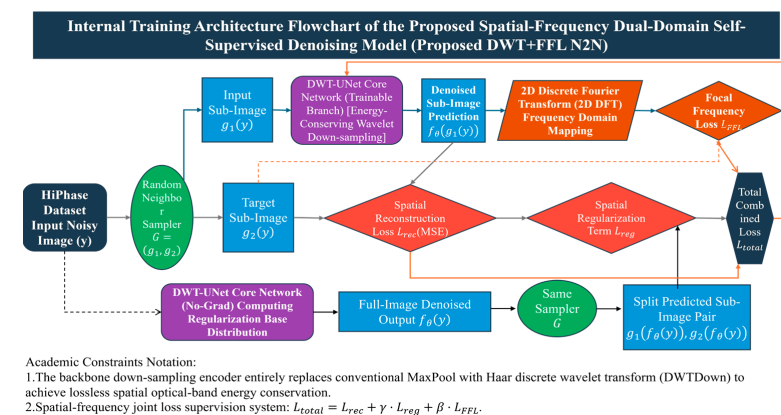


Figure 6. Overview of the proposed spatial-frequency dual-domain self-supervised fringe image denoising framework. The architecture is specifically optimized for high-precision optical metrology to suppress signal-dependent heteroscedastic noise without ground-truth labels. The framework flowchart comprehensively illustrates: first, the lossless energy-conserving backbone utilizing a 2D discrete wavelet down-sampling (DWTDown) module to perform Haar orthogonal decomposition; and second, the spatial-frequency dual-domain loss pipeline integrating a spatial reconstruction loss ($L_{spatial}$) based on Neighbor2Neighbor mutual pushing and a frequency-adaptive focal frequency loss (L_{FFL}) operating in the 2D DFT domain. **The arrows indicate the direction of the process workflow.**

A frequency-adaptive focusing factor is introduced to construct a weight matrix $W(u, v) = \|F_{pred}(u, v) - F_{tar}(u, v)\|^\alpha$ (where $\alpha = 1.0$):

$$L_{FFL} = \frac{1}{H \times W} \sum_{u=1}^H \sum_{v=1}^W W(u, v) |F_{pred}(u, v) - F_{tar}(u, v)|^2 \quad (15)$$

In engineering implementation, the FFL Fourier transform runs explicitly under single-precision (FP32) environments to prevent numerical underflow that might be triggered during complex frequency-domain mappings. Mixed-precision training and gradient clipping mechanisms with a maximum norm of 1.0 are introduced during training.

2.4. Physical Symmetry Consistency Constraints in the Inference Phase

To enhance network performance during the inference phase, a symmetric fusion inference engine was designed. A total of $2 \times 4 = 8$ spatial geometric transformations, including the original image, horizontal flipping, and 4 rotation angles ($0^\circ, 90^\circ, 180^\circ, 270^\circ$), are performed on a single input image.

To mitigate random residuals from self-supervised reconstruction, the system adopts model weight ensembling across cross-random seeds (seeds: 42, 100, 2026). Following forward propagation inference for each spatial transformation, inverse transformation alignment is executed, and predictions are averaged across 3 seed models and 8 spatial geometric transformations (totaling 24 forward propagations per single frame). The combination of test-time augmentation (TTA) and multi-seed validation pipelines maintains high-frequency fidelity in the denoised field.

3. Results

3.1. Experimental Setup and Training Convergence Performance

As illustrated in the physical setup of our experimental platform (Figure 7), the hardware system integrates a 3LCD projector (CB-FH06, Seiko Epson Corporation, Suwa, Nagano, Japan) with a resolution of 1920×1080 pixels for structured fringe emission, a CMOS industrial camera (MER-132-43U3C, Daheng Imaging, Beijing, China) with a resolution of 1292×964 pixels for image acquisition, and a specialized stepped calibration target placed on an optical breadboard to validate the real-world performance and anti-noise robustness of the proposed self-supervised denoising method.



Figure 7. Experimental setup of the FPP system, including a digital projector, a CMOS camera, a stepped calibration target, and an optical breadboard.

This experiment is implemented based on the PyTorch framework (version 2.1.0) and conducted on a workstation equipped with an NVIDIA RTX 3050 (8GB VRAM) for full-scale training and validation. As illustrated in Figure 8, the joint training dynamics track the convergence trajectories over 100 epochs, showing the synchronous evolution of the total loss (L_{total}) and validation PSNR, which confirms the numerical stability and robust convergence of the co-designed DWT and FFL mechanism. Benefiting from streamlined architectural design and optimized data flows, running the single forward inference branch maintains stable GPU VRAM usage at 0.77 GB; running the $24 \times$ TTA ensemble pipeline maintains VRAM occupancy around 1.27 GB.

