Learning-Based Modulation Enhancement for High-Dynamic-Range Fringe Projection Profilometry

Kejiang Chen[1], Pei Zhou[1], Jiangping Zhu[1, 2*]
[1]College of Computer Science, Sichuan University, Chengdu, China
[2]School of Information Science and Technology, Tibet University, Lhasa China

***Abstract*—Fringe Projection Profilometry (FPP) is a key technique for optical 3D surface measurement, yet it remains highly sensitive to high dynamic range (HDR) reflective surfaces, such as metallic objects. Strong specular reflections often lead to overexposure, while insufficient illumination causes underexposure, resulting in degraded fringe modulation, phase distortions, and incomplete 3D reconstructions. These issues severely limit the accuracy and reliability of FPP in industrial scenarios. Existing solutions, including surface treatment and multi-exposure HDR imaging, are either invasive or computationally expensive, restricting their practical deployment. To overcome these challenges, we propose a learning-based modulation enhancement network that explicitly models and corrects modulation distortions in HDR fringe patterns. The proposed framework introduces a Modulation Offset Estimation (MOE) module to accurately estimate modulation offsets in over- and under-exposed regions, followed by a Modulation Intensity Fusion (MIF) module that adaptively integrates the corrected modulation features to generate enhanced fringe images. This design enables effective restoration of fringe modulation while preserving spatial consistency across HDR regions.**

Extensive experiments demonstrate that the proposed method significantly improves reconstruction accuracy and robustness, especially under challenging HDR conditions. For standard block reconstruction, the proposed method achieves an MAE of 0.0030 mm and an SD of 0.0038 mm, which are comparable to those of the multi-exposure fusion method (MAE = 0.0029 mm, SD = 0.0036 mm). These results indicate that the proposed method offers a practical, non-invasive, and efficient solution for high-precision three-dimensional measurement of highly reflective surfaces.

***Index Terms*—deep learning, modulation enhancement, HDR surface, fringe projection profilometry.**

I. Introduction

Fringe Projection Profilometry (FPP) has been widely applied in 3D reconstruction [1], defect detection [2], and visual measurement [3]–[5]. However, its measurement accuracy is affected by high dynamic range (HDR) issues. To address the HDR problem, researchers have proposed various solutions including spraying diffuse reflective materials [6], employing polarizers [7], [8], or using photometric stereo [9], [10]. However, these methods increase the complexity of the measurement system [11], [12].

Several researchers have proposed multi-exposure fusion. Multi-exposure fusion methods capture multiple sets of fringe patterns under different exposure times and fuse them into a composite pattern. Zhang et al. [13] developed HDRS algorithm which synthesizes high-quality fringe patterns by integrating sequences captured at varying exposure levels. Feng et al. [14] proposed a grayscale histogram-based method for optimal exposure pixel fusion. Du et al. [15] introduced a phase-domain information fusion method that eliminates abnormal phase points in phase domain to fuse phase maps. Although multi-exposure fusion achieves relatively superior reconstruction quality, its efficiency is low.

Deep learning has also been applied to FPP. Sam Van Der Jeught et al. [16] demonstrated the feasibility of CNN-based FPP methods. Nguyen et al. [17] further validated the strong capability of U-Net [18] encoder–decoder architectures for single-frame depth estimation. Based on this architecture, several improved variants have been proposed, such as ResUNet [19], AttentionUNet [20], and RAUNet [21], which enhance feature representation by introducing residual connections and attention mechanisms. These designs improve gradient propagation and enable the network to focus on informative regions, and have shown effectiveness in depth estimation and fringe-related tasks under complex imaging conditions. Cheng et al. [22] developed a method based on modulation analysis and single-frame networks. Zhu et al. [23] incorporated the discrete wavelet transform into the deep learning framework. Song et al. [24] proposed a phase demodulation method based on a full-scale connection network, SE-FSCNet. While these methods achieve promising reconstruction results under standard exposure conditions, they generally lack mechanisms to explicitly handle HDR-induced artifacts in fringe patterns, such as local over- or under-exposure, and may fail to fully recover the modulation offsets across different intensity regions.

To enable the network to effectively restore problematic regions, this paper proposes a dual-input modulation enhancement network that improves HDR fringe patterns quality inspired by Li's [25] method of offset estimation and correction. The network first employs UNet to extract brightness and darkness feature maps from input patterns, subsequently generating pseudo modulation features. Modulation offset estimation (MOE) module then estimates and corrects the offsets between the bright-dark modulation feature maps and the pseudo modulation features. Finally, modulation intensity

*Corresponding author: zjp16@scu.edu.cn

fusion (MIF) module optimizes and integrates these modulation offsets to produce the enhanced fringe patterns. Our main contributions are as follows:

1) A novel neural network method is proposed to enhance images with both overexposed and underexposed regions by modeling modulation intensity distribution shifts.
2) Two new modules are introduced:
   - Modulation Offset Estimation (MOE) module: Responsible for estimating and correcting modulation offsets in overexposed and underexposed regions.
   - Modulation Intensity Fusion (MIF) module: Adjusts the modulation offsets in overexposed and underexposed regions to generate enhanced fringe images.

