PAPER

Robust 3D reconstruction of highly reflective objects via structured light

Measurement Science and Technology

PAPER

RECEIVED
16 December 2025

REVISED
11 February 2026

ACCEPTED FOR PUBLICATION
18 March 2026

PUBLISHED
30 March 2026

Robust 3D reconstruction of highly reflective objects via structured light

Chengcheng Li[1,2], **Huiying Xu**[1,*], **Yan Zhang**[4], **Tao Wang**[4], **Lan Yu**[5], **De'ang Su**[1], **Zhendong Chen**[1] and **Xinzhong Zhu**[1,3]

[1] Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua 321004, People's Republic of China
[2] Modern Science and Technology College, China Jiliang University, Yiwu 322000, People's Republic of China
[3] Wenzhou University, Wenzhou 325035, People's Republic of China
[4] Hailiang Technology Group, Zhuji 311814, People's Republic of China
[5] Jinhua Vocational and Technical University, Jinhua 321016, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: xhy@zjnu.edu.cn

Keywords: 3D precision measurement, high-reflective, structured light, deep learning

Abstract

Achieving accurate three-dimensional (3D) measurement on metallic and similar components in industrial environments remains a major challenge in fringe projection profilometry, primarily due to severe phase information loss caused by overexposure in highly reflective regions and underexposure in low-reflectance areas commonly present on such surfaces. These issues often lead to incomplete or distorted depth reconstruction. To address this, we propose a novel fringe restoration framework, called multi-attention fusion network, designed to enhance the robustness and precision of structured light-based 3D reconstruction under complex reflectance conditions. A key component of our approach is the reflective-aware multi-attention module, which emphasizes critical phase features while suppressing noise from saturated pixels. This enables effective compensation for fringe distortion in reflective regions. To overcome data scarcity, we further introduce a mask-guided adaptive fringe fusion strategy, which augments training data by replacing degraded regions with high-quality patches guided by semantic segmentation masks. Experiments on challenging objects, such as aero-engine blades, demonstrate that our method reduces the mean absolute error of depth maps from 0.191 to 0.045 and improves valid point cloud coverage from 70.8 % to 98.4 %. These results demonstrate the effectiveness of the proposed method in improving the robustness, completeness, and relative accuracy of structured-light-based 3D reconstruction under high-reflectivity conditions.

1. Introduction

Structured light (SL) three-dimensional (3D) measurement technology [1], as a non-contact and active optical sensing method, has emerged as a vital solution in the field of high-precision metrology. It plays a critical role in industrial component inspection [2], automated reverse engineering [3], and other domains where accurate dimensional measurement is essential. The core principle of SL lies in integrating optical projection and imaging-based sensing: it projects encoded structured patterns onto the target surface and captures the reflected or modulated light using a calibrated imaging sensor. The captured fringe images are then decoded and processed through phase-to-height mapping and triangulation to reconstruct high-resolution 3D point clouds. This tight coupling between structured illumination, optical sensing, and computational reconstruction enables SL systems to achieve micron-level accuracy with high robustness against surface texture and ambient lighting. In particular, fringe projection profilometry [4] and Gray code encoding [5] have become the predominant techniques, owing to their sub-pixel phase

sensitivity and resilience to noise–making them highly suitable for precision instrumentation and real-time measurement applications.

However, when measuring highly reflective surfaces (e.g. metallic, plated, or mirror-like materials), conventional SL systems encounter significant challenges [6]. Specular reflections in high-reflectance areas often exceed the dynamic range of image sensors, causing local saturation and phase information loss, which leads to incomplete or distorted 3D point clouds. Conversely, low-reflectance regions may suffer from underexposure, thereby degrading reconstruction accuracy.

Early studies addressed this issue using physical surface modification or optical compensation techniques. For instance, spraying diffuse powder [7] improves surface diffusivity but compromises object integrity and efficiency. Polarization-based filtering [8] requires complex calibration and can reduce the signal-to-noise ratio (SNR) of the system.

Subsequent research explored adaptive fringe projection [9] and multi-exposure fusion (MEF) techniques [10], which dynamically adjust the projector's intensity or camera's exposure settings to suppress overexposure in reflective regions. For example, Sun and Zhang [11] constructed a projector intensity model with sub-pixel mapping to generate adaptive fringe patterns. Xu *et al* [12] designed an illumination template based on highlight mapping, which was multiplied with standard fringes to adjust projection. Tang *et al* [13] introduced a reflectance-based threshold model for automatically selecting optimal exposure times, while Lee Park [14] optimized multi-exposure setups to minimize the number of exposures within a limited range, improving measurement efficiency. Tan *et al* [15] proposed a hetero-spectral fringe multiplexing strategy that separates spectral channels to enable independent phase retrieval for each measured object, thereby achieving accurate and complete 3D measurements without interference from mutual reflection artifacts. In addition, Tan *et al* [16] combined Fourier analysis with the Hilbert transform and numerical simulations to model the relationship among phase error, saturation level, and the number of phase-shifting steps, enabling effective correction of saturation-induced phase errors. However, these methods often require multiple projections and acquisitions, making them complex and unsuitable for real-time industrial applications.

In recent years, deep learning-based approaches have achieved significant progress in fringe enhancement and image restoration. Lai and Chiang [17] developed a fringe inpainting system based on generative adversarial networks [18] (GANs) to recover locally missing fringe details. Song *et al* [19] proposed DcMcNet, a deformable convolution (DC)-based multi-scale network for repairing saturated regions in both sinusoidal and Gray-code patterns. Li *et al* [20] combined a pix2pix GAN with the DC-HNet architecture to progressively remove specular highlights and reconstruct accurate 3D shapes from fringe images. Li *et al* [21] proposed SS-E2E, which composed of a multipath branch auxiliary supervision network and guided by the fringe-phase physical model (MPS XNet) to predict the absolute phase. Tan *et al* [22] developed a cyclic-GAN based on a focused projection strategy to achieve domain translation from binary to sinusoidal fringes, allowing high-quality phase retrieval in both focused regions and non-ideal defocused areas. Furthermore, Wang *et al* [23] proposed a deep self-supervised MEF algorithm. Their SMEF network was trained without ground-truth labels to effectively remove exposure-related artifacts by fusing fringe images captured at different exposures.

In addition to the challenges posed by high-reflective surfaces, the limited volume and diversity of SL training datasets remain a significant bottleneck. To address this issue, recent research has explored various data construction strategies that aim to balance realism with manageable data acquisition costs. Existing approaches can be broadly categorized into two types: synthetic datasets and real-captured datasets. For synthetic data, two primary strategies are commonly adopted. The first utilizes simulation software such as MATLAB to generate fringe images. Yang *et al* [24] introduced random Gaussian noise into synthetic fringe patterns to mimic real-world disturbances. Li *et al* [25] used MATLAB toolboxes to create 3000 raster images with randomized deformation and noise (referred to as Li-3000) to support absolute phase prediction using a ResUNet architecture. The second strategy involves building virtual scanning environments in 3D rendering platforms like Blender. Zhu *et al* [26] constructed a large-scale synthetic dataset with configurable scene parameters and object poses to enhance rendering efficiency and diversity. Kim *et al* [27] developed a dataset called SynthFringe (Golden-18 000), based on 500 CAD models rotated 60° around a fixed axis, producing 18 000 images with black backgrounds, significantly improving geometric and visual variation within the training set. For real-world data, Nguyen and Wang [28] released a dataset containing approximately 1500 fringe-to-height samples of real statues, covering the full SL processing pipeline. Li *et al* [25] constructed another dataset with 980 images of statues and plastic parts, captured against a white background. However, real datasets under high dynamic range (HDR) conditions remain scarce, particularly in the context of specular surfaces.

In summary, while notable progress has been made, existing methods still struggle to effectively handle complex surface reflections and adapt to HDR scenarios. Moreover, most approaches rely heavily on large-scale annotated data, whereas currently available public SL datasets are limited in scale and insufficient in representing high-reflectivity conditions. This gap restricts model generalization and practical applicability. Therefore, improving fringe restoration quality in defect-prone regions under data-constrained conditions has become a pressing challenge in the field. It should be noted that structured-light 3D measurement is fundamentally constrained by optical processes such as specular reflection and sensor dynamic range, and this work focuses on mitigating locally correctable fringe degradations rather than eliminating all reflection-induced physical errors.

To address these limitations from a measurement instrumentation perspective, this work proposes a novel measurement enhancement framework that leverages reflective-aware multi-attention (RAMA) and mask-guided adaptive fringe fusion (MGAF). Our approach tackles the problem from two complementary perspectives. Specifically, a mask-guided adaptive fusion strategy is developed to integrate high-quality fringe structures from reference images, compensating for saturated or underexposed low-dynamic-range (LDR) regions in the target image. And, a feature enhancement module that combines multiple attention mechanisms is introduced to strengthen the network's capability in modeling local detail and global context, ultimately improving fringe restoration robustness. Unlike prior methods focused solely on computer vision metrics, our approach is designed to enhance the reliability and completeness of the SL-based 3D measurement pipeline, with special attention to measurement repeatability, depth accuracy, and point cloud integrity.

The major contributions of this work are summarized as follows:

- MGAF: We propose a novel cross-image fusion strategy that enhances fringe completeness in high-reflectivity scenarios by compensating for missing phase information in locates non-phase regions (NPRs). A U-Net-based segmentation network is employed to generate LDR masks in the target image, which are then used to guide the spatial alignment and fusion of high-quality fringe regions extracted from other reference images. This strategy enriches the training data's diversity and improves the robustness of downstream learning tasks.
- RAMA: We designed a powerful module, integrating channel attention, spatial attention, and Transformer-based multi-scale feature enhancement. RAMA enables simultaneous modeling of local and global contexts, reinforces edge and texture features, and effectively suppresses saturation-induced noise in reflective regions. This leads to precise restoration of defective areas in fringe images.
- Extensive experiments on representative datasets (including 3D-printed parts, metallic surfaces, and aero-engine blades) demonstrate the effectiveness of our method. The proposed approach significantly reduces the mean absolute error (MAE) in depth estimation within reflective regions and notably increases point cloud coverage, highlighting its robustness and potential for practical deployment in industrial scenarios.

