applied optics

Underwater structured-light 3D imaging method based on FP-DiffNet

Lei Lu,[1] Yuheng Wang,[2] Zhilong Su,[3,*] Haojun Zhang,[2] Wei Pan,[4] and Peng Li[1]

[1]Institute for Complexity Science, Henan University of Technology, Zhengzhou 450001, China
[2]College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
[3]Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai University, Shanghai 200444, China
[4]Department of Research and Development, OPT Machine Vision Tech Co., Ltd., Dongguan 523860, China
*szloong@shu.edu.cn

Received 5 January 2026; revised 24 February 2026; accepted 2 March 2026; posted 2 March 2026; published 17 March 2026

Underwater structured-light three-dimensional (3D) imaging is essential for the precise reconstruction of submerged objects in scientific and engineering applications. However, the observed fringe patterns are often degraded by underwater light attenuation, scattering, and refractive distortion, leading to fringe blurring and aliasing and, consequently, reduced phase-reconstruction accuracy. To address these issues, we analyze the propagation characteristics of structured light in underwater environments and their effects on fringe image quality. With that, we engineer a diffusion-model-based fringe restoration neural framework, referred to as FP-DiffNet, in which fringe restoration is formulated as a probability-driven iterative denoising process. Specifically, a U-Net model is trained to progressively map degraded fringe observations to clear fringe patterns, incorporating physics-guided constraints and an adaptive noise-annealing mechanism to effectively separate scattering-induced noise while preserving fringe structures during the reverse diffusion process. The proposed framework is evaluated in terms of fringe image quality, wrapped-phase accuracy, and 3D reconstruction performance. Under extremely high turbidity, it achieves a PSNR of 42.42 dB and a wrapped-phase MAE of 0.0354 rad. By combining diffusion-driven phase reconstruction with dynamic compensation, sub-millimeter absolute measurement is achieved in turbid water, with a 3D reconstruction RMSE below 0.1 mm. Importantly, no prior modeling of water optical parameters is required, enabling robust and practical underwater 3D imaging.

https://doi.org/10.1364/AO.589434

1. INTRODUCTION

As a representative technique in three-dimensional (3D) reconstruction, structured-light 3D imaging has played an important role in industrial inspection, medical imaging, and robotic vision because it enables non-contact measurement with high accuracy and high efficiency [1–3]. The underlying principle is active optical triangulation. Specifically, encoded grating fringe patterns are projected onto the target surface by a digital projector, and the deformed fringes modulated by the surface geometry are synchronously captured by a high-resolution industrial camera. Three-dimensional point-cloud coordinates are then obtained via phase-to-height mapping. As a classic method in this paradigm, phase-shifting profilometer (PSP) sequentially projects multiple sinusoidal fringe patterns with fixed phase offsets. A phase-shifting algorithm is used to suppress ambient-light interference and extract the wrapped phase, after which a phase-unwrapping algorithm reconstructs the continuous phase distribution. Micrometer-level measurement accuracy can be achieved, making PSP highly practical for precision manufacturing and reverse engineering [4–6].

Notably, the theoretical foundations and experimental validation of existing structured-light 3D reconstruction methods have largely been established under air as the measurement medium. When the medium changes, light propagation across refractive interfaces is altered, which induces systematic measurement errors. In underwater structured-light imaging, multiple degradation factors coexist, including refraction, scattering, and absorption. Refraction at the water–air interface causes a nonlinear deviation between the measured fringe phase and the true height, while Mie scattering induced by suspended particles markedly reduces fringe-image contrast [7]. In addition, exponential attenuation in water results in the rapid decay of effective signal intensity with increasing propagation distance [8,9]. In extreme conditions such as highly turbid water, the coupling of these effects results in complex phase-error accumulation, which can cause local point-cloud loss or geometric distortion in 3D measurements.

Addressing underwater fringe degradation is, therefore, beneficial for extending structured-light techniques to cross-medium measurement scenarios.

Conventional approaches to underwater image degradation can be broadly categorized into restoration methods, represented by polarization imaging [10] and stereo imaging [11], and enhancement methods, represented by histogram equalization [12] and color correction [13]. However, these approaches exhibit clear limitations-restoration methods rely heavily on prior modeling of water optical parameters, whereas enhancement methods may introduce color bias and additional noise [14–16]. With the rapid advancement of deep learning, its strengths in extracting and representing complex data features have provided an alternative pathway for underwater image restoration and enhancement. Accordingly, deep-learning-based solutions have gained increasing attention for addressing key challenges in underwater 3D reconstruction, including light-transport distortion and medium-scattering interference [17,18]. Current studies can be broadly grouped into two routes: physics-prior-based underwater image restoration and data-driven end-to-end model construction for underwater image restoration and enhancement [19].