II. METHODS

A. Principle of Phase-shifting Fringe

A typical FPP system consists of a projector and a camera. Projector projects phase-shifting fringe patterns onto the target object surface, and the camera captures the deformed fringe patterns which are then processed to reconstruct the 3D point cloud. The fringe patterns projected is

$$I_i^p(x_p, y_p) = A^p + B^p cos(\frac{2x_p\pi}{T} - \frac{2\pi(i-1)}{N}), \quad (1)$$

and deformed fringe pattern captured is

$$I_i^c(x_c, y_c) = A^c + B^c cos(\phi(x_c, y_c) - \frac{2\pi(i-1)}{N}). \quad (2)$$

Where (x_p, y_p) and (x_c, y_c) denote the coordinates in projector and camera, respectively. $I_i^p(x_p, y_p)$ and $I_i^c(x_c, y_c)$ represent the intensity in i-th ($i = 1, 2, \ldots, N$) fringe image, A^p and A^c denotes the background intensity, T is the width of one period. In this study, the four-step phase shifting method ($N = 4$) is adopted, with a fringe frequency of 64 ($T = 64$). B^p and B^c represent the modulation intensity, $\phi(x_c, y_c)$ indicates the wrapped phase to be solved, which is calculated by

$$\phi(x_c, y_c) = arctan\left[\frac{\sum_{i=1}^{N} I_i^c(x_c, y_c) sin\frac{2\pi(i-1)}{N}}{\sum_{i=1}^{N} I_i^c(x_c, y_c) cos\frac{2\pi(i-1)}{N}}\right]. \quad (3)$$

After calculating the wrapped phase $\phi(x_c, y_c)$, the absolute phase is computed by Zhu's [26] method, followed by the reconstruction of the object surface employing Zhou's [27] phase-height mapping model.

B. Datasets Generation

This paper employs digital twin [28] to generate training datasets in Blender. Fig.1(a) shows the real measurement system. The system consists of a projector (DLP LightCrafter 4500) with a resolution of 912 × 1140 and a CMOS camera (MER2-502-79U3M) with a resolution of 2448 × 2048. Fig.1(b) shows the simulation system. The simulation system consists of a camera and a projector, with a camera focal length of 110 mm and an angle of 22.5° between the camera and the projector, consistent with the real measurement system. The 3D models are sourced from the Thingi10K [29] dataset which contains 10,000 different types of 3D models. Fig. 1(c) shows some input data and corresponding labels in our dataset. We generated 3200 sets of training data and divided them into training, validation, and test sets with a ratio of 8:1:1.

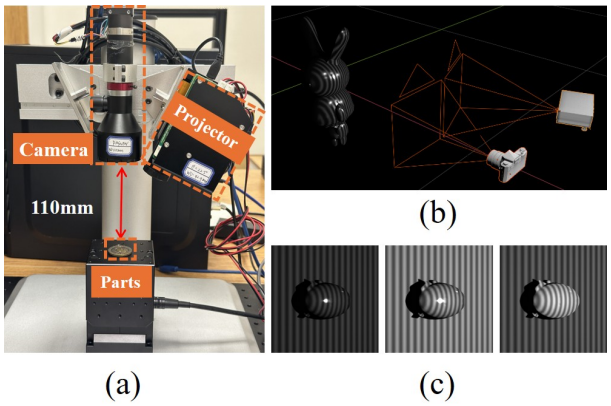

Fig. 1. (a) The real measurement system. (b) Side view of the scene layout in Blender. (c) Dataset example, from left to right: low exposure input, high exposure input, and label.

C. Network Structure

In the modulation enhancement network shown in Fig. 2(a), HDR image sequences captured at lower and higher exposure times, I_u and I_o, are input separately. Subsequently, I_u and I_o are inputted into UNet to extract their corresponding background intensities A_u and A_o, respectively. Then, the brightening modulation feature sequence B_b and darkening modulation feature sequence B_d corresponding to I_u and I_o are calculated as follows:

$$B_b = \frac{I_u}{A_u} = \frac{I_u}{f(I_u)}, \quad (4)$$

$$B_d = 1 - \frac{I_o}{A_o} = 1 - \frac{I_o}{f(I_o)}, \quad (5)$$

where $f(\cdot)$ represents the background intensity feature map extracted by UNet. To estimate the modulation offset of the HDR fringe images, the pseudo-modulation intensity enhancement is first performed by combining B_b, B_d with I_u, I_o, generating the pseudo-modulation intensity-enhanced image sequence I_p, as follows:

$$I_p = w_u \cdot I_u + w_o \cdot I_o + w_b \cdot B_b + w_d \cdot B_d, \quad (6)$$

where w_u, w_o, w_b, and w_d are learnable parameters. Then, I_p will be used as reference information, and two MOE modules will be employed to guide the generation of the modulation offset feature sequences O_b between B_b and I_p (as well as O_d between B_d and I_p), as follows:

$$O_b = s(B_b, I_p), \quad (7)$$

$$O_d = s(B_d, I_p), \quad (8)$$

where $s(\cdot)$ represents the modulation offset estimation mapping. Finally, O_b, O_d combined with I_p are input into the MIF