2. Basic principle

As illustrated in figure 1, the general pipeline of deep learning-based SL 3D reconstruction for highly reflective objects can be summarized as follows: (1) SL patterns are actively projected onto the object's surface, and phase distortions caused by object geometry are captured. (2) A deep learning model is employed to restore overexposed fringe regions with missing information [29]. (3) The wrapped phase is demodulated and then unwrapped to obtain the absolute phase. (4) Finally, based on the calibrated system parameters, the 3D coordinates of the object surface are computed using triangulation [30].

The system hardware consists of three main modules: a projection unit, an image acquisition unit, and a computing/control unit. The calibration of the system follows Zhang's method [31], which includes internal and external parameters of both the camera and projector, as well as the mapping of the light plane geometry. Phase compensation algorithms are also employed to correct nonlinear distortions such as gamma correction and lens aberration.

In operation, the projector emits modulated structured patterns (e.g. Gray codes, sinusoidal phase-shift fringes, or multi-frequency heterodyne patterns), which are captured by the camera. Phase-shift fringe images with overexposed regions are input into a deep learning network for restoration. The phase-shifting method is then used to calculate the wrapped phase from multiple fringe images. A phase unwrapping algorithm in the temporal or spatial domain eliminates phase ambiguities, resulting in a pixel-wise absolute phase map. The absolute phase, combined with system calibration parameters, is used

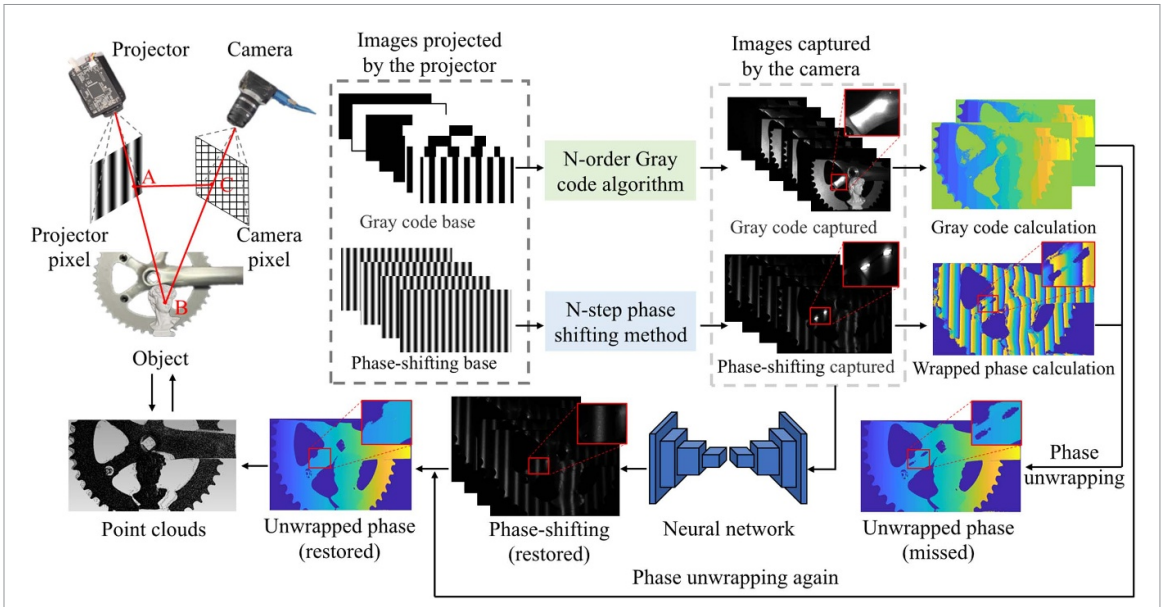

Figure 1. Principle of 3D reconstruction of highly reflective objects with structured light based on deep learning.

in triangulation to recover the 3D point cloud of the object surface [32]. Compared to multi-frequency heterodyne schemes [33], the Gray code with phase-shifting method [34] offers higher encoding efficiency and stronger robustness, and is thus adopted in this study. Below, we introduce key components of the system pipeline. In particular, the deep learning-based restoration of overexposed fringe regions is detailed in section 3.

2.1. Gray code with phase-shifting decoding principle

In SL 3D reconstruction, the hybrid encoding strategy of Gray code combined with phase-shifting enables both coarse and fine phase retrieval. Gray code determines the fringe order (coarse phase), while the phase-shifting technique retrieves the wrapped phase with sub-pixel accuracy. The absolute phase is then derived by fusing both components. The core principles are as follows:

1) Wrapped phase extraction: a set of phase-shifted fringe images is projected, and the wrapped phase is calculated based on intensity differences. Taking the four-step phase-shifting method as an example, the intensity of the ith projected fringe image $I_i(u,v)$ at pixel (u,v) is defined as:

$$I_i(u,v) = A + B\cos\left[\varphi(u,v) + \frac{i\pi}{2}\right], \quad i = 0,1,2,3. \tag{1}$$

Here, A and B are the background and modulation intensities, and $\varphi(u,v)$ is the wrapped phase. The wrapped phase can then be computed as [35]:

$$\varphi(u,v) = \arctan\left(\frac{I_3 - I_1}{I_0 - I_2}\right). \tag{2}$$

The output of this formula lies within the range $(-\pi,\pi)$, leading to inherent phase ambiguity. To resolve this, a Gray code sequence is used for phase unwrapping.

2) Gray code decoding: Gray codes are constructed such that adjacent codes differ by only one bit, and are converted to binary using XOR operations. In low-frequency phase unwrapping, each fringe period is assigned a unique Gray code identifier. For instance, a 5-bit Gray code sequence can divide the projection region into 32 fringe orders. The projector sequentially emits 5 Gray code patterns, and the binary value at each pixel is determined via thresholding. In this study, we adopt a 4+5 scheme, using four-step phase-shifting for wrapped phase extraction and five-bit Gray coding for unwrapping, resulting in the final absolute phase $\phi(u,v)$.

2.2. Reflection effects in SL measurement

In SL 3D reconstruction, the surface reflectance of the target significantly affects the fidelity of captured fringe patterns. Diffuse surfaces typically preserve fringe structures well, allowing stable phase decoding

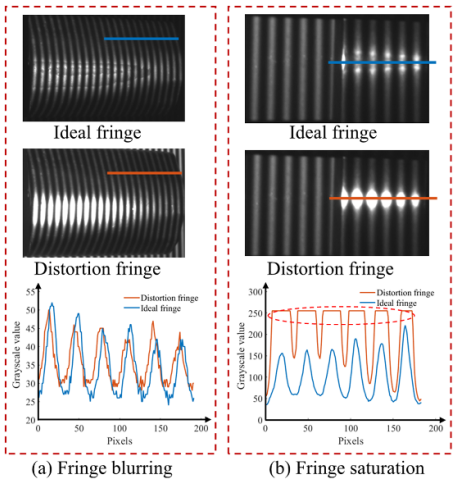

Figure 2. The sinusoidal fringe pattern of a highly reflective object appears in both high and low reflectivity regions.

and depth recovery. Ideally, SL assumes Lambertian reflection with a linear imaging response, ensuring that sinusoidal patterns maintain consistent gray-level modulation across the surface.

However, real-world surfaces such as metal or ceramic often exhibit specular reflections, strong highlights, or sharp reflectance transitions. These effects can lead to local saturation, underexposure, or occlusion shadows, severely distorting fringe intensity distributions. In saturated regions, intensity values are clipped and amplitude information is lost. In underexposed regions, SRNs degrade, and fringe patterns become indistinct or unreadable. Transitional regions between high and low reflectivity can introduce waveform distortions and discontinuities. Such artifacts result in significant phase decoding errors, undermining both the decodability and robustness of SL patterns, and ultimately compromising 3D reconstruction quality.

To systematically analyze these distortions, we randomly select a column of pixels from three types of regions (overexposed, underexposed, and normal) and plot their intensity curves, as shown in figure 2. Results indicate two main types of degradation under HDR conditions:

1) Underexposure in dark regions: in figure 2(a), highlights falling in dark fringe zones elevate local brightness and reduce fringe contrast, blurring edges and obscuring structural cues. Additionally, low-reflectivity surfaces may produce weak reflections that are indistinguishable from noise, making fringe patterns unresolvable and leading to signal dropout or decoding failure.

2) Saturation in bright regions: as shown in figure 2(b), when specular highlights align with the bright fringes of sinusoidal patterns, sensor saturation (e.g. reaching 255 in 8-bit images) forms flattened peak plateaus. This destroys the sinusoidal periodicity and continuity, causing severe phase extraction errors.

In conclusion, sinusoidal fringe patterns are highly susceptible to structural distortions on reflective surfaces, significantly impacting phase calculation and 3D reconstruction. Therefore, it is essential to perform fringe restoration and enhancement prior to phase decoding, in order to recover ideal encoding characteristics and ensure the robustness and accuracy of the entire SL measurement system.