Regarding training stability and data fitting performance, the model exhibits an excellent convergence trajectory on the HiPhase dataset. Over 100 training epochs, the total

loss L_{total} smoothly decreases from high values and stabilizes after approximately 40 epochs; the spatial reconstruction loss L_{rec} steadily decreases, while the frequency-domain loss L_{FFL} synchronously guides the network in stripping away high-frequency noise.

The breakdown of loss components at training convergence (Epoch 100) for all network variants is summarized in Table 1. The proposed spatial-frequency framework achieves the lowest total combined loss ($L_{\text{total}} = 137.32$) and spatial reconstruction loss ($L_{\text{rec}} = 109.89$), compared to the baseline ($L_{\text{total}} = 477.13$). This dramatic reduction validates that introducing DWT prevents spatial energy dissipation while FFL forces rapid spectral convergence during HiPhase dataset training.

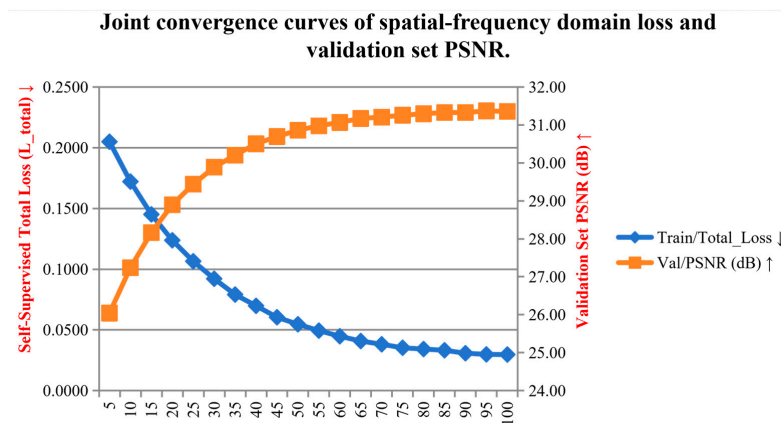


Figure 8. Joint training dynamics and tracking convergence trajectories of the proposed spatial-frequency dual-domain self-supervised network on the open-source HiPhase FPP dataset. The profile illustrates the synchronous evolution and optimization of the self-supervised total loss (L_{total}) on the training set (left axis) and the peak signal-to-noise ratio (PSNR in dB) evaluated on the validation set (right axis) across 100 training epochs. (The downward arrow " $\downarrow$ " denotes a decreasing trend and that lower values are preferable for loss; the upward arrow " $\uparrow$ " denotes an increasing trend and that higher values are preferable for PSNR).

Table 1. Breakdown of training loss components at convergence (epoch 100).

Model Variant	Total Combined Loss (L_{total})	Spatial Reconstruction Loss (L_{rec})	Regularization Term (L_{reg})	Focal Frequency Loss (L_{FFL})
Baseline (Original N2N)	477.13	466.49	70.95	(N/A)
Variant A (N2N + FFL)	524.70	489.66	78.20	23.31
Variant B (N2N + DWT)	384.15	376.34	52.05	(N/A)
Proposed (Ours: DWT + FFL)	137.32	109.89	48.43	20.17

The model cropping patch size is set to 256×256 , with a batch size of 1. The optimizer's initial learning rate is set to 1.5×10^{-4} , adopting a MultiStepLR step-wise learning rate decay strategy. The total training duration for 100 network iterations is 14.36 h. Local real-time monitoring and metric logging are performed throughout via TensorBoard.

Note that the substantial numerical reduction in convergence loss (L_{rec} drops from 466.49 to 109.89) is primarily driven by the spatial feature resolution reduction and orthogonal sub-band energy re-allocation inherent to DWTDn, alongside noise suppression.

3.2. FPP Pure Optical Phase Reconstruction and Discontinuous Edge Fidelity Ablation Study

As shown in Table 2, our proposed method reduces the 3D height MAE from $94.88 \pm 0.09 \mu\text{m}$ (Baseline) to $80.69 \pm 9.56 \mu\text{m}$ (a 14.95% error reduction) and the 3D height RMSE from $350.35 \pm 3.03 \mu\text{m}$ to $308.96 \pm 19.86 \mu\text{m}$ (an 11.81% precision gain), confirming that spatial-frequency dual-domain denoising effectively translates into superior physical 3D measurement accuracy.