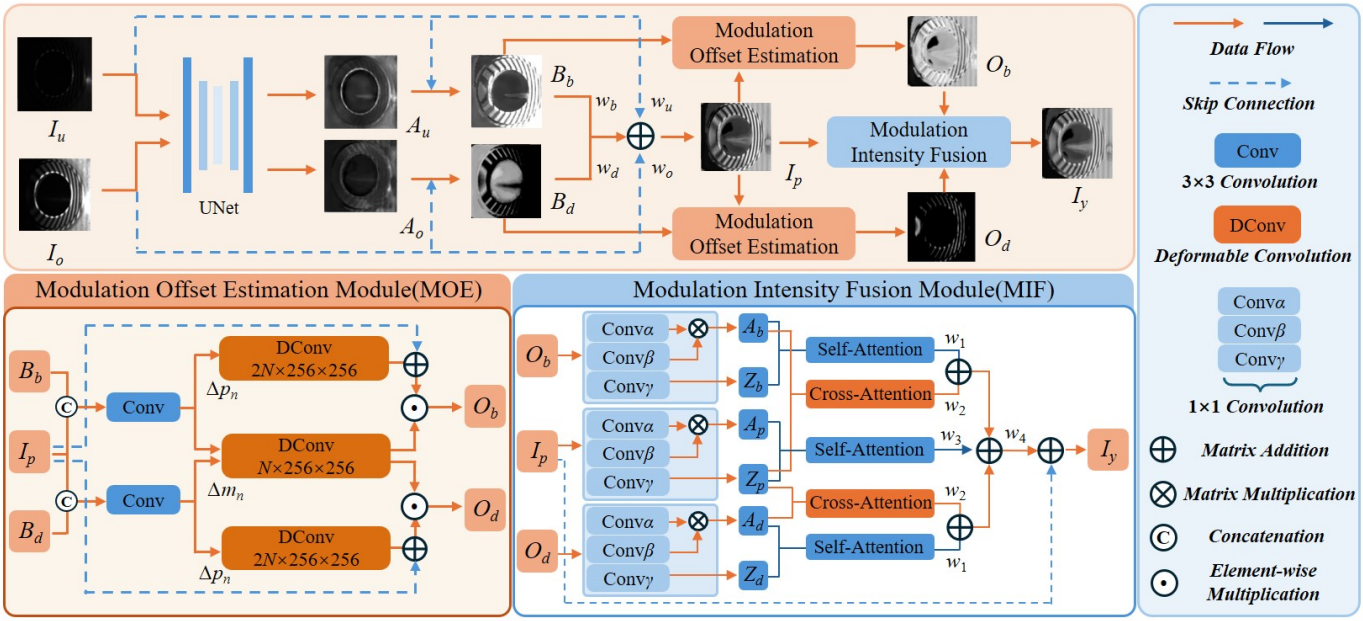

Fig. 2. (a) Architecture of the network. UNet extracts the bright-dark features A_o/A_u from I_o/I_u, based on which the modulation feature B_d/B_b are computed. Then I_o, I_u, B_d and B_b are fused to obtain pseudo modulation sequence I_p. I_p and B_b/B_d are fed into the MOE to derive the modulation offset O_b/O_d. Finally, MIF processes O_b, O_d, and I_p by attention mechanism to generate enhanced pattern I_y. (b) Structure of MOE module. The MOE module generates modulation offset feature sequences O_b and O_d by establishing a modulation offset model for HDR images. (c) Structure of MIF module. The MIF module ultimately generates enhanced fringe patterns through attention mechanism and feature fusion.

module to correct the modulation intensity offsets in different regions, ultimately yielding the modulation-enhanced I_y, as follows:

$$I_y = c(I_p, O_b, O_d), \tag{9}$$

where $c(\cdot)$ represents the modulation intensity fusion mapping.

1) Modulation Offset Estimation Module: The MOE module estimates local modulation offsets and spatial offsets by comparing the bright-dark features with the pseudo-modulation reference. To accommodate the local spatial variations of modulation information and the changes in modulation intensity within fringe patterns, the MOE module employs a deformable convolution that is scalable to the modulation intensity space. The module introduces spatial offsets and modulation offsets into the standard convolution, enabling the convolution kernel to adaptively adjust both position and modulation intensity amplitude within an $N \times N$ pixel region, effectively capturing the complex variations of modulation offset in fringe images. Specifically, the MOE module first concatenates the pseudo-modulation intensity enhanced image sequence I_p and the brightening/darkening modulation feature sequences B_b/B_d. Then, two independent 3×3 convolution kernels are used to extract the spatial offset Δp_n and the modulation offset Δm_n. Therefore, the deformable convolution calculation process combining the spatial domain and modulation intensity space can be expressed as [30]:

$$y = \sum_{p_n \in \mathcal{R}} \omega_n \cdot x(p_0 + p_n + \Delta p_n) \cdot \Delta m_n. \tag{10}$$

Where $\mathcal{R} = \{(-1,-1),(-1,0),...,(1,1)\}$ represents the coordinates of a 3×3 square grid. n is the enumerator for the elements in $\mathcal{R}$ indicating the n-th position, and N is the length of $\mathcal{R}$ ($N = 9$ for a regular 3×3 kernel). x is input for the convolution operation. p_0, p_n, and Δp_n are 2D variables representing spatial positions, ω_n represents the learnable parameter of the convolution kernel, y is the output of the deformable convolution operation, including O_b and O_d. By explicitly modeling both spatial and modulation deviations, the MOE module ensures accurate local compensation.

2) Modulation Intensity Fusion Module: The MIF module combines these modulation offset features with HDR fringe images through attention operations, effectively correcting the offset in over-exposure and under-exposure regions. The MIF module takes O_b, O_d and I_p as input and allocates independent attention mapping paths to effectively fuse them. The attention mechanism extracts the attention matrices A_i and the corresponding weight matrices Z_i ($i \in \{b, p, d\}$) for each input sequence, allowing the module to selectively enhance features that are consistent across input sequences while suppressing noise and redundant information. Each attention mapping path contains three independent convolution kernels (denoted as Conv_α, Conv_β, and Conv_γ). The fusion is performed as:

$$f_j = w_1 \cdot A_j \otimes Z_j + w_2 \cdot A_j \otimes Z_p, \tag{11}$$

where, $j \in \{b, d\}$, w_1 and w_2 are learnable parameters, $\otimes$ represents matrix multiplication, f_j represents the process of fusion. Finally, f_b and f_d are combined with I_p to fuse and

obtain the modulation-enhanced fringe image sequence I_y, as follows:

$$I_y = w_4 \cdot (f_b + f_d + w_3 \cdot A_p \otimes Z_p) + I_p, \tag{12}$$

where w_3 and w_4 are learnable parameters. By leveraging self-attention and cross-attention mechanism, the MIF module effectively aligns and fuses local modulation patterns, enhancing HDR regions while preserving structural consistency.