3. Network building

In this section, we present the overall architecture of multi-attention fusion network (MA-FusionNet), which aims to restore distorted fringe patterns and achieve accurate phase reconstruction under high-reflectivity conditions. The network adopts a RAMA-based restoration backbone equipped with DC to handle nonlinear distortions caused by specular reflections. The DC layer first adapts the sampling positions to capture complex geometric variations, while the RAMA blocks integrate convolutional and Transformer-based attention to recover both local details and global consistency. In addition, a MGAF strategy is introduced during training to enhance data diversity by fusing high-quality fringe patches

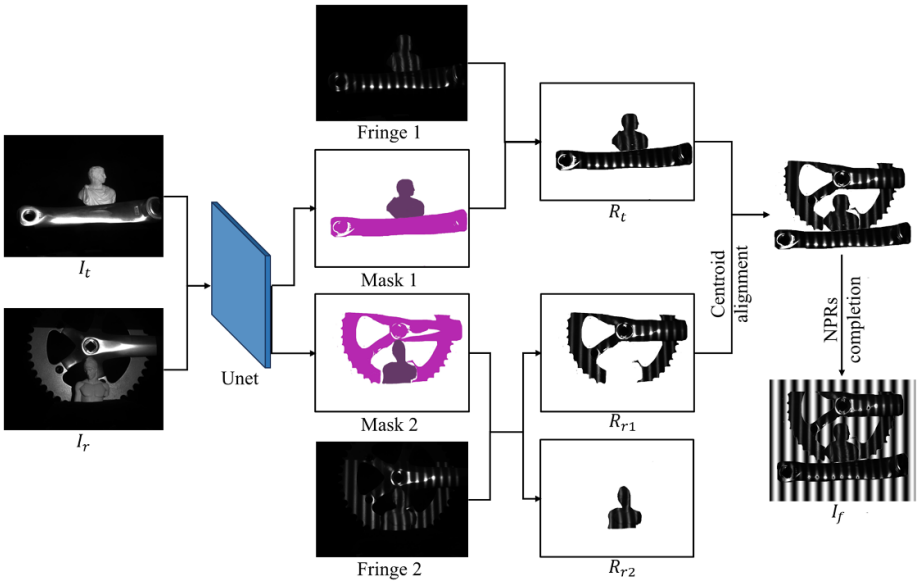

Figure 3. Principle diagram of structured light dataset enhancement method based on MGAF.

from reference images into degraded regions. This augmentation allows the model to learn reflection-aware correspondences between surface reflectance and fringe modulation, improving generalization to real high-reflectivity scenes. Through the synergy of the DC and RAMA modules, MA-FusionNet reconstructs physically consistent and spatially continuous fringe patterns, achieving high-fidelity phase recovery under complex reflective conditions.

It worth that, although Gray code patterns may be affected by intensity saturation, their binary decoding makes them relatively robust to moderate distortions, whereas errors in phase-shift fringes directly propagate and are amplified during phase unwrapping; therefore, the proposed method focuses on correcting phase-related distortions, which have a much larger impact on reconstruction accuracy, while remaining extensible to Gray code correction if needed. The following subsections describe each module in detail, beginning with the MGAF-based dataset augmentation strategy.

3.1. SL dataset augmentation via MGAF

To address the limited generalization capability of deep learning models caused by data scarcity in SL 3D reconstruction of highly reflective objects, we proposed MGAF strategy. The core idea is to leverage high-quality fringe patterns from reference images to adaptively restore NPRs regions in the target image, thereby providing more robust inputs for subsequent fringe reconstruction. The overall workflow is illustrated in figure 3, and consists of the following four main steps:

Step 1: NPRs identification and mask generation. Given a target fringe image $I_{\mathrm{t}} \in \mathbb{R}^{H\times W}$ for enhancement, a semantic segmentation network based on U-Net is first used to generate a binary foreground-background mask $M \in \{0,1\}^{H\times W}$, defined as:

$$M(x,y) = \begin{cases} 1, & \text{if } (x,y) \text{ belongs to the foreground object region,} \\ 0, & \text{if } (x,y) \text{ belongs to the background.} \end{cases} \tag{3}$$

Within the foreground region, a statistical analysis of pixel intensity is then performed to identify the NPRs R_{t}, i.e. the area lacking valid phase information, which serves as the target for restoration.

Step 2: Reference region extraction via centroid alignment. To restore R_{t}, a spatially aligned fringe-exposed region R_{r} is extracted from a reference image $I_{\mathrm{r}} \in \mathbb{R}^{H\times W}$. Specifically, we compute the centroid $C_{\mathrm{t}} = (x_t, y_t)$ of the target NPRs R_{t} as:

$$C_{\mathrm{t}} = \frac{1}{|R_{\mathrm{t}}|} \sum_{(x,y)\in R_{\mathrm{t}}} (x,y)\,. \tag{4}$$

Using C_{t} as the center, a slightly larger region is cropped from I_{r} to form R_{r}, ensuring that the selected patch contains complete and high-quality fringe structures.

Step 3: Adaptive cropping strategy. As the reference region R_r is typically larger than the target region R_t, we introduce an adaptive cropping strategy to extract a sub-region $R_r^{cropped}$ from R_r that matches the size of R_t, defined as:

$$R_r^{cropped} = \text{center\_crop}\left(R_r, \text{size} = |R_t|\right). \tag{5}$$

Here, center_crop$(\cdot)$ denotes a centered cropping operation with respect to C_t, ensuring spatial alignment between the reference and target patches.

Step 4: Fusion and image restoration. The final fused image $I_f \in \mathbb{R}^{H\times W}$ is generated under the guidance of the mask M, where the NPRs R_t is selectively filled using the reference content:

$$I_f(x,y) = \begin{cases} R_r^{resized}(x,y), & \text{if } (x,y) \in R_t, \\ I_t(x,y), & \text{otherwise.} \end{cases} \tag{6}$$

This selective replacement ensures that high-quality fringe information from the reference image is used to compensate for degraded regions in the target, thereby mitigating the negative impact of invalid pixels during training and improving overall input quality.

In addition, the reference image I_r is preferentially selected to maximize the spatial overlap with the non-phase-shift regions of the target image I_t, while simultaneously containing sufficient phase-shift information for effective data augmentation. In practice, I_r is chosen to cover a large area of valid fringe patterns in I_t and provide rich phase cues to guide the MGAF process. During Step 2, a high-quality region with valid fringe structures is cropped from I_r based on the spatial correspondence of the degraded area in I_t. The extracted patch is then geometrically aligned and fused into the defective region under the guidance of the segmentation mask. Essentially, the reference object functions as a data donor, enabling cross-image compensation that enriches the diversity and completeness of the training data. The fused regions preserve phase continuity and spatial correspondence with the original degraded areas, ensuring metrological consistency during 3D reconstruction.

Overall, the proposed MGAF strategy combines semantic segmentation for precise localization with geometric alignment for structurally consistent fusion. It achieves cross-image fringe completion while maintaining global consistency and enhancing local texture details, significantly improving the robustness of the network in handling HDR conditions.

3.2. Fringe image restoration network based on the RAMA module

Based on the physical reflection effects [36] in structured-light measurement, reflection-induced degradations can be broadly divided into two categories. The first category includes local effects such as saturation, underexposure, and fringe contrast loss, which primarily disrupt fringe visibility and phase completeness. The second category includes global illumination effects, such as mutual and cross-reflections, which alter light transport paths and introduce non-local distortions. In practical structured-light 3D reconstruction of highly reflective objects, the most frequently encountered degradations are dominated by the first category. Specular reflection and exposure imbalance often lead to locally missing or severely distorted phase regions in the captured fringe images, which in turn significantly impacts the accuracy of phase computation and subsequent 3D reconstruction. To address the degradation of fringe patterns caused by such reflective disturbances, we propose a fringe image restoration network based on a RAMA module. It should be emphasized that the proposed RAMA module do not attempt to reconstruct physically unobservable information caused by complex light transport phenomena. Instead, they aim to learn statistically consistent mappings to restore locally corrupted fringe patterns, assuming that the dominant degradations arise from exposure-related effects rather than multi-path reflection. As illustrated in figure 4, the network aims to restore damaged sine fringe patterns by jointly modeling local details and global context, thereby enhancing phase recovery in missing regions.

3.2.1. Multi-scale feature fusion

To effectively exploit multi-modal information, we use both the captured fringe image $I_s \in \mathbb{R}^{H\times W}$ and the absolute phase map $\Phi_a \in \mathbb{R}^{H\times W}$ decoded by the system as joint inputs to the network. The fringe image encodes surface-related intensity variations, while the absolute phase map provides global, continuous phase information that aids restoration in degraded regions. Specifically, two independent convolutional encoders E_s and E_ϕ are used to extract features from I_s and Φ_a, respectively: $F_s = E_s(I_s), \quad F_\phi = E_\phi(\Phi_a)$. The resulting features are then concatenated to form a high-dimensional multimodal representation: $F = \text{Concat}(F_s, F_\phi)$. This fusion strategy, employed during training, enhances the network's perception of different reflective regions and improves restoration stability. During inference,

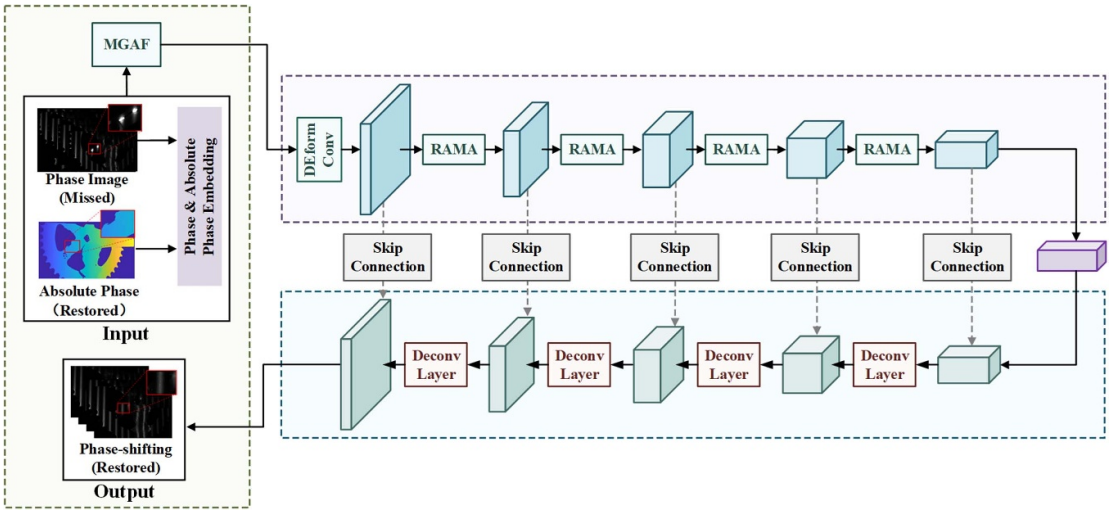

Figure 4. Schematic diagram of fringe image restoration network based on RAMA module.