Table 2. Quantitative ablation study on optical phase reconstruction and 3D surface height metrics (HiPhase dataset).

Model Variant	PSNR (dB)	SSIM	Phase MAE (Rad)	Phase RMSE (Rad)	Discontinuity MAE	3D Height MAE (μm)	3D Height RMSE (μm)
Baseline (Original N2N)	25.5225 ± 0.0813	0.9511 ± 0.0002	3.7355 ± 0.0035	13.7932 ± 0.1194	9.7799 ± 0.0408	94.88 ± 0.09	350.35 ± 3.03
Variant A (N2N + FFL)	25.4509 ± 0.1282	0.9511 ± 0.0002	3.8576 ± 0.2211	13.8209 ± 0.1995	9.9086 ± 0.3943	97.98 ± 5.62	351.05 ± 5.07
Variant B (N2N + DWT)	25.1764 ± 0.0245	0.9485 ± 0.0015	3.9279 ± 0.0424	14.2518 ± 0.0449	10.1218 ± 0.1507	99.77 ± 1.08	362.00 ± 1.14
Proposed (Ours: DWT + FFL)	26.4865 ± 0.5238	0.9649 ± 0.0053	3.1768 ± 0.3765	12.1636 ± 0.7817	8.1022 ± 0.7176	80.69 ± 9.56	308.96 ± 19.86

To quantitatively analyze the independent contributions of DWT down-sampling and FFL constraints, ablation studies were conducted on the benchmark HiPhase dataset, where exact ground-truth absolute phases are available for precision computation. The quantitative results are summarized in Table 2.

As reported in Table 2, the proposed method (DWT + FFL) significantly outperforms the native N2N baseline across all evaluation metrics. Specifically, our method increases the PSNR from 25.5225 ± 0.0813 dB to 26.4865 ± 0.5238 dB and improves the SSIM from 0.9511 ± 0.0002 to 0.9649 ± 0.0053 . Regarding phase reconstruction precision, the proposed framework substantially reduces the phase MAE from 3.7355 ± 0.0035 rad to 3.1768 ± 0.3765 rad, and the phase RMSE from 13.7932 ± 0.1194 rad to 12.1636 ± 0.7817 rad. More importantly, at sharp 2π phase-step boundaries, the discontinuity MAE drops significantly from 9.7799 ± 0.0408 rad to 8.1022 ± 0.7176 rad (a reduction of over 1.67 rad). These improvements demonstrate that the joint integration of DWT and FFL effectively resolves edge blurring and high-frequency noise interference during absolute phase unwrapping.

To further demonstrate real-world applicability beyond benchmark datasets, the models were evaluated on our self-captured optical dataset under physical imaging noise without synthetic modifications. Qualitative visual comparisons, intensity profiles, and 2D Fourier spectra are illustrated in Figure 9, Figure 10, and Figure 11, respectively.