D. Loss Function

The overall loss function L is composed of two terms, L_p and L_y, which supervise the generation of the pseudo-modulation sequence I_p and the final modulation-enhanced sequence I_y, respectively. The first term, L_p , encourages I_p to approximate the ground truth gt and is defined as the L_1 loss:

$$L_p = \|I_p - gt\|_1. \tag{13}$$

For the modulation-enhanced output I_y, we adopt a combination of three complementary loss functions L_y to ensure both local fidelity and global structural consistency. The L_1 loss enforces pixel-wise accuracy, the cosine similarity loss L_{cos} preserves the relative intensity distribution, and the structural similarity loss L_{ssim} [31] encourages the reconstruction of local textures and contrast:

$$L_1 = \|I_y - gt\|_1, \quad L_{cos} = 1 - \frac{I_y \cdot gt}{\|I_y\|_2 \|gt\|_2}, \tag{14}$$

$$L_{ssim} = 1 - \frac{(2\mu_y\mu_{gt} + C_1)(2\sigma_{ygt} + C_2)}{(\mu_y^2 + \mu_{gt}^2 + C_1)(\sigma_y^2 + \sigma_{gt}^2 + C_2)}, \tag{15}$$

where μ_y and μ_{gt} denote the mean values of I_y and gt, σ_y and σ_{gt} denote their variances, and σ_{ygt} denotes the covariance. These three losses are combined as a weighted sum:

$$L_y = \lambda_1 L_1 + \lambda_{cos} L_{cos} + \lambda_{ssim} L_{ssim}, \tag{16}$$

with hyper-parameters $\lambda_1 = 1$, $\lambda_{cos} = 0.5$, and $\lambda_{ssim} = 1$. The overall training objective is then expressed as:

$$L = \lambda_p L_p + \lambda_y L_y, \tag{17}$$

where λ_p and λ_y set to 0.5 which balance the supervision of the pseudo-modulation generation and the final modulation-enhanced output. By combining these losses, the network is guided to produce outputs that are accurate at the pixel level, maintain consistent intensity distributions, and preserve local structural details.

III. EXPERIMENTS

Our experiments are conducted on one NVIDIA GeForce RTX 4090 GPU using PyTorch version 2.6.0. We use an Adam optimizer with an initial learning rate of 1e-4 and train the models for 200 epochs. We adopt mini-batch training with a batch size of 2. The initial learning rate is set to 1e-4, and the Adam optimizer is employed to guide gradient-based updates. To further improve convergence and model performance, a learning rate annealing strategy is applied during training, dynamically adjusting the learning rate at different stages. This approach helps prevent premature convergence and reduces the likelihood of getting trapped in local minima.

A. Phase Parsing Validation

We conducted an analysis of the intermediate results within the reconstruction pipeline. As illustrated in Fig. 3, the proposed network demonstrates significant restoration capabilities in both under-exposed and over-exposed regions. The model can accurately distinguish and enhance HDR regions, demonstrating strong discriminative capability in extracting features from different regions, including over-exposed, under-exposed, and normally exposed areas. As shown in Fig. 3(a–c), the proposed method enhances the grayscale intensity in under-exposed regions, while suppressing over-exposed regions. Meanwhile, the missing fringe information is effectively restored, bringing the global fringe intensity back to a reference exposure range and improving the modulation in HDR regions. To further analyze the enhancement effect, a row-wise grayscale profile is extracted from the high-exposure image and the corresponding enhanced image, as illustrated in Fig. 3(d). The blue curve represents the grayscale cross-section of the original image traversing the HDR region, whereas the orange curve corresponds to the enhanced image. The results show that the original sinusoidal fringe distribution is highly uneven, exhibits abrupt variations, and is contaminated by noise. After processing with the proposed method, the fringe intensity is restored to a more reasonable range, and the profile becomes significantly smoother, indicating that the sinusoidal characteristics of the fringe pattern are effectively recovered with improved noise robustness. Figure 3(e) presents the absolute phase map. Except for shadow regions caused by object occlusion, all other areas, including HDR regions, are correctly unwrapped with a smooth and monotonic phase variation. This confirms that no phase unwrapping errors are introduced in the subsequent reconstruction process. Experimental results across different cases demonstrate that the proposed method is capable of repairing HDR fringe patterns while providing effective modulation compensation and correction.

B. Comparative Experiments

For comparative experiments, we selected several representative methods, including UNet [18], RAUNet [21], SE-FSCNet [24], single-exposure approach (SPF) [32], and multi-exposure fusion (MEF) method [15], in order to evaluate reconstruction performance. In this subsection, a coin, a thread, and a rivet are selected as the test objects for comparative experiments. The coin surface exhibits a non-uniform reflectivity distribution, where high and low reflectivity regions are randomly interleaved across complex curved surfaces. For the thread, the reflectivity is more spatially concentrated, primarily at the junctions between threads, and is characterized by a highly complex surface geometry. The rivet features a relatively simple geometry, with high and low reflectivity

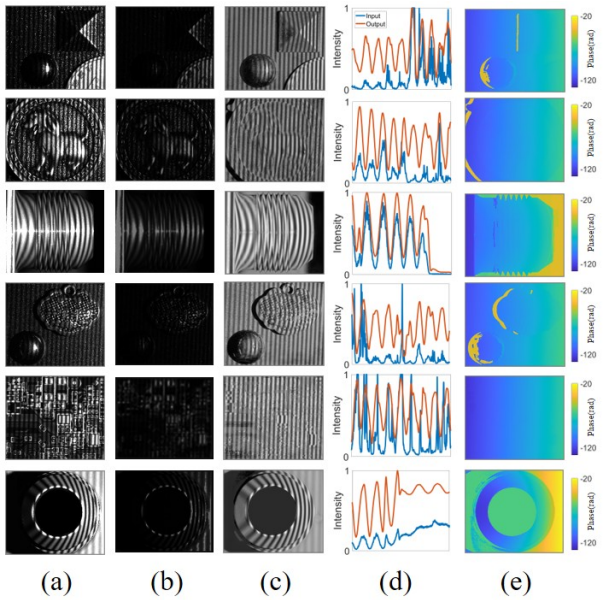

Fig. 3. (a) Over-exposure image. (b) Under-exposure image. (c) Enhanced fringe image. (d) A comparison of the grayscale values of a certain row between the enhanced image and the overexposed image. (e) The corresponding absolute phase.

regions uniformly distributed across different sloped planes. These three types of components facilitate a comprehensive evaluation of the modeling capabilities of various methods under diverse HDR scenarios, ranging from simple to complex geometries and from uniform to non-uniform reflectivity distributions.