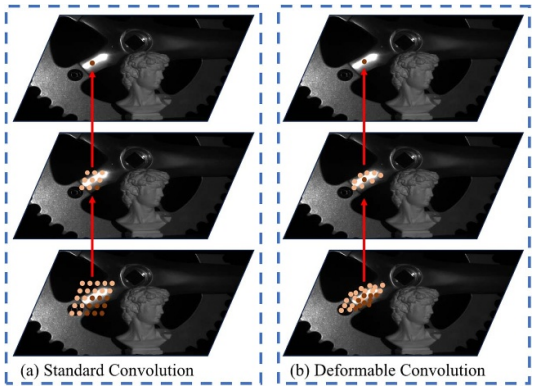

Figure 5. The working principle of deformable convolution processing for high reflective areas. Deformable convolution allows flexible deformations of the convolution kernel by adding an offset to each position.

however, only the fringe image is used as input to ensure model practicality and avoid dependency on phase maps.

3.2.2. DC module

Reflective regions typically exhibit significant nonlinear distortions and local fringe deformations. Conventional convolutional kernels, which use fixed sampling grids, often struggle to accurately model such irregular structures. To enhance the network's capacity for capturing complex geometric variations, a DC module is introduced prior to the RAMA module, as shown in figure 5. This module adapts the sampling positions of the convolutional kernel via learnable offsets Δp_k, enabling flexible spatial adaptation and structural alignment. The operation can be expressed as:

$$y(p_0) = \sum_{k=1}^{K} w_k \cdot x(p_0 + p_k + \Delta p_k). \tag{7}$$

Here, x denotes the input feature map, w_k the convolution weights, and p_k the standard sampling locations. In high-reflection areas, where specular highlights or local light accumulation induce severe nonlinear intensity distortions, the DC effectively captures these irregular patterns, providing accurate contextual information for subsequent attention-based feature fusion.

3.2.3. RAMA

The extracted features are fed into the proposed RAMA module, a parallel attention architecture composed of a channel attention module (CAM), a spatial attention module (SAM), and a Transformer

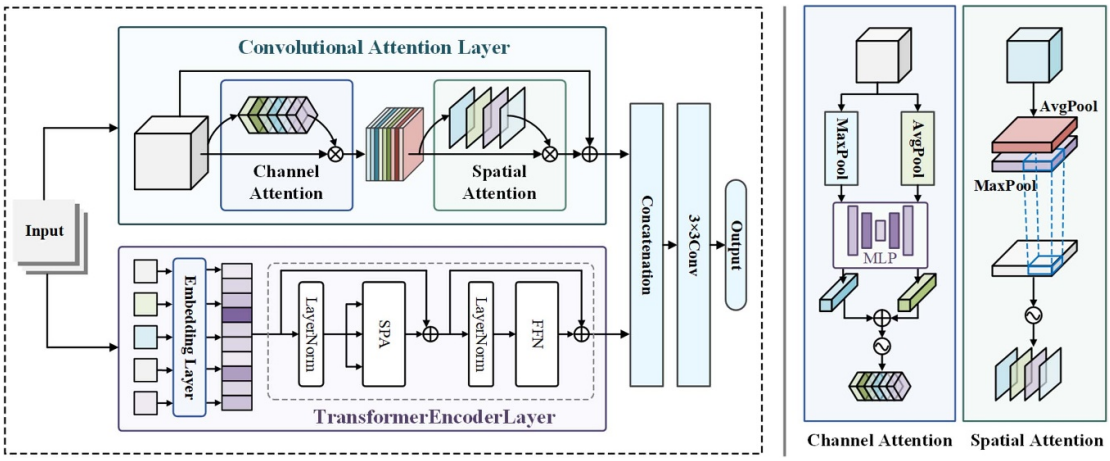

Figure 6. Schematic diagram of RAMA structure. Our proposed RAMA block which contains two parallel streams, i.e. convolutional attention and Transformer layer respectively.

encoder. As illustrated in figure 6, RAMA integrates the local feature extraction capabilities of convolutional neural networks with the global context modeling strengths of Transformer architectures. The core idea is to bridge the limitations of each mechanism by combining convolutional operations with self-attention, thereby effectively addressing degradation from local reflections and discontinuities in phase information caused by varying surface reflectivity.

Convolutional attention mechanism: Conventional convolutional neural networks typically focus on local receptive fields and struggle to model long-range dependencies, which are essential for recovering corrupted fringe regions. To enhance the network's sensitivity to reflective artifacts, we introduce convolution-based attention by incorporating both spatial and channel attention mechanisms. Formally, they are defined as:

$$\begin{aligned} F_{\mathrm{SAM}} &= \sigma\left(\mathrm{Conv}_{3\times 3}\left(F\right)\right) \odot F, \\ F_{\mathrm{CAM}} &= \sigma\left(\mathrm{MLP}\left(\mathrm{GAP}\left(F\right)\right)\right) \cdot F, \end{aligned} \tag{8}$$

$$F_{\mathrm{local}} = F_{\mathrm{SAM}} + F_{\mathrm{CAM}}. \tag{9}$$

Here, σ denotes the sigmoid activation, $\mathrm{GAP}(\cdot)$ is global average pooling, and $\odot$ represents element-wise multiplication. The spatial attention adjusts the pixel-wise importance based on location, while channel attention adaptively reweights the importance across feature channels, improving the model's ability to capture subtle local features in highly reflective regions.

Multi-head self-attention mechanism: To compensate for the limited receptive field of convolutions, the RAMA module incorporates a parallel Transformer branch based on the pyramid vision transformer architecture. It employs a spatial reduction attention (SRA) mechanism to capture global dependencies efficiently while reducing computational cost and memory overhead:

$$F_{\mathrm{global}} = \mathrm{Transformer}\left(F;\ \mathrm{SRA}\right). \tag{10}$$

This self-attention pathway models long-range contextual relationships, which helps to restore globally coherent phase structures in discontinuous regions affected by severe reflections.

Parallel fusion architecture: RAMA fuses local and global information through a parallel dual-path structure. Each RAMA block contains two sub-networks: one performing convolutional attention and the other executing Transformer-based self-attention. Their outputs are merged through a 3×3 convolutional layer:

$$F_{\mathrm{RAMA}} = \mathrm{Conv}_{3\times 3}\left(F_{\mathrm{local}} \oplus F_{\mathrm{global}}\right). \tag{11}$$

Here, $\oplus$ denotes the channel-wise concatenation. This design allows the network to jointly leverage fine-grained local details and holistic global patterns, substantially improving the accuracy and consistency of fringe restoration, particularly in complex reflective scenarios.

3.2.4. Network architecture for fringe restoration based on RAMA

In SL 3D reconstruction of highly reflective objects, specular reflections often lead to unstructured, cross-scale defective regions in the captured fringe images. These corrupted areas may span different spatial frequencies and structural hierarchies. To enhance the network's ability to model reflective defects across multiple scales, we design a pyramid-style encoder–fusion–decoder architecture that progressively extracts, fuses, and restores missing phase-shift fringe information.

The encoder consists of five stages, each performing spatial downsampling on the input fringe image I_s to capture features from local to global contexts, denoted as $I_r = \mathcal{N}_\theta(I_s)$, where $\mathcal{N}_\theta$ represents the restoration network parameterized by θ. Specifically, the first stage employs DC to enhance the network's sensitivity to nonlinear distortions caused by specular reflections. By dynamically adjusting the receptive field, DC adapts to local geometric deformations. The subsequent four stages adopt the proposed RAMA module to perform feature extraction and fusion. RAMA combines local attention mechanisms with global semantic context modeling, enabling the network to learn rich spatial information at different resolutions. The outputs of all encoding stages are aggregated into a hierarchical pyramid structure, allowing the network to address texture loss and structural distortion at multiple scales.

In the decoding phase, we introduce skip connections to propagate features from the encoder to the decoder, along with a reflection-aware enhancement mechanism that enables coarse-to-fine feature restoration. These connections preserve low-level details while introducing high-level semantic constraints, significantly improving the stability and accuracy of fringe reconstruction under complex reflective conditions.

3.3. Loss function design

To effectively guide the network in recovering missing fringe patterns in highly reflective regions while ensuring structural consistency and edge continuity. And we design a multi-component loss function composed of reconstruction error, structural fidelity, and edge-aware constraints. The total loss is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{ssim}} + \lambda_3 \mathcal{L}_{\text{grad}}. \tag{12}$$

Here, λ_1, λ_2, and λ_3 are weighting coefficients.

3.3.1. Reconstruction loss $\mathcal{L}_{rec}$

To ensure pixel-wise fidelity between the restored output and the ground truth fringe pattern, we employ the mean squared error loss:

$$\mathcal{L}_{\text{rec}} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\hat{I}_t^{(i,j)} - I_{gt}^{(i,j)} \right)^2. \tag{13}$$

Here, $\hat{I}_t$ is the network-generated fringe image and I_{gt} denotes the corresponding defect-free ground truth image.

3.3.2. Structural similarity loss $\mathcal{L}_{ssim}$

Accurate reconstruction in reflective regions requires not only pixel precision but also structural consistency. We employ the structural similarity index (SSIM) as a perceptual loss to enforce local structural fidelity:

$$\mathcal{L}_{\text{ssim}} = 1 - \text{SSIM}\left(\hat{I}_t, I_{gt}\right). \tag{14}$$

This loss captures contrast and structural variations, particularly enhancing the preservation of fringe orientation and fine textures.