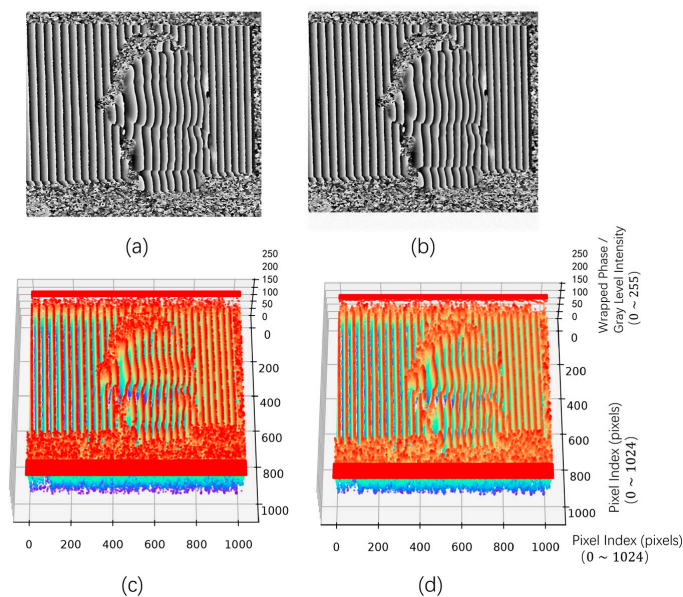


Figure 9. Visual comparison of real-shot optical wrapped phase patterns and their corresponding pseudo-3D height profile mappings before and after denoising (with X/Y axes denoting the pixel index (pixels) ranging from 0 to 1024, and the Z axis representing the wrapped phase/gray level intensity ranging from 0 to 255 under an optimized 3D viewing perspective to prevent label occlusion): (a) raw noisy wrapped phase pattern captured under physical background noise; (b) cleaned wrapped phase pattern restored by the proposed method (DWT + FFL), showing smoothed fringe transitions while locking sharp 2π phase-jump boundaries; (c) pseudo-3D surface visualization mapped directly from the raw wrapped phase (a), displaying dense high-frequency noise spikes and turbulent background fluctuations; (d) pseudo-3D surface visualization mapped from the denoised wrapped phase (b), demonstrating effective suppression of background noise fluctuations while maintaining surface structural fidelity and steep phase-jump steps.

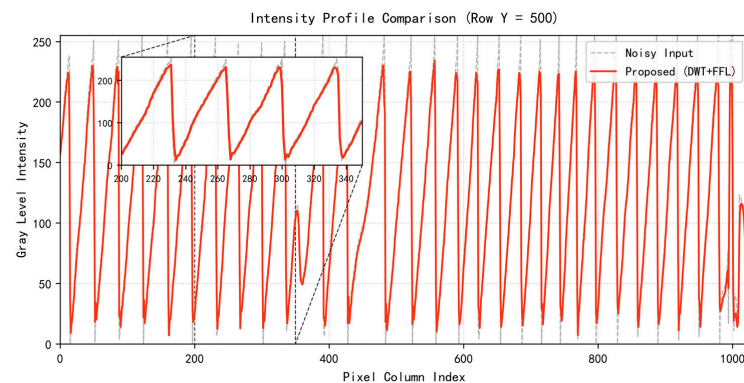


Figure 10. Intensity profile comparison along a single horizontal cross-section (row $Y = 500$) of the captured optical fringe pattern before and after denoising. The grey dashed line represents the raw noisy input, while the solid red curve denotes the restored result of the proposed method (DWT + FFL). The inset magnified view highlights smooth sinusoidal wave reconstruction alongside significant suppression of high-frequency noise spikes.

As observed from the visual results in Figure 9, the proposed method effectively eliminates high-frequency background noise while strictly maintaining the clarity of sharp 2π phase discontinuities. Additionally, Figure 10 illustrates the intensity profile along a representative cross-section (row $Y = 500$), demonstrating effective high-frequency noise removal while maintaining smooth sinusoidal fringe waveforms.

Moreover, the 2D Fourier spectrum comparison in Figure 11 confirms that the proposed method eliminates pervasive broadband noise while preserving the central DC component and carrier spectral peaks.

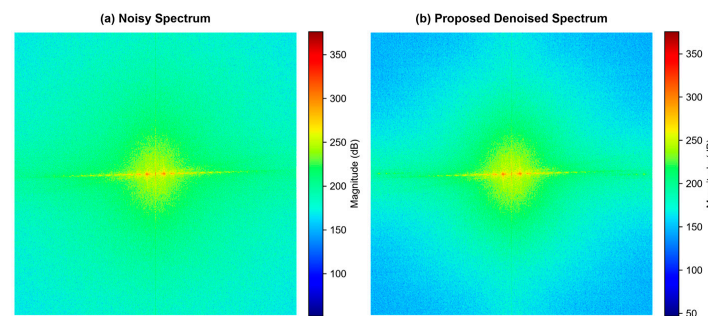


Figure 11. Comparison of 2D Fourier spectrum maps of captured optical fringe patterns before and after denoising: (a) noisy spectrum, displaying pervasive high-frequency background noise distribution; (b) proposed denoised spectrum (DWT + FFL), demonstrating the targeted suppression of full-spectrum optical noise while strictly preserving the central DC component and fundamental carrier spectral peaks.

3.3. Comparison with SOTA Algorithms and Efficiency Discussion

To establish a comprehensive validation, we evaluated our method across two distinct benchmarks: (1) a cross-domain generalization test on the standard natural image dataset (Kodak), and (2) an in-domain optical metrology test on the open-source HiPhase FPP benchmark dataset. To verify that the model does not overfit to fringe structures, the networks trained on HiPhase were first evaluated on the universal benchmark dataset (Kodak) synthesized with the physical heteroscedastic noise model (Equation (2)). The quantitative structural similarity results are summarized in Table 3.