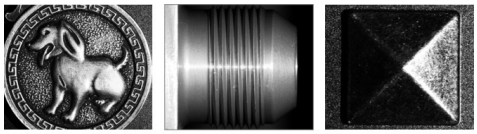

Fig. 4. Representative test objects for comparative experiments: (a) A coin with non-uniform reflectivity distribution; (b) A thread with complex HDR surfaces; (c) A rivet with uniform reflectivity distribution.

1) Reconstruction Point Cloud Comparison: We conducted reconstruction comparisons on coin, thread, and rivet regions, which involve HDR surfaces and complex geometries. Since ground truth is unavailable for real-world measurements, the results of MEF [15] are used as a reference for comparison. Fig. 5 shows the reconstruction results of the comparison methods. In the coin region, UNet, SE-FSCNet, and RAUNet exhibit noticeable artifacts, mainly in the form of periodic fluctuations in the reconstructed point clouds, indicating limitations in handling HDR scenes. The SPF method demonstrates high fidelity with the reference in well-exposed regions, but in low-light or high-light areas, the point cloud is incomplete due to low signal-to-noise ratio and insufficient modulation. By contrast, the proposed method produces smoother reconstructions with improved consistency. For thread part, the compared methods often suffer from trailing or streaking effects, resulting in blurred geometry, whereas the proposed method preserves detailed thread profiles. Similar periodic oscillations are also observed in the rivet region for the compared methods, while the proposed method achieves a surface topology that is closer to the reference.

The results demonstrate that single-exposure methods are inherently limited in HDR scenarios, as a fixed exposure setting cannot simultaneously preserve proper exposure across all regions, thereby restricting their reconstruction performance on HDR surfaces. Although deep learning–based methods improve robustness to a certain extent, the compared approaches do not sufficiently incorporate the physical characteristics of fringe patterns for targeted restoration. Their predominantly data-driven learning strategy leads to limited generalization when facing severe exposure variations, which manifests as periodic oscillations and instability in the reconstructed point clouds. For complex surface structures, such as thread part, the lack of explicit constraints on the fidelity between the restored and original fringe images further degrades reconstruction quality. As a result, artifacts, streaking effects, and blurred geometric details are observed in the final point clouds, indicating that the restored fringe patterns are not fully consistent with the underlying physical imaging process. In contrast, the proposed method explicitly exploits fringe modulation as a physically meaningful attribute. By first estimating modulation deviations in HDR regions and subsequently performing targeted modulation correction, the proposed approach effectively compensates for exposure-induced degradation. In addition, multiple skip connections are employed to fuse the original fringe images with pseudo-modulation feature maps at different scales, which facilitates accurate fringe restoration while preserving structural and intensity consistency with the source images. Consequently, the proposed method achieves high-fidelity reconstruction in HDR regions and improves the stability and reliability of the final 3D point clouds.

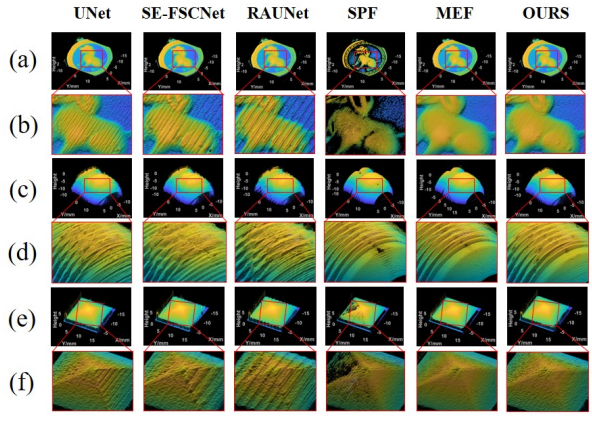

Fig. 5. Reconstruction results comparison, where (a), (c), and (e) show the full reconstructions of coin, thread, and rivet, respectively, and (b), (d), and (f) present magnified views of the HDR regions in the corresponding objects.

2) Reconstruction Error Comparison: To validate the reconstruction performance of the proposed method, error maps are visualized for deep learning–based comparative methods. For the coin error map (Fig. 6), the errors in RAUNet and SE-FSCNet exhibit pronounced periodic fluctuations, while UNet shows lower overall error, though some regions still have significant inaccuracies. In contrast, the proposed method achieves lower overall and local errors, resulting in higher reconstruction accuracy. For the thread part error map (Fig. 7), the comparison methods show errors exceeding 0.2 mm at the thread regions, indicating insufficient reconstruction capability for complex surfaces, leading to substantial errors. In contrast, the proposed method performs well in both smooth and complex surface areas. For the rivet error map (Fig. 8), UNet and SE-FSCNet show smaller reconstruction errors in low-intensity regions with less fluctuation, but in high-intensity regions, the errors significantly increase and exhibit periodic large errors. RAUNet also shows periodic large errors globally. The proposed method, however, maintains low errors in both bright and dark regions without periodic fluctuations. This suggests that even for simple HDR surfaces, comparison methods have limitations, while the proposed method exhibits stronger generalization and robustness for both simple and complex HDR surfaces, leading to better reconstruction performance. This improvement is attributed to MOE and MIF modules: MOE accurately estimates modulation offsets, and MIF corrects them, with both modules working synergistically to ensure higher fidelity in the reconstruction results.