3.3.3. Gradient consistency loss $\mathcal{L}_{grad}$

Fringe images affected by specular reflection often exhibit blurred or broken edges. To promote edge continuity and geometric coherence, we introduce a gradient-based loss to match edge structures:

$$\mathcal{L}_{\text{grad}} = \frac{1}{HW} \sum_{i,j} \left| \nabla_x \hat{I}_t^{(i,j)} - \nabla_x I_{gt}^{(i,j)} \right| + \left| \nabla_y \hat{I}_t^{(i,j)} - \nabla_y I_{gt}^{(i,j)} \right|. \tag{15}$$

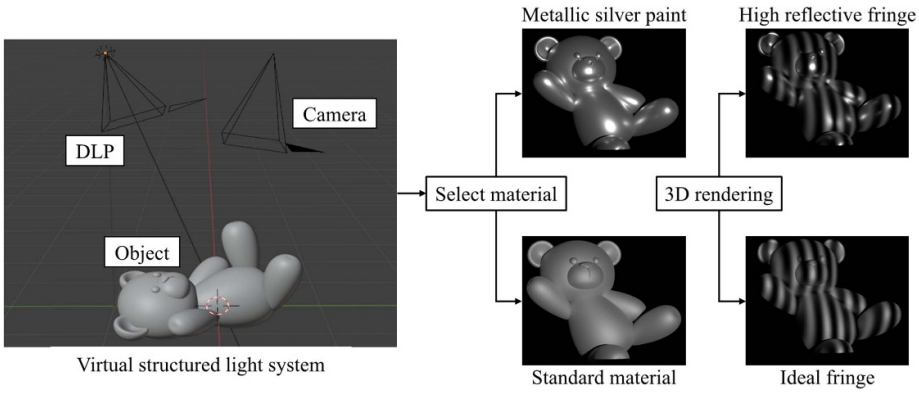

Figure 7. Virtual dataset construction method.

Here, ∇_x and ∇_y denote the horizontal and vertical image gradients, respectively, which emphasize precise fitting along the edges.

The combined multi-component loss jointly optimizes for pixel-level accuracy, structural perception, and edge continuity. This design effectively guides the network to reconstruct complete and coherent fringe patterns under challenging reflective disturbances, providing reliable input for subsequent phase decoding and 3D reconstruction tasks.

4. Experiment

4.1. Dataset preparation

4.1.1. Virtual simulation dataset

To effectively train the proposed MA-FusionNet, we construct a large-scale and high-quality dataset using a simulated SL scanning environment. As illustrated in figure 7, we leverage Blender, a 3D modeling and rendering software, to build a digital twin of a real-world SL system. This simulation environment replicates the physical configuration of the projector, camera, and target object, enabling the generation of diverse and photorealistic fringe patterns. The dataset generation process comprises two major stages:

1) **Model import and preprocessing**: We import 3D models in STL format into the Blender environment and adjust their spatial pose and geometry using transformation tools to conform to the predefined field of view and depth range. A total of 300 objects from various categories are selected from public model repositories. To simulate different surface reflectance conditions, we vary material parameters, including specular intensity and roughness.

2) **Fringe pattern generation and data acquisition**: Using Blender's built-in 'SL Scanning' plugin, we generate phase-encoded fringe sequences based on Gray code and sinusoidal phase-shifting principles. The user interface allows parameterization of fringe frequency, projection orientation (horizontal/vertical), and image resolution (1920×1080). During scanning, the system automatically performs synchronized projection and image acquisition of multi-frequency fringe patterns. The collected data is hierarchically organized into directories containing raw fringe images, unwrapped phase maps, and corresponding ground truth depth maps.

For each object, two image sets are acquired: one with a metallic silver-painted surface that generates strong specular reflections (used as the input to MA-FusionNet), and the other with a uniformly coated diffuse surface (used as the ground truth). The resulting virtual dataset consists of high-reflection fringe images as network input and ideal fringe patterns as supervisory targets for training MA-FusionNet.

4.1.2. Real-world dataset

To validate the performance of MA-FusionNet in real-world scenarios, we used the same dataset construction method as in [37]. We construct a SL measurement system, as shown on the left of figure 1. The hardware platform includes both data acquisition and processing modules. Specifically, we utilize a TI LightCrafter DLP3010 projector (1280×720 resolution) paired with a Hikvision MV-CA020-10UM

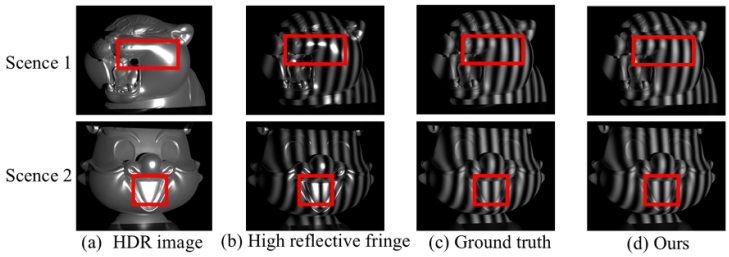

Figure 8. High reflectivity object prediction performance of Blender simulation dataset.

industrial camera (1624 × 1240 resolution, 8-bit grayscale). The optical path incorporates a Hikvision MVL-MF1228M-8MP fixed-focus lens (focal length: 12 mm) to capture high-resolution fringe patterns from various reflective targets.

The target objects include aerospace turbine blades, bicycle cassettes, and plaster statues with diffuse reflectance, simulating challenging HDR measurement scenes. Each object is captured under two different surface conditions: (1) a high-reflection set acquired directly from the original surface (used as network input), and (2) a diffuse set acquired after spraying the surface with white matte powder (used as the ground truth fringe image). The sprayed matte surface produces high-quality fringe images, from which the absolute phase is calculated and used as the metrologically reliable ground truth in this study. Additionally, multi-exposure imaging is employed to synthesize HDR fringe patterns following the method of Song *et al* [38], which estimates the camera response and fuses radiance maps in the gradient domain for accurate 3D reconstruction of specular surfaces.

4.2. Training

We implement our network using the PyTorch framework and train it on a server equipped with an NVIDIA GeForce RTX 3090 GPU. The training process spans 100 epochs and takes approximately 2 h. The network is optimized using the Adam optimizer with default momentum parameters and the batch size is set to 64. The initial learning rate is set to 1×10^{-4} and is gradually decayed during training using a step-based learning rate scheduler. The virtual dataset mentioned in section 4 is used for training, comprising 3000 samples. Among these, 70 % are allocated for training, 20 % for validation, and 10 % for testing. The performance of MA-FusionNet is evaluated from multiple perspectives. First, we assess its ability to restore depth from highly reflective objects. Since reflective distortions can severely degrade phase reconstruction and lead to depth artifacts, we evaluate the reconstructed phase and resulting point cloud quality as performance metrics. Second, we conduct ablation studies to examine the impact of incorporating the MASF-based dataset augmentation strategy on accuracy and reconstruction robustness. We also demonstrate the specific contributions of DC and the RAMA module to the overall performance of our network.

4.3. Qualitative result analysis

To comprehensively evaluate the performance of MA-FusionNet in fringe image restoration and phase recovery under highly reflective conditions, we conduct qualitative experiments on both synthetic and real-world datasets. In this section, we employ two classical phase decoding techniques, the three-step, four-step and twelve-step phase-shifting methods to process fringe images before and after restoration. To further demonstrate the superiority of our approach, we compare MA-FusionNet with two state-of-the-art representative networks: a GAN-based method [17] and DcMcNet [19].

Figure 8 presents the predicted phase-shift fringe patterns on the simulated dataset. As illustrated, MA-FusionNet effectively eliminates the depth fluctuations and surface discontinuities caused by specular reflection distortions. On the synthetic dataset, the predicted fringe patterns closely resemble the ground truth, exhibiting high structural consistency and visual fidelity.

To further verify the real-world applicability of MA-FusionNet, we visualize the restoration results on several representative scenes from the real-world dataset, as shown in figure 9. The figure compares the absolute phase recovery results computed from the predicted phase-shift fringe images. Specifically, the purple box highlights the phase reconstruction results in the NPRs, while the red box indicates regions with intense specular reflections.

The results reveal that conventional methods suffer from phase decoding failures in both high-reflective regions and NPRs. In contrast, MA-FusionNet is capable of adaptively identifying corrupted

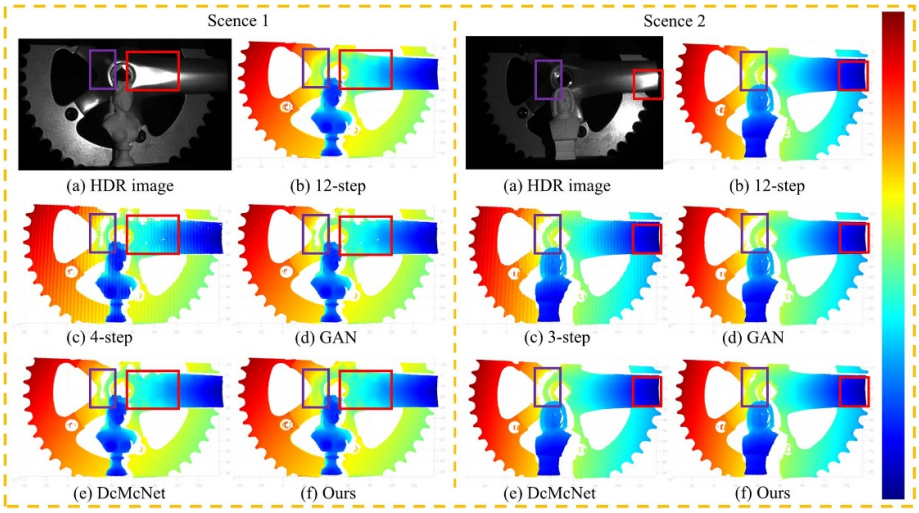

Figure 9. Comparison of absolute phase calculation results of phase shift fringes calculated/predicted by different methods in typical HDR scenarios.