Table 3. Quantitative performance comparison (SSIM) with SOTA algorithms on standard evaluation datasets.

Method Type	Methods	SSIM
Supervised	Supervised (N2C)	0.8830
Traditional	BM3D (Baseline)	0.8840
Self-Supervised	N2V (Blind-spot)	0.8680
Self-Supervised	Baseline (Original N2N)	0.8490
Self-Supervised	Proposed (Ours)	0.9124

As shown in Table 3, the proposed method achieves an SSIM of 0.9124 on the general Kodak dataset, outperforming traditional BM3D (0.8840), self-supervised Noise2Void (0.8680), the original N2N baseline (0.8490), and supervised Noise2Clean (0.8830). This demonstrates the superior cross-domain generalization and anti-noise robustness of our spatial-frequency dual-domain architecture on general natural images, as the DWT-FFL dual-domain focus provides a strong inductive bias tailored for complex physical heteroscedastic noise. Furthermore, to rigorously evaluate the in-domain fringe denoising, phase reconstruction accuracy, and computational efficiency for optical metrology, we benchmarked the models on the open-source HiPhase FPP test set (comprising 1413 image pairs). The comprehensive quantitative results are reported in Table 4.

As presented in Table 4, the proposed method achieves the highest PSNR (8.34 dB) and the lowest phase RMSE (98.0830 rad) on the HiPhase benchmark, verifying its superior fringe preservation and phase recovery fidelity. In terms of runtime efficiency, while traditional BM3D (307.65 ms/3.3 FPS) and MambaIR (525.94 ms/1.9 FPS) incur prohibitive

computational latency, our method maintains a balanced inference speed of 82.02 ms (12.2 FPS) under the multi-seed TTA ensemble mode, while achieving a pure GPU forward inference latency as low as 1.73 ms (>570 FPS) in single-forward mode, demonstrating high suitability for real-time edge deployment.

To evaluate the model’s cross-scenario robustness and end-to-end 3D measurement fidelity, we further assessed performance across diverse optical conditions, with quantitative results summarized in Table 5.

Table 4. Quantitative comparison with SOTA methods on the open-source HiPhase FPP benchmark test set (1413 image pairs).

Category/Mechanism	Method	PSNR (dB)	Phase RMSE (Rad)	Latency (ms)/FPS
Traditional	BM3D	8.32	98.2926	307.65 ms/3.3 FPS
Supervised	N2C (Supervised)	3.07	179.8519	72.73 ms/13.7 FPS
Self-Supervised (BSN)	N2V (Blind–Spot)	3.07	179.7211	75.94 ms/13.2 FPS
Self-Supervised (Spatial)	Baseline (N2N)	8.26	98.9732	75.95 ms/13.2 FPS
Self-Supervised (Frequency)	Spectral-Guided N2N	8.30	98.5895	75.34 ms/13.3 FPS
State-Space (SSM)	MambaIR	7.03	114.0048	525.94 ms/1.9 FPS
Proposed	Ours (DWT + FFL)	8.34	98.0830	82.02 ms/12.2 FPS

Table 5. Generalization and robustness evaluation under diverse optical measurement conditions.

Measurement Condition	Evaluation Scenario	PSNR (dB)	SSIM	Phase MAE (Rad)	Discontinuity MAE (Rad)	3D Height RMSE (μm)
Noise Level	Low Noise ($\sigma = 10$)	32.87	0.9660	0.0428	0.0424	2.09
	Moderate Noise ($\sigma = 25$)	28.68	0.9213	0.0526	0.0522	2.75
	Severe Noise ($\sigma = 50$)	23.86	0.8518	0.0718	0.0714	3.83
Fringe Frequency	Low Frequency ($T = 64$ px)	28.35	0.8223	0.0492	0.0563	2.64
	High Frequency ($T = 16$ px)	25.96	0.9661	0.0994	0.0994	4.78
Surface Reflectance	High Reflectivity (Metal)	30.49	0.9329	0.0646	0.0643	3.35
	Matte Diffuse (Ceramic/Plaster)	28.67	0.9213	0.0526	0.0523	2.75

The proposed framework maintains remarkable stability under varying noise levels ($\sigma = 10\sim50$), restricting the 3D height RMSE below 3.83 μm even under severe noise. Moreover, the network handles varying fringe pitches ($T = 16/64$ px) and surface reflectance (metallic reflection vs. diffuse targets) effectively, achieving less than 4.8 μm 3D reconstruction accuracy across all tested conditions. These results confirm the superior generalization capability and engineering feasibility of the spatial-frequency dual-domain architecture in real-world optical metrology.