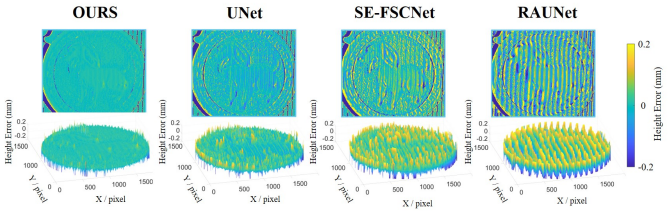

Fig. 6. Visualization of reconstruction errors for the coin. The color scale for error is set within the range of -0.2 to 0.2 mm.

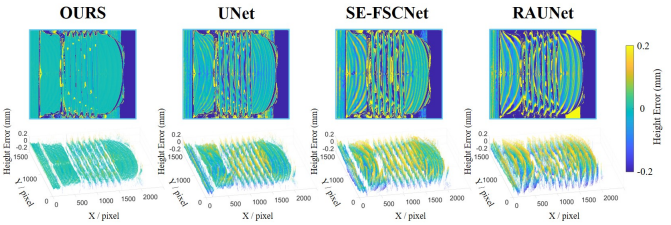

Fig. 7. Visualization of reconstruction errors for the thread. The color scale for error is set within the range of -0.2 to 0.2 mm.

To quantitatively assess the performance, we used Mean Absolute Error (MAE), Mean Relative Error (MRE), Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR) and Standard Deviation (SD) metrics, as summarized in Table I. Our method consistently outperforms existing techniques, particularly in SSIM and PSNR, demonstrating superior structural preservation and sharpness. The lower MAE and MRE values indicate that our method produces more accurate reconstructions, especially in areas with fine details like threads and rivet. These improvements are attributed to our network architecture. The MOE module adapts to spatial and modulation intensity variations in HDR fringe images through deformable convolutions, which reduces artifacts such as trailing and oscillation seen in other methods. The MIF module then refines the image by fusing modulation features, enhancing the reconstruction's accuracy and sharpness, as reflected in the higher PSNR values. In conclusion, the proposed method's superior performance in all metrics is driven by the effective handling of HDR challenges.

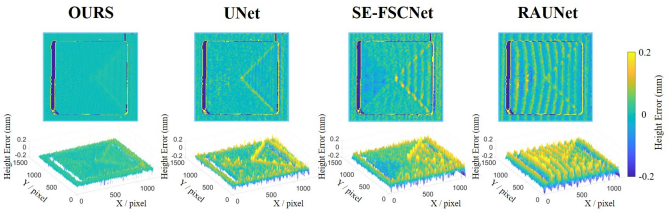

Fig. 8. Visualization of reconstruction errors for the rivet. The color scale for error is set within the range of -0.2 to 0.2 mm.

TABLE I
QUANTITATIVE COMPARISON OF RECONSTRUCTION ACCURACY FOR ALL METHODS, EVALUATED USING MAE, MRE, SSIM, PSNR, AND SD METRICS ON COIN, THREAD, AND RIVET PARTS.

		UNet	SE-FSCNet	RAUNet	OURS
Coin	SSIM	0.9742	0.9714	0.9638	**0.9814**
	MRE(mm)	0.0042	0.0068	0.0109	**0.0021**
	MAE(mm)	0.1863	0.0303	0.0488	**0.0095**
	SD(mm)	0.0296	0.0447	0.0631	**0.0155**
	PSNR(dB)	17.9872	16.2503	15.5916	**22.7045**
Thread Part	SSIM	0.9055	0.8848	0.8680	**0.9491**
	MRE(mm)	0.0118	0.0159	0.0214	**0.0089**
	MAE(mm)	0.1088	0.1470	0.1736	**0.0823**
	SD(mm)	0.0389	0.0517	0.0593	**0.0161**
	PSNR(dB)	13.7251	13.1645	11.8749	**17.6533**
Rivet	SSIM	0.9668	0.9642	0.9556	**0.9821**
	MRE(mm)	0.0022	0.0036	0.0046	**0.0015**
	MAE(mm)	0.0124	0.0196	0.0256	**0.0082**
	SD(mm)	0.0236	0.0359	0.0450	**0.0148**
	PSNR(dB)	22.1925	21.9800	21.4252	**23.1578**

C. Comparison of Reconstruction Accuracy for Standard Parts

To quantitatively evaluate the reconstruction accuracy of different methods, experiments were conducted on a standard metallic gauge block with a height of 1 mm and a tolerance of $\pm 1 \mu m$, as shown in Fig. 9. The surface of the gauge block exhibits high reflectivity. The reconstructed point clouds are illustrated in Fig. 10. The compared deep learning–based methods still suffer from periodic fluctuations in the reconstructed point clouds to varying degrees. The

single-exposure method yields results comparable to multi-exposure fusion in normally exposed regions; however, in underexposed and overexposed areas, it is affected by HDR issues and also exhibits periodic point cloud oscillations. In contrast, the proposed method produces a more complete reconstruction, closely matching the results obtained by multi-exposure fusion.

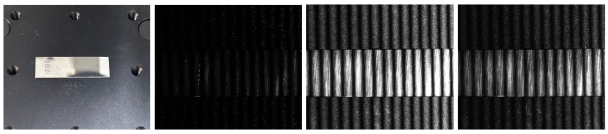

Fig. 9. The metal calibration block $1 \pm 0.001mm$ and the corresponding fringe patterns (from left to right: input images of two exposure levels for the model, images used by the SPF method).