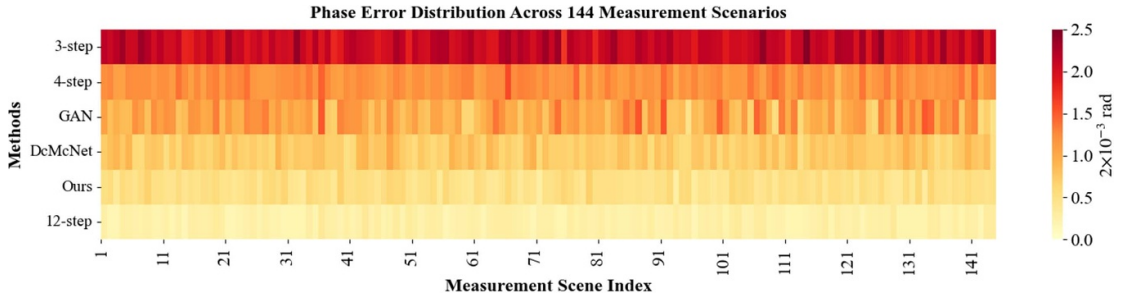

Figure 10. Comparison of absolute phase solution/prediction errors of different methods.

areas and effectively reconstructing complete fringe structures using both local details and global contextual cues. The restored phase maps exhibit smooth and natural transitions at boundaries, without noticeable texture discontinuities, thereby avoiding phase decoding errors in saturated regions validating the robustness and accuracy of our approach in challenging reflective environments.

4.4. Quantitative result analysis

We designed and conducted a comprehensive set of quantitative experiments to evaluate the accuracy of different methods in absolute phase computation. Specifically, we measured the average phase error across 144 testing scenarios to compare the robustness of each method under diverse conditions. To ensure that the error calculation focused only on valid regions, we set the modulation threshold B to 1.5, which automatically excluded low-quality areas where phase accuracy could not be guaranteed. Figure 10 presents the distribution of phase reconstruction errors across the 144 test scenes. The horizontal axis represents the scene index, while the vertical axis denotes the average absolute error (AAE), computed over all valid pixels. The AAE is defined as follows:

$$\phi_{\mathrm{AAE}} = \frac{1}{P}\sum_{i=1}^{P}\left|\phi_i^{\text{12-step}} - \phi_i\right|. \tag{16}$$

Here, i indexes valid pixels and P denotes the total number of valid pixels in the scene. $\phi_i^{\text{12-step}}$ represents the reference (ground truth) absolute phase obtained from the standard 12-step phase-shifting method after applying a matte coating on the object surface to suppress specular reflections.

As illustrated in figure 10, the heatmap depicts the distribution of phase reconstruction errors across 144 measurement scenarios for six representative methods. The color gradient indicates the magnitude of the AAE, where darker tones represent higher errors. Compared with conventional and learning-based baselines, our proposed method exhibits a consistently lighter color band across all scenes, revealing its

Table 1. Comparison of the point cloud coverage rate C for different methods.(%).

Method	Scene 1	Scene 2	Scene 3	Scene 4	Mean value
3-step	69.4	67.2	73.1	71.6	70.8
4-step	77.5	73.7	83.6	82.9	79.9
GAN	87.3	86.4	88.9	89.7	88.1
DcMcNet	95.3	94.5	94.2	94.1	94.5
12-step	96.2	95.5	95.4	93.3	95.1
Ours	**98.6**	**98.0**	**98.8**	**98.3**	**98.4**

remarkable robustness and stability under diverse reflective and geometric conditions. The error values remain uniformly low without noticeable outliers, highlighting the strong generalization capability of MA-FusionNet.

Quantitatively, our approach reduces the mean AAE by 75.4 % relative to the classic three-step phase-shifting method, while requiring fewer input images. It also achieves a 52.3 % reduction compared with the GAN-based approach. These results demonstrate that MA-FusionNet not only ensures higher reconstruction accuracy but also maintains computational efficiency, making it particularly suitable for real-time or resource-constrained structured-light measurement scenarios where both precision and speed are essential.

Table 1 presents a comparison of the point cloud coverage achieved by different methods. The quantitative evaluation metric, coverage C, is defined as:

$$C = \frac{P_m}{P_b} \times 100\%, \tag{17}$$

Here, P_m denotes the number of points in the reconstructed point cloud, and P_b represents the number of points in the reference point cloud. The reference point cloud is generated by spraying a thin layer of white powder onto the object surface to eliminate reflectivity, thereby ensuring high-quality stripe acquisition and accurate 3D reconstruction.

As summarized in table 1, the proposed method consistently achieves the highest point cloud coverage rate across all evaluated scenes, demonstrating its superior ability to recover valid 3D points in highly reflective regions. Traditional phase-shifting methods (3-step and 4-step) suffer from severe information loss caused by saturation and underexposure, resulting in sparse and incomplete point clouds, with average coverage rates below 80%. Although increasing the number of phase-shifting steps (12-step) improves robustness to noise, it remains limited by irreversible fringe saturation and thus cannot fully recover missing surface regions. Compared with deep learning-based approaches, our method also shows a clear advantage. The GAN-based method improves coverage by learning local fringe completion, but still leaves noticeable gaps in strongly reflective areas, leading to an average coverage of 88.1%. DcMcNet further enhances reconstruction completeness through multi-scale feature modeling and achieves a higher mean coverage of 94.5%. However, residual saturated regions and incomplete phase restoration still prevent full surface recovery. In contrast, the proposed method achieves an average point cloud coverage rate of 98.4%, outperforming DcMcNet by approximately 3.9 percentage points (98.4% vs 94.5%). This improvement is consistently observed across all scenes, indicating strong robustness under varying reflectivity conditions. The significantly higher coverage demonstrates that the proposed method is more effective in compensating for saturation-induced information loss, enabling more complete surface reconstruction and denser point clouds. These results confirm that the proposed approach provides substantial benefits for structured-light 3D reconstruction of highly reflective objects, particularly in scenarios where reconstruction completeness is critical.

Furthermore, in a structured-light 3D reconstruction system, the final reconstructed point cloud represents the true surface geometry of the object and constitutes the most critical output from a practical application perspective. To evaluate the impact of fringe restoration on the final geometric reconstruction, we adopt a calibrated absolute phase-to-height mapping scheme based on triangulation to generate 3D point clouds from the recovered phase maps. In addition to the quantitative evaluations, we further compare the proposed method with two representative approaches: a traditional multiple-exposure method (ME), which relies on exposure fusion to alleviate saturation, and a recent CNN-based fringe restoration method, TC-Net [37]. These comparisons allow us to assess the effectiveness of the proposed approach at the point-cloud level, reflecting its practical performance in structured-light 3D reconstruction under high-reflectivity conditions.

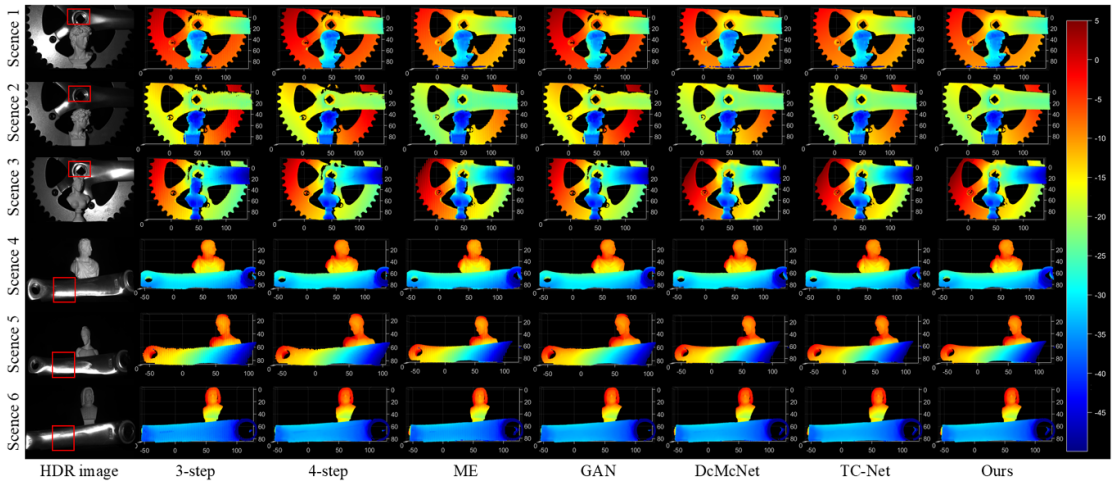

Figure 11. The reconstructed point clouds for four representative scenes.

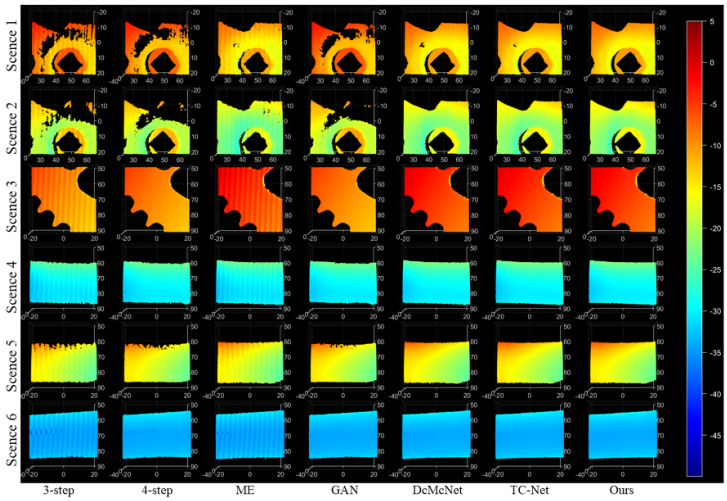

Figure 12. The locally enlarged reconstruction results for four representative scenes in highly reflective regions.