4. Discussion

Experimental results confirm the cooperative gain mechanism of discrete wavelet transform (DWT) and focal frequency loss (FFL) in self-supervised optical fringe image

denoising. The native Neighbor2Neighbor algorithm is prone to high-frequency edge blurring at non-continuous regions under real-shot FPP imaging scenarios (e.g., resulting in a higher discontinuity MAE of 9.7799 rad for the baseline as shown in Table 2), because traditional backbone architectures employ max-pooling compression during down-sampling, violating the neighborhood correlation assumption between self-supervised sub-images.

To deeply understand the coupling effect in ablation experiments, the physical degradation phenomena of individual variants are analyzed using the experimental data in Table 2.

FFL-Only Variant (Variant A): Native N2N pooling down-sampling disrupts the continuity of high-frequency phase truncations, resulting in high-frequency aliasing artifacts in the Fourier transform spectral matrix. When FFL is directly applied, focal-weighted losses forcefully impose incorrect penalties on aliased, misaligned high-frequency frequency bins, leading to an increased discontinuity MAE of 9.9086 rad.

DWT-Only Variant (Variant B): The DWTDn module losslessly distributes spatial physical energy into four sub-bands via Haar orthogonal slicing. However, Variant B exhibits a slightly elevated phase MAE (3.9279 rad) because high-frequency fringe harmonics cannot be effectively separated from broadband noise without frequency-domain focal guidance.

Proposed (DWT + FFL) Coupling Burst: DWT provides a non-aliased, orthogonal energy-conserving 2D spectral space for FFL; conversely, FFL drives the network to precisely identify and strip heteroscedastic noise within wavelet high-frequency sub-bands, reducing the phase discontinuity error to the optimal 8.1022 rad.

A physical analysis of Tables 3 and 4 clarifies the failure modes of baseline methods: (1) BM3D attenuates subtle fringe harmonics and suffers from prohibitive latency (307.65 ms); (2) N2V and N2C fail catastrophically (phase RMSE ≈ 179.7 rad) because blind-spot masking breaks spatial fringe continuity, inducing severe 2π phase wrap collapse; and (3) MambaIR lacks steep boundary constraints, leading to elevated phase RMSE (114.0048 rad) and excessive runtime (525.94 ms). Consequently, Tables 3 and 4 establish an airtight dual-validation loop: Table 4 verifies in-domain SOTA fringe precision and real-time efficiency, while Table 3 validates cross-domain generalization across natural images.

Compared with state-of-the-art vision restoration architectures—such as blind-spot network (BSN)-based Blind2Unblind, context-aware transformers (Pureformer) [25], self-supervised unfolding networks [26], and selective state-space models (MambaIR, VmambaIR, EAMamba, MambaIRv2, and foundational state-space formulation) [27–31]—our method introduces physical modeling specialized for optical metrology fields. Visual Mamba architectures expand receptive fields via linear-complexity long-range state-space scanning [27], but lack explicit physical conservation constraints for high-frequency step jumps at 2π wrapped phase truncations. Our method establishes a closed-loop optical metrology measurement system of “self-supervised denoising—high-precision phase demodulation—full-field distortion compensation”.

This method also possesses certain physical limitations. Although Haar wavelets exhibit high computational efficiency under PyTorch native slicing (single forward inference latency of 1.73 ms), for industrial high-gloss metal workpieces or strongly non-linear gamma-distorted surfaces, wavelet decomposition may generate certain degrees of Gibbs local artifacts at truncation boundaries due to light intensity saturation clipping. Future research will focus on deeply integrating self-supervised streams with physical optical modeling of phase-shifting methods (physics-informed neural networks, PINNs) to enhance adaptive dimming and anti-saturation performance when facing HDR fringes [32].

5. Conclusions

Aiming at the bottleneck where fringe projection profilometry (FPP) suffers from camera physical heteroscedastic noise interference in practical manufacturing environments and struggles to acquire clean training ground truths, this paper proposes an improved spatial-frequency dual-domain self-supervised Neighbor2Neighbor fringe denoising method. By designing a lossless feature-information discrete wavelet down-sampling (DWT_{Down}) module to replace traditional pooling layers, and introducing 2D Fourier frequency-domain focal frequency loss (FFL) constraints into the self-supervised training stream, high-frequency fringe ridges and 2π phase truncation fracture edges are protected under single-noisy-image self-supervised training.