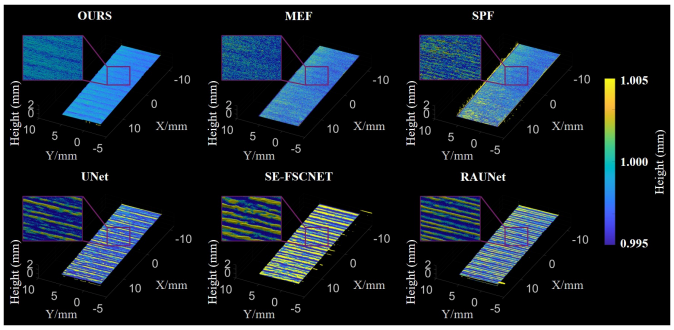

Fig. 10. Reconstructed point clouds of the gauge block. Deep learning methods show periodic fluctuations, while the single-exposure method fails in HDR regions. The proposed method achieves higher accuracy, closely matching the MEF results.

The quantitative reconstruction error metrics are summarized in Table II. Due to periodic fluctuations in the reconstructed point clouds, the comparison methods exhibit relatively large errors, indicating their limited capability in handling HDR regions. The single-exposure method also shows considerable reconstruction errors, as a single exposure cannot adequately address all HDR regions. In contrast, the proposed method achieves a MAE of 0.0030 mm with a standard deviation of 0.0038 mm, which is close to MAE and standard deviation of the multi-exposure fusion approach. These results demonstrate that the proposed method provides higher reconstruction accuracy and more effective handling of HDR regions compared with the comparison methods.

TABLE II
QUANTITATIVE COMPARISON OF RECONSTRUCTION ACCURACY FOR ALL METHODS, EVALUATED USING MAE AND SD ON GAUGE BLOCK.

	UNet	SE-FSCNet	RAUNet	SPF	MEF	OURS
MAE(mm)	0.0074	0.0128	0.0086	0.0063	0.0029	**0.0030**
SD(mm)	0.0094	0.0161	0.0107	0.0074	0.0036	**0.0038**

D. Ablation Experiments

To validate the effectiveness of MOE, MIF, and loss function, we performed the ablation experiments. The reconstruction results are shown in Fig. 11. Model A removes the MOE module, and due to the absence of the dark and bright modulation offset information estimated by the MOE, although the MIF module can still correct distribution shifts through guided pseudo-modulation features, the reconstruction results exhibit slight periodic fluctuations. Moreover, the lack of dark/bright offset information in the MIF inputs limits Model A's ability to handle overexposed and underexposed regions effectively. Model B omits both the MOE and MIF modules, resulting in no correction for the distribution shift in the pseudo-modulation features. This leads to a significant distribution deviation between the enhanced images and the ground-truth fringe patterns, causing pronounced periodic fluctuations in the reconstruction results. Model C changes the loss function to MSE Loss. Experimental results show that its reconstruction performance is poor, with significant artifacts on the object surface, indicating that this model lacks the ability to properly handle HDR surface reconstruction.

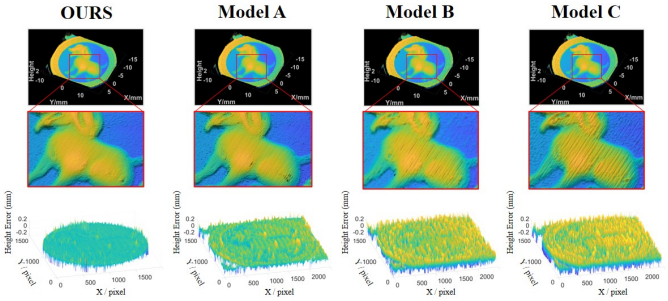

Fig. 11. Point cloud reconstruction results of different models in ablation experiment.

The SSIM, MRE and MAE for each experiment are shown in Table III. The proposed method explicitly estimates dark and bright modulation offsets using the MOE module and further corrects the distribution shift of pseudo-modulation features through the MIF module. In addition, a carefully designed combination of loss functions provides complementary supervision, leading to more accurate and stable reconstruction results. As a result, the proposed approach achieves an SSIM of 0.9814, while the MAE and MRE are reduced to 0.0095 mm and 0.0021 mm, respectively. Ablation studies further demonstrate the effectiveness of the proposed modules and loss function design, validating the superior reconstruction performance of the proposed method.

TABLE III
THE SSIM, MRE AND MAE FOR EACH EXPERIMENTS

	A	B	C	OURS
SSIM	0.9740	0.9739	0.9732	**0.9814**
MRE(mm)	0.0023	0.0038	0.0041	**0.0021**
MAE(mm)	0.0103	0.0174	0.0186	**0.0095**
SD(mm)	0.0174	0.0303	0.0321	**0.0155**

IV. Conclusion

In this paper, we proposed a deep learning approach for fringe projection profilometry. The method first extracts bright and dark features from the input fringe patterns, after which the MOE module estimates and compensates for modulation offsets. The MIF module then optimally fuses and refines these offset features to enhance the overall modulation of the fringe patterns. Experimental results demonstrate that the proposed method achieves high-accuracy reconstruction across high-dynamic-range regions. Comparative experiments show clear improvements over existing methods, and ablation studies confirm that the inclusion of the MOE and MIF modules, along with the carefully designed loss functions, significantly contributes to the enhanced reconstruction performance. These results validate the effectiveness and robustness of the proposed framework for high-precision FPP measurement.

V. Acknowledgment

This work was supported by the National Natural Science Foundation of China (grant number 62571355, 62101364); Provincial science foundation of Sichuan (grant number 2025ZNSFSC0534, 2025ZNSFSC1445).

References

[1] J. Zhu, J. Luo, G. Yin, and P. Zhou, "An expandable weighted spatial-temporal binarization encoding method maximizing intensity information utilization for 3d imaging," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[2] M. Niu, K. Song, L. Huang, Q. Wang, Y. Yan, and Q. Meng, "Unsupervised saliency detection of rail surface defects using stereoscopic images," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2271–2281, 2021.

[3] Z. Wang, J. Dong, M. Yin, Y. Zhu, H. Zheng, L. Xie, and G. Yin, "3-d reconstruction and measurement of blade profiles with laser-scanning sensor via multiview registration based on dynamic encoding of feature-coordinate information," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 2, pp. 1663–1674, 2024.