Figure 11 illustrates the reconstructed 3D point clouds for seven representative scenes obtained using different fringe restoration strategies. It can be observed that methods relying on limited exposure compensation or single-stage restoration (e.g. 3-step, 4-step, and ME) suffer from noticeable geometric distortions in highly reflective regions, which manifest as depth outliers, surface ripples, and broken contours. Although learning-based approaches such as GAN, DcMcNet, and TC-Net alleviate some of these issues, residual depth inconsistencies and local discontinuities are still apparent, particularly around specular highlights and sharp geometric transitions. In contrast, the proposed MA-FusionNet produces more complete and coherent point clouds across all scenes. By effectively restoring saturated fringe information prior to phase recovery, MA-FusionNet significantly suppresses depth artifacts induced by overexposure, resulting in smoother surfaces, sharper structural edges, and improved global geometric consistency. These results indicate that the proposed method achieves a more faithful reconstruction of the true object geometry, especially under challenging high-reflectivity conditions.

To further analyze reconstruction quality in critical regions, figure 12 presents locally enlarged views of the reconstructed surfaces corresponding to highly reflective areas. The close-up comparisons reveal that traditional and existing CNN-based methods still exhibit surface degradation, including local holes, uneven depth fluctuations, and staircase-like artifacts caused by incomplete or inaccurate phase recovery. Such errors are particularly evident along object boundaries and curved surfaces, where phase discontinuities are easily amplified during triangulation. By contrast, MA-FusionNet maintains smooth and continuous surface profiles in these challenging regions, while preserving fine geometric details

Table 2. The comparison of MAE and SSIM for the depth maps.

	Scene 1					Scene 2				
	3-step	12-step	GAN	DcMcNet	Ours	3-step	12-step	GAN	DcMcNet	Ours
MAE	0.2103	0.0556	0.0931	0.0617	**0.0478**	0.1982	0.0589	0.0873	0.0542	**0.0425**
SSIM	0.3215	0.8124	0.6732	0.7884	**0.8405**	0.3458	0.8479	0.7026	0.8207	**0.8612**
	Scene 3					Scene 4				
	3-step	12-step	GAN	DcMcNet	Ours	3-step	12-step	GAN	DcMcNet	Ours
MAE	0.1859	0.0520	0.0854	0.0586	**0.0451**	0.1937	0.0505	0.0912	0.0634	**0.0489**
SSIM	0.3372	0.7986	0.6665	0.7903	**0.8361**	0.3598	0.8263	0.6877	0.8145	**0.8596**
	Scene all (Avg)					Scene all (Absolute error)				
	3-step	12-step	GAN	DcMcNet	Ours	3-step	12-step	GAN	DcMcNet	Ours
MAE	0.1956	0.0562	0.0897	0.0605	**0.0467**	±0.0154	±0.0141	±0.0159	±0.0123	**±0.0110**
SSIM	0.3428	0.8184	0.6858	0.7942	**0.8556**	±0.0265	±0.0232	±0.0253	±0.0239	**±0.0218**

and sharp transitions. The reconstructed surfaces show significantly reduced noise and improved structural integrity, demonstrating the robustness of the proposed fringe restoration strategy. These visual results further confirm that MA-FusionNet enables high-fidelity point-cloud reconstruction and offers clear advantages for practical structured-light 3D measurement applications involving objects with strong specular reflections.

To quantitatively evaluate the reconstruction quality, we adopt the MAE (mm) and the SSIM as evaluation metrics. MAE quantifies the average absolute deviation between the predicted and ground truth depth (or phase) maps, while SSIM assesses their perceptual consistency in terms of luminance, contrast, and structure. Lower MAE and higher SSIM values indicate more accurate and visually consistent reconstruction results. It note that MAE and coverage metrics are used only for relative method comparison, not as absolute metrological accuracy indicators.

Table 2 summarizes the quantitative evaluation results of all compared methods in terms of MAE and SSIM on four representative scenes as well as the overall average performance. As can be observed, the proposed MA-FusionNet consistently achieves the lowest MAE and the highest SSIM across all scenes, indicating superior depth accuracy and structural fidelity under high-reflectivity conditions. Traditional phase-shifting methods (e.g. 3-step) exhibit large MAE values and very low SSIM scores, primarily due to severe phase errors caused by saturation and underexposure in reflective regions. Although increasing the number of phase-shifting steps (12-step) improves robustness and reduces noise to some extent, these methods remain fundamentally limited by irreversible fringe information loss and thus cannot fully recover accurate depth in distorted regions. Compared with deep learning-based baselines, the proposed method also demonstrates clear advantages. GAN-based approaches significantly reduce the error compared to traditional methods by learning local fringe completion; however, they still suffer from unstable restoration and residual artifacts, leading to relatively higher MAE and lower SSIM. DcMcNet further improves reconstruction quality through multi-scale feature modeling and achieves better accuracy, but its performance remains constrained when saturation is severe or spatial continuity is strongly disrupted. In contrast, MA-FusionNet achieves an average MAE of 0.0467, which is approximately 2–3 times lower than that of other deep learning methods, while traditional methods produce errors nearly an order of magnitude larger. Moreover, the proposed method also yields the highest SSIM value (0.8556 on average), which is closest to the ideal value of 1, demonstrating superior preservation of structural consistency and surface continuity. The reduced standard deviation of the absolute error further indicates that MA-FusionNet provides more stable and reliable reconstruction results across different scenes and reflectivity conditions.

It is worth noting that comparable reconstruction accuracy has also been reported in prior studies employing alternative measurement strategies, such as MEF or polarized-light imaging. Multi-exposure methods effectively suppress saturation by adaptively combining images captured at different exposure levels; however, they require multiple projections and acquisitions, which significantly reduce temporal efficiency and limit applicability in dynamic or real-time measurement scenarios. Polarized-light techniques can partially mitigate specular reflections, but they rely on precise optical calibration and often reduce the overall SNR due to light attenuation through polarizers. The proposed MA-FusionNet achieves competitive or superior reconstruction performance without introducing additional optical

Table 3. Ablation study results of the core module of MA-FusionNet.

Number	MGAF	DC	RAMA	MAE	SSIM
(a)	×	×	✓	0.0613	0.7634
(b)	✓	×	✓	0.0611	0.8129
(c)	×	✓	✓	0.0792	0.8291
(d)	✓	✓	×	0.0893	0.7335
(e)	✓	✓	✓	**0.0484**	**0.8632**

hardware or increasing acquisition complexity, making it more suitable for practical structured-light 3D reconstruction applications under high-reflectivity and high-dynamic-range conditions.

4.5. Ablation study

To validate the effectiveness of each key component in the proposed framework, we conduct a systematic ablation study by explicitly defining different network configurations. Each configuration corresponds to a specific modification of the full model and is summarized in table 3. All models are trained under identical settings unless otherwise specified.

Configuration (a): w/o MGAF. In this configuration, the MGAF data augmentation strategy is removed, and the network is trained directly on the original dataset using the same batch size and optimization parameters. Compared with the full model, the prediction accuracy of the phase-shifted fringes degrades significantly, indicating that MGAF effectively enriches data diversity, stabilizes training, and improves prediction accuracy.

Configuration (b): w/o DC. To evaluate the contribution of the DC module, we replace it with a standard convolutional layer while keeping all other components unchanged. This modification leads to a 0.127 increase in MAE and a 0.0503 decrease in SSIM, demonstrating that DC effectively captures local geometric distortions in fringe patterns and improves the modeling of nonlinear disturbances caused by high reflectivity.

Configuration (c): w/o RAMA. In this setting, the RAMA module is substituted with standard convolutional blocks. The resulting reconstructions exhibit blurred and discontinuous stripe boundaries, especially in high-reflectivity regions. This confirms that RAMA plays a critical role in jointly modeling local details and global consistency, significantly enhancing fringe continuity and contextual constraints.

Configuration (d): w/o DC and RAMA. To further analyze the combined effect of the DC and RAMA modules, both components are removed simultaneously while MGAF is retained. This configuration results in a more pronounced degradation in both MAE and SSIM compared with removing either module individually, indicating that DC and RAMA provide complementary benefits in modeling local distortions and global contextual consistency.

Configuration (e): Full model. The full model integrates all proposed components, including MGAF, DC, and RAMA. It consistently achieves the best performance across all evaluation metrics, demonstrating the synergistic effect of the proposed modules.

We calculated the average MAE and SSIM for 10 scenes. The quantitative results of the ablation experiments are summarized in table 3. As shown, the complete model achieves the best performance across all metrics, highlighting the complementary nature and synergistic effect of the proposed modules.

5. Conclusion

5.1. Conclusion

This paper proposes MA-FusionNet, a robust and effective fringe restoration framework tailored for 3D SL measurement of highly reflective objects. The method is composed of two key modules: the MGAF strategy and the Reflective-RAMA module. The former enhances dataset diversity and phase completeness by fusing reliable fringe information from reference images, while the latter combines local convolutional attention with Transformer-based global context modeling to effectively restore structurally distorted regions.

Extensive experiments on both simulated and real-world datasets validate the superior performance of the proposed method. Quantitatively, MA-FusionNet achieves significant improvements in MAE and SSIM for depth maps, while also obtaining the highest point cloud coverage across all evaluated methods. Qualitatively, the restored fringe patterns exhibit high visual fidelity and structural consistency, even in the presence of severe specular reflections. The ablation study further demonstrates the complementary effects of the MGAF, DC, and RAMA modules, substantiating the robustness and generalizability of the framework. Collectively, these results highlight MA-FusionNet's potential for real-world deployment in industrial inspection, digital twin modeling, and reverse engineering applications under HDR conditions.