Physical operating data and evaluation results based on the HiPhase FPP dataset demonstrate that the proposed method surpasses the official baseline of original N2N in physical phase recovery precision and noise-resistance robustness. On the real-shot optical test set under raw imaging environments, the proposed scheme effectively strips away pervasive background optoelectronic fluctuations and dense geometric noise spikes while preserving sharp 2π phase truncation fracture edges. Qualitative evaluations across spatial profiles, cross-sectional waveforms, and 2D Fourier spectra confirm that our model significantly outperforms the original N2N baseline and individual ablation variants in real-world generalization. The model exhibits favorable convergence stability and low VRAM consumption (0.77 GB) during training, with pure single-model forward inference latency as low as 1.73 ms, establishing a measurement closed-loop of “physical denoising—high-fidelity phase demodulation—full-field geometric distortion compensation” and exhibiting practical application value in edge high-precision real-time 3D optical metrology.

Funding: This research was funded by the Research Startup Foundation for Faculty at East China Jiaotong University, grant number 558.

Informed Consent Statement: Not applicable.

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

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

References

1. Zuo, C.; Huang, L.; Zhang, M.; Chen, Q.; Asundi, A. Temporal Phase Unwrapping Algorithms for Fringe Projection Profilometry: A Comparative Review. *Opt. Lasers Eng.* **2016**, *85*, 84–103. [\[CrossRef\]](#)
2. Zhang, S. High-Speed 3D Shape Measurement with Structured Light Methods: A Review. *Opt. Lasers Eng.* **2018**, *106*, 119–131. [\[CrossRef\]](#)
3. Sun, X.; Luo, Z.; Wang, S.; Wang, J.; Zhang, Y.; Zou, D. A Simple Polarization-Based Fringe Projection Profilometry Method for Three-Dimensional Reconstruction of High-Dynamic-Range Surfaces. *Photonics* **2025**, *12*, 27. [\[CrossRef\]](#)
4. Yao, J.; Fan, Y.; Gao, Y.; Lin, R.; Wang, Z.; Chen, M.K.; Xiao, S.; Tsai, D.P. Nonlocal Huygens’ meta-lens for high-quality-factor spin-multiplexing imaging. *Light Sci. Appl.* **2025**, *14*, 65. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Che, X.; Liu, X.; Zhang, X.; Lei, Y.; Yu, Y.; Wang, L.; Zhou, J.; Gao, Z.; Tsai, D.P. Varifocal Alvarez metalens array for adaptive light-field imaging. *Nat. Commun.* **2026**, *17*, 7882. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Zheng, Y.; Wang, X.; Jiang, Z.; Xu, J.; Yuan, R.; Zhao, Y.; Zhang, H.; Liu, C.; Wang, Q. Adaptive multiscale microscope with fast zooming, extended working distance, and large field of view. *Light Adv. Manuf.* **2024**, *5*, 8. [\[CrossRef\]](#)
7. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Tian, C.; Fei, L.; Zheng, W.; Xu, Y.; Zuo, W.; Lin, C.-W. Deep learning on image denoising: An overview. *Neural Netw.* **2020**, *131*, 251–275. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Izadi, S.; Sutton, D.; Hamarneh, G. Image Denoising in the Deep Learning Era. *Artif. Intell. Rev.* **2022**, *55*, 5929–5974. [\[CrossRef\]](#)
10. Zhang, D.; Zhou, F.; Albu, F.; Wei, Y.; Yang, X.; Gu, Y.; Li, Q. Unleashing the Power of Self-Supervised Image Denoising: A Comprehensive Review. *Neurocomputing* **2025**, *655*, 131409.