[4] Z. Zhu, P. Zhou, D. Wang, L. Cheng, and J. Zhu, "Online multi-relational clustering with dominant view mining," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, pp. 29232–29240, 2026.

[5] Z. Zhu, P. Zhou, W. Du, S. Min, and J. Zhu, "Unsupervised semantic discovery via global and local semantic alignment in multimodal clustering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, pp. 35248–35256, 2026.

[6] D. Palousek, M. Omasta, D. Koutny, J. Bednar, T. Koutecky, and F. Dokoupil, "Effect of matte coating on 3d optical measurement accuracy," *Optical Materials*, vol. 40, pp. 1–9, 2015.

[7] S. K. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *International Journal of Computer Vision*, vol. 21, no. 3, pp. 163–186, 1997.

[8] B. Salahieh, Z. Chen, J. J. Rodriguez, and R. Liang, "Multi-polarization fringe projection imaging for high dynamic range objects," *Optics express*, vol. 22, no. 8, pp. 10064–10071, 2014.

[9] Z. Song, Z. Song, J. Zhao, and F. Gu, "Micrometer-level 3d measurement techniques in complex scenes based on stripe-structured light and photometric stereo," *Optics Express*, vol. 28, no. 22, pp. 32978–33001, 2020.

[10] F. A. Saiz, I. Barandiaran, A. Arbelaiz, and M. Graña, "Photometric stereo-based defect detection system for steel components manufacturing using a deep segmentation network," *Sensors*, vol. 22, no. 3, p. 882, 2022.

[11] S. Wei, C. Zhou, S. Wang, K. Liu, X. Fan, and J. Ma, "Colorful 3-d imaging using an infrared dammann grating," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1641–1648, 2016.

[12] H. Lin, J. Gao, G. Zhang, X. Chen, Y. He, and Y. Liu, "Review and comparison of high-dynamic range three-dimensional shape measurement techniques," *Journal of Sensors*, vol. 2017, no. 1, p. 9576850, 2017.

[13] S. Zhang and S.-T. Yau, "High dynamic range scanning technique," *Optical Engineering*, vol. 48, no. 3, pp. 033604–033604–7, 2009.

[14] S. Feng, Y. Zhang, Q. Chen, C. Zuo, R. Li, and G. Shen, "General solution for high dynamic range three-dimensional shape measurement using the fringe projection technique," *Optics and Lasers in Engineering*, vol. 59, pp. 56–71, 2014.

[15] J. Du, F. Yang, H. Guo, J. Zhu, and P. Zhou, "Parameter selection on a multi-exposure fusion method for measuring surfaces with varying reflectivity in microscope fringe projection profilometry," *Applied Optics*, vol. 63, no. 13, pp. 3506–3517, 2024.

[16] S. Van der Jeught and J. J. Dirckx, "Deep neural networks for single shot structured light profilometry," *Optics express*, vol. 27, no. 12, pp. 17091–17101, 2019.

[17] H. Nguyen, Y. Wang, and Z. Wang, "Single-shot 3d shape reconstruction using structured light and deep convolutional neural networks," *Sensors*, vol. 20, no. 13, p. 3718, 2020.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[19] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE visual communications and image processing (VCIP)*, pp. 1–4, IEEE, 2017.

[20] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[21] Z.-L. Ni, G.-B. Bian, X.-H. Zhou, Z.-G. Hou, X.-L. Xie, C. Wang, Y.-J. Zhou, R.-Q. Li, and Z. Li, "Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments," in *International Conference on Neural Information Processing*, pp. 139–149, Springer, 2019.

[22] X. Cheng, Y. Tang, K. Yang, L. Liu, and C. Han, "Single-exposure height-recovery structured illumination microscopy based on deep learning," *Optics Letters*, vol. 47, no. 15, pp. 3832–3835, 2022.

[23] X. Zhu, Z. Han, L. Song, H. Wang, and Z. Wu, "Wavelet based deep learning for depth estimation from single fringe pattern of fringe projection profilometry," *Optoelectronics Letters*, vol. 18, no. 11, pp. 699–704, 2022.

[24] Z. Song, J. Xue, W. Lu, R. Jia, Z. Xu, and C. Yu, "Se-fscnet: full-scale connection network for single-shot phase demodulation," *Optics Express*, vol. 32, no. 9, pp. 15295–15314, 2024.

[25] Y. Li, K. Xu, G. P. Hancke, and R. W. Lau, "Color shift estimation-and-correction for image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25389–25398, 2024.

[26] J. Zhu, F. Yang, J. Hu, and P. Zhou, "High dynamic reflection surface 3d reconstruction with sharing phase demodulation mechanism and multi-indicators guided phase domain fusion," *Optics Express*, vol. 31, no. 15, pp. 25318–25338, 2023.

[27] P. Zhou, J. Zhu, X. Su, Z. You, H. Jing, C. Xiao, and M. Zhong, "Experimental study of temporal-spatial binary pattern projection for 3d shape acquisition," *Applied Optics*, vol. 56, no. 11, pp. 2995–3003, 2017.

[28] Y. Zheng, S. Wang, Q. Li, and B. Li, "Fringe projection profilometry by conducting deep learning from its digital twin," *Optics express*, vol. 28, no. 24, pp. 36568–36583, 2020.

[29] Q. Zhou and A. Jacobson, "Thingi10k: A dataset of 10,000 3d-printing models," *arXiv preprint arXiv:1605.04797*, 2016.

[30] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets v2: More deformable, better results," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9308–9316, 2019.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[32] S. Zhang, "Rapid and automatic optimal exposure control for digital fringe projection technique," *Optics and Lasers in Engineering*, vol. 128, p. 106029, 2020.