5.2. Limitations and future work

Despite the demonstrated effectiveness, several limitations remain to be addressed:

1) Dependence on reference images in MGAF: the current fusion-based data augmentation strategy assumes the availability of multiple reference images with high-quality fringe patches. In scenarios where sufficient reference diversity is lacking, such as in single-shot or real-time applications, MGAF's applicability may be limited.
2) Computational overhead of RAMA: while the integration of Transformer blocks improves global context modeling, it inevitably introduces additional computational and memory burdens. This may hinder deployment in lightweight or real-time systems with stringent latency constraints.
3) Generalization to ultra-complex geometries: although tested on various reflective objects, further exploration is needed to validate the performance on geometrically complex or textureless surfaces, where fringe deformation becomes more nonlinear and irregular.
4) Performance on complex geometries: although the proposed method shows improved robustness on reflective objects, handling complex geometries such as free-form surfaces with sharp edges remains challenging. Abrupt depth discontinuities and severe self-occlusions may still introduce local phase ambiguities, especially under strong specular reflections. While the DC and RAMA modules help alleviate these issues by modeling local distortions and global context, further improvements, such as edge-aware constraints or adaptive sampling strategies, are left for future work.

To address these limitations, future research will consider incorporating self-supervised or unsupervised domain adaptation techniques, designing lightweight network variants for edge deployment, and exploring multimodal data fusion (e.g. polarization or ToF data) to further enhance robustness under extreme reflection conditions. Additionally, leveraging temporal redundancy in fringe video sequences may offer promising directions for dynamic surface measurement.

Acknowledgments

This work was supported by the National Natural Science Foundation of China(62376252); Zhejiang Province Leading Geese Plan(2025C02025,2025C01056); Zhejiang Province Province-Land Synergy Program(2025SDXT004-3); General research projects of Zhejiang Provincial Department of Education(Y202557906).

Data availability statement

The synthetic dataset and the associated generation pipeline are closely tied to our ongoing and future research directions and are therefore available upon reasonable request from the authors. The real-world experiments that support the findings of this study are included within the article. The data that support the findings of this study are available upon reasonable request from the authors.

Author contributions

Chengcheng Li 0009-0008-8243-8618
Conceptualization (equal), Data curation (equal), Resources (equal), Software (equal), Validation (equal)

Yan Zhang
Visualization (equal)

Tao Wang
Data curation (equal), Resources (equal)

Lan Yu
Formal analysis (equal), Investigation (equal), Software (equal)

De'ang Su
Methodology (equal), Writing – original draft (equal)

Zhendong Chen
Formal analysis (equal), Visualization (equal)

Xinzhong Zhu 0000-0002-0033-5260
Data curation (equal), Formal analysis (equal), Visualization (equal)

References

[1] Geng J 2011 Structured-light 3D surface imaging: a tutorial *Adv. Opt. Photon.* **3** 128–60
[2] Chen S, Tao W, Zhao H and Lv N 2021 A review: application research of intelligent 3D detection technology based on linear-structured light *Trans. on Intelligent Welding Manufacturing: Volume III No. 4 2019* pp 35–45
[3] Wang L, Cai J, Zhang Y, Chen D, Ge L and Zhang Y 2024 Application of structured light 3D reconstruction technology in industrial automation scenarios in the context of digital transformation *SHS Web of Conferences* vol 181 (EDP Sciences) p 03026
[4] Van der Jeught S and Dirckx J J 2016 Real-time structured light profilometry: a review *Opt. Lasers Eng.* **87** 18–31
[5] Savage C 1997 A survey of combinatorial Gray codes *SIAM Rev.* **39** 605–29
[6] Zhao X, Yu T, Liang D and He Z 2024 A review on 3D measurement of highly reflective objects using structured light projection *The Int. J. Adv. Manuf. Technol.* **132** 4205–22
[7] Palousek D, Omasta M, Koutny D, Bednar J, Koutecky T and Dokoupil F 2015 Effect of matte coating on 3D optical measurement accuracy *Opt. Mater.* **40** 1–9
[8] Cai Z, Liu X, Pedrini G, Osten W and Peng X 2020 Structured-light-field 3D imaging without phase unwrapping *Opt. Lasers Eng.* **129** 106047
[9] Lin H, Gao J, Mei Q, Zhang G, He Y and Chen X 2017 Three-dimensional shape measurement technique for shiny surfaces by adaptive pixel-wise projection intensity adjustment *Opt. Lasers Eng.* **91** 206–15
[10] Cheng T, Qin L, Li Y, Hou J and Xiao C 2024 An adaptive multiexposure scheme for the structured light profilometry of highly reflective surfaces using complementary binary Gray code *IEEE Trans. Instrum. Meas.* **73** 1–12
[11] Sun J and Zhang Q 2022 A 3D shape measurement method for high-reflective surface based on accurate adaptive fringe projection *Opt. Lasers Eng.* **153** 106994
[12] Xu P, Liu J and Wang J 2023 High dynamic range 3D measurement technique based on adaptive fringe projection and curve fitting *Appl. Opt.* **62** 3265–74
[13] Tang S, Zhang X, Li C and Gu F 2019 High dynamic range three-dimensional shape reconstruction via an auto-exposure-based structured light technique *Opt. Eng., Bellingham* **58** 064108
[14] Lee S and Park H 2024 Automatic setting of optimal multi-exposures for high-quality structured light depth imaging *IEEE Access* **12** 152786–97
[15] Tan J, Zou T, Chen H, Su W and He Z 2026 Spectral division multiplexing-based region projection for 3D measurement under mutual reflection *Opt. Lett.* **51** 488–91
[16] Tan J, Su W, He Z, Bai Y, Dong B and Xie S 2022 Generic saturation-induced phase error correction for structured light 3D shape measurement *Opt. Lett.* **47** 3387–90
[17] Lai B-Y and Chiang P-J 2023 Improved structured light system based on generative adversarial networks for highly-reflective surface measurement *Opt. Lasers Eng.* **171** 107783
[18] Purwono P, Wulandari A N E, Ma'arif A and Salah W A 2025 Understanding generative adversarial networks (gans): a review *Control Syst. Optim. Lett.* **3** 36–45
[19] Song X, Zhang S and Wu Y 2024 An accurate measurement of high-reflective objects by using 3D structured light *Measurement* **237** 115218
[20] Li W, Yan Y, Lin H and Feng Z 2025 Three-dimensional surface structure reconstruction of reflective objects using multi-stage deep learning *Opt. Rev.* **32** 1–13
[21] Li Y, Li Z, Chen W, Zhang C, Wang H, Wang X, Gui W, Gao W, Liang X and Li X 2025 Reliable 3D reconstruction with single-shot digital grating and physical model-supervised machine learning *IEEE Trans. Instrum. Meas.* **74** 5505413
[22] Tan J, Liu J, Wang X, He Z, Su W, Huang T and Xie S 2024 Large depth range binary-focusing projection 3D shape reconstruction via unpaired data learning *Opt. Lasers Eng.* **181** 108442
[23] Wang S, Zhou F, Zhang W, Wang X and Zhang L 2025 Deep self-supervised learning network for multi-exposure image fusion in fringe structured-light 3D reconstruction *IEEE Trans. Instrum. Meas.* **74** 5025012
[24] Yang G, Yang M, Zhou N and Wang Y 2022 High dynamic range fringe pattern acquisition based on deep neural network *Opt. Commun.* **512** 127765
[25] Li Y, Li Z, Zhang C, Han M, Lei F, Liang X, Wang X, Gui W and Li X 2023 Deep learning-driven one-shot dual-view 3-D reconstruction for dual-projector system *IEEE Trans. Instrum. Meas.* **73** 1–14
[26] Zhu X, Zhang Z, Hou L, Song L and Wang H 2022 Light field structured light projection data generation with blender *2022 3rd Int. Conf. on Computer Vision, Image and Deep Learning & Int. Conf. on Computer Engineering and Applications (CVIDL & ICCEA)* (IEEE) pp 1249–53

[27] Kim W-H, Kim B, Chi H-G and Hyun J-S 2024 Novel approach for fast structured light framework using deep learning *Image Vis. Comput.* **150** 105204
[28] Nguyen H and Wang Z 2021 Accurate 3D shape reconstruction from single structured-light image via fringe-to-fringe network *Photonics* **8** 459
[29] Wu Z, Wang J, Jiang X, Fan L, Wei C, Yue H and Liu Y 2023 High-precision dynamic three-dimensional shape measurement of specular surfaces based on deep learning *Opt. Express* **31** 17437–49
[30] Bans-Akutey A and Tiimub B M 2021 Triangulation in research *Acad. Lett.* **2** 1–7
[31] Zhang Z 2002 A flexible new technique for camera calibration *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 1330–4
[32] Meza J, Vargas R, Romero L A, Zhang S and Marrugo A G 2020 What is the best triangulation approach for a structured light system? *Proc. SPIE* **11397** 65–74
[33] Hyun J-S and Zhang S 2016 Enhanced two-frequency phase-shifting method *Appl. Opt.* **55** 4395–401
[34] Sansoni G, Carocci M and Rodella R 1999 Three-dimensional vision based on a combination of Gray-code and phase-shift light projection: analysis and compensation of the systematic errors *Appl. Opt.* **38** 6565–73
[35] Qican Z and Zhoujie W 2020 Three-dimensional imaging technique based on Gray-coded structured illumination *Infrared Laser Eng.* **49** 0303004
[36] He X D, Torrance K E, Sillion F X and Greenberg D P 1991 A comprehensive physical model for light reflection *ACM SIGGRAPH Comput. Graph.* **25** 175–86
[37] Cao J, Wang X, Zhang X and Tu D 2025 Fringe projection profilometry with lightweight transformer-CNN network for high-dynamic-range scenes *Meas. Sci. Technol.* **36** 105202
[38] Song Z, Jiang H, Lin H and Tang S 2017 A high dynamic range structured light means for the 3D measurement of specular surface *Opt. Lasers Eng.* **95** 8–16