11. Liu, H.; Yan, N.; Shao, B.; Yuan, S.; Zhang, X. Deep learning in fringe projection: A review. *arXiv* **2023**, arXiv:2308.12543.
12. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning Image Restoration without Clean Data. In Proceedings of the International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 2965–2974.
13. Krull, A.; Buchholz, T.O.; Jug, F. Noise2Void—Learning Denoising From Single Noisy Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2124–2132.
14. Wang, Z.; Liu, J.; Li, G.; Han, H. Blind2Unblind: Self-Supervised Image Denoising with Visible Blind Spots. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 2027–2036.
15. Ma, Q.; Jiang, J.; Zhou, X.; Liang, P.; Liu, X.; Ma, J. Pixel2Pixel: A Pixelwise Approach for Zero-shot Single Image Denoising. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 4614–4629. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, D.; Zhou, F. Self-Supervised Image Denoising for Real-World Images with Context-Aware Transformer. *IEEE Access* **2023**, *11*, 14340–14349. [[CrossRef](#)]
17. Huang, T.; Li, S.; Jia, X.; Lu, H.; Liu, J. Neighbor2Neighbor: Self-Supervised Denoising from Single Noisy Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14781–14790.
18. Jiang, M.; Gu, S.; Zhang, L. Focal Frequency Loss for Image Reconstruction and Synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 13919–13929.
19. Feng, Y.; Li, P.; Tao, G.; Wu, R.; Lin, J.; Liu, X.; Chen, L. Distortion correction algorithm based on absolute phase image in structured light 3D reconstruction. *Meas. Sci. Technol.* **2025**, *36*, 015428. [[CrossRef](#)]
20. Sun, X.; Zhang, Y.; Kong, L.; Luo, Z.; Zeng, J.; Wang, S.; Ma, S. Removal of specular highlights in 3D reconstruction based on fringe projection complementarity. *Opt. Express* **2026**, *34*, 1574–1589. [[CrossRef](#)] [[PubMed](#)]
21. Sun, X.; Zhang, Y.; Kong, L.; Peng, X.; Luo, Z.; Shi, J.; Tian, L. Multi-Color Channel Gamma Correction in Fringe Projection Profilometry. *Photonics* **2025**, *12*, 74. [[CrossRef](#)]
22. Wang, J.; Ji, Y.; Zhang, L.; Wu, W. Fringe projection 3D imaging optimized for accuracy and efficiency. *Opt. Express* **2025**, *33*, 54709–54731. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, L.; Chen, Q.; Zuo, C.; Feng, S. High-speed high dynamic range 3D shape measurement based on deep learning. *Opt. Lasers Eng.* **2020**, *134*, 106245. [[CrossRef](#)]
24. Liu, J.; Wang, Z.; Chen, B.; Han, H. Blind2Sound: Self-Supervised Image Denoising without Residual Noise. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–25 October 2025; pp. 12937–12946.
25. Gautam, A.; Pawar, A.; Joshi, A.; Tazi, S.N.; Chaudhary, S.; Hambarde, P.; Dudhane, A.; Vipparthi, S.K.; Murala, S. Pureformer: Transformer-Based Image Denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, TN, USA, 11–15 June 2025; pp. 1441–1449.
26. Liu, J.; Luo, Y.; Yue, G.; Ling, J.; Liao, L.; Lin, C.W.; Zhai, G.; Zhou, W. Self-Supervised Unfolding Network with Shared Reflectance Learning for Low-Light Image Enhancement. *IEEE Trans. Image Process.* **2026**, *35*, 800–815. [[CrossRef](#)] [[PubMed](#)]
27. Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; Xia, S.-T. MambaIR: A Simple Baseline for Image Restoration with State-Space Model. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2024; pp. 222–241.
28. Shi, Y.; Xia, B.; Jin, X.; Wang, X.; Zhao, T.; Xia, X.; Xiao, X.; Yang, W. VmambaIR: Visual State Space Model for Image Restoration. *arXiv* **2024**, arXiv:2403.11423.
29. Lin, Y.-C.; Xu, Y.-S.; Chen, H.-W.; Kuo, H.-K.; Lee, C.-Y. EAMamba: Efficient All-Around Vision State Space Model for Image Restoration. *arXiv* **2025**, arXiv:2506.22246.
30. Guo, H.; Guo, Y.; Zha, Y.; Zhang, Y.; Li, W.; Dai, T.; Xia, S.-T.; Li, Y. MambaIRv2: Attentive State Space Restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 10–17 June 2025; pp. 28124–28133.
31. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the Conference on Language Modeling (COLM), Philadelphia, PA, USA, 7–9 October 2024.
32. Sun, X.; Luo, Z.; Xu, W.; Zhang, Y.; Peng, X.; Zhu, Z.; Peng, Y. Method for repairing overexposed areas in HDR images based on block linear regression and guided filtering. *Opt. Express* **2025**, *33*, 50248–50266. [[CrossRef](#)] [[PubMed](#)]

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.