When combined with physical modeling, WaterGAN was proposed by Li *et al.* [20] to translate RGB-D images into underwater images via unsupervised learning. By incorporating a light-transport model into adversarial training, WaterGAN enables underwater color restoration and depth estimation without annotated data. Mousavi *et al.* [21] proposed iDehaze, a two-stage deep-learning framework for underwater image dehazing and color correction. In iDehaze, large-scale synthetic data generated by a physical simulation model are used to train a dehazing network, and a color-transfer model maps the dehazed outputs to a target color domain, thereby improving sharpness and color fidelity. Nevertheless, when deep learning is coupled with explicit physical models, the imposed constraints may reduce adaptability. In contrast, data-driven end-to-end methods that do not explicitly depend on prior knowledge often exhibit superior adaptability [22]. For example, Xinya *et al.* [23] proposed a deep-learning-based underwater fringe-projection 3D measurement method that introduces a fringe pattern enhancement convolutional neural network (FPENet) to optimize low-contrast, high-noise fringe images, effectively alleviating turbidity-induced quality degradation and phase errors. Jianxin *et al.* [24] proposed a lightweight generative adversarial network with dual attention (AL-GAN), in which PatchGAN is used as the discriminator and MobileNet is adopted as the generator, substantially reducing the parameter count and improving inference efficiency while enhancing underwater detail restoration. However, such approaches treat neural networks as black-box end-to-end mappings. The generated results often lack explicit constraints for physical consistency, which may distort restored fringe patterns and consequently introduce systematic errors into high-precision phase retrieval.

To overcome this limitation, a fringe pattern restoration framework based on deep diffusion models, termed FP-DiffNet (Fringe Pattern Diffusion Network), is proposed in this study. Generative deep learning has been effectively integrated into structured-light imaging to boost system performance [25], among which diffusion models exhibit remarkable advantages in training stability and geometric fidelity. Under a Markov-chain formulation, iterative denoising can decouple medium-scattering noise from target geometric features, providing a principled basis for mitigating pixel-level distortions under non-uniform illumination and promoting underwater optical metrology toward higher robustness and practicality [26,27]. In FP-DiffNet, physics guidance is incorporated into the diffusion process to jointly leverage data-driven adaptability and physically consistent constraints. Specifically, an adaptive noise-annealing mechanism is introduced to progressively decouple scattering noise during reverse diffusion. Meanwhile, physical constraints and a periodic-feature modulation mechanism are employed to enforce the correct intensity distribution and periodic structure in the restored fringes, thereby enabling more reliable downstream phase estimation. Building on this design, the pipeline of our method comprises three main stages: fringe projection and acquisition, FP-DiffNet-based fringe restoration to suppress underwater degradation, and phase retrieval followed by 3D reconstruction from the recovered phase. With this approach, sub-millimeter absolute measurement accuracy is achieved in turbid water, establishing a robust paradigm for underwater 3D imaging.

2. DIFFUSION-MODEL-BASED FRINGE RESTORATION ALGORITHM

A. Principle of Underwater Structured-Light Imaging

In underwater structured-light imaging, optical propagation is affected by multiple coupled effects. The radiometric degradation caused by scattering and attenuation can be characterized by the classical Jaffe–McGlamery (J–M) model [28]. According to J–M theory, underwater light can be decomposed into three components: direct transmission I_d, forward scattering I_f, and backward scattering I_b, as illustrated in Fig. 1. The direct component $I_d(z)$ is not scattered, and its intensity attenuation follows a modified Beer–Lambert law [29]:

$$I_d(z) = I_0 e^{-[a(\lambda)+b_b(\lambda)+b_f(\lambda)]z} \cdot \frac{1}{(1+\eta z)^2}, \tag{1}$$

where $a(\lambda)$ denotes the absorption coefficients, $b_b(\lambda)$ and $b_f(\lambda)$ denote the backward-scattering coefficients and forward-scattering coefficients, respectively, and $1/(1+\eta z)^2$ accounts for geometric spreading. Forward-scattered light is produced by small-angle scattering from suspended particles, which induces spatial blurring of the optical field. This effect can be modeled by a point spread function (PSF) represented as a Gaussian kernel:

$$h_{\mathrm{PSF}}(r) = \frac{1}{2\pi\sigma_s^2} e^{-r^2/(2\sigma_s^2)}, \quad \sigma_s \propto b_f z, \tag{2}$$

where σ_s quantifies the blur strength. Its magnitude increases with the forward-scattering coefficient $b_b(\lambda)$ and the propagation distance z, reflecting the low-pass filtering effect of water turbidity on the spatial-frequency content of the captured image. Backward-scattered light I_b is generated by large-angle scattering and forms background noise. It is commonly modeled by a Poisson-Gaussian mixed distribution:

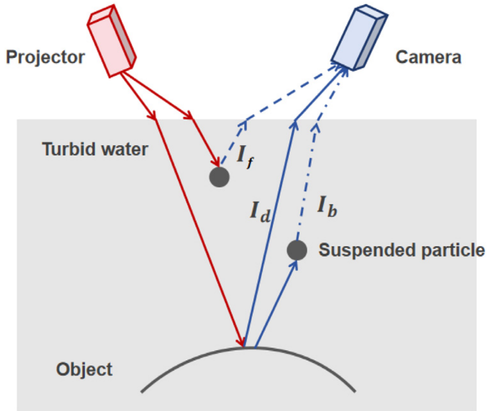

Fig. 1. Schematic illustration of the light propagation in turbid underwater environments.

$$I_b \sim \mathcal{P}\left(k I_0 e^{-b_b z}\right) + \mathcal{N}\left(0, \sigma_{read}^2\right). \quad (3)$$

The scattering anisotropy of suspended particles in turbid water can be described by the Henyey–Greenstein phase function [30]:

$$p(\theta) = \frac{1}{4\pi} \frac{1-g^2}{\left(1+g^2-2g\cos\theta\right)^{3/2}}, \quad (4)$$

where the anisotropy factor $g \approx 0.9$ indicates the dominance of forward scattering, which further leads to low-pass-filter-like degradation of the fringe patterns.

In addition, the refractive-index mismatch at the water–air interface ($n_w \approx 1.33$, $n_a = 1.0$) causes geometric distortion. The refracted optical path should be corrected using Snell's law [31]:

$$n_w \sin\theta_i = n_a \sin\theta_t. \quad (5)$$

As indicated by the above relationship, refraction introduces depth-dependent geometric displacement of light rays $\Delta x = z \cdot \tan(\theta_i - \theta_t)$, which directly induces 3D reconstruction errors.

By integrating the above effects, the observed signal of a projected structured-light fringe pattern after underwater propagation can be modeled as

$$I_{obs}(x) = I_d(x) \otimes h_{\mathrm{PSF}} + I_b(x) + G(x+\Delta x), \quad (6)$$

where $\otimes$ denotes the convolution operator, and $G(x+\Delta x)$ denotes the refractive-distortion function. The coupling of these degradation mechanisms can lead to phase discontinuities, phase-unwrapping failures, and nonlinear errors in depth mapping during phase retrieval, thereby significantly reducing the accuracy of underwater 3D imaging. For clarity, the key symbols involved in the physical modeling and diffusion process are summarized in Table 1.

B. Diffusion-Model-Based Fringe Image Restoration

To accurately model the mixed noise and phase distortion in fringe patterns and to generate high-precision phase distributions, a deep diffusion-model-based fringe restoration framework, termed FP-DiffNet (Fringe Pattern Diffusion Network), is proposed in this study. Given the characteristics of the task, the denoising diffusion probabilistic model (DDPM) [32,33] is adopted as the core formulation. The forward process explicitly simulates the coupled degradation of mixed noise and phase distortion in fringe images, and the noise-scheduling parameters can be adaptively adjusted according to the spectral characteristics of fringes. The reverse process restores fringe patterns under physics-guided constraints. In particular, gradient regularization and a periodic-feature modulation mechanism are incorporated to enforce smooth intensity variation and suppress discontinuous jump artifacts [34–36]. Compared with conventional approaches, the proposed generation scheme integrated with physical constraints substantially improves fringe-structure integrity and reconstruction robustness.

Table 1. Key Symbols Used in the Proposed Framework

Symbol	Description
I_d, I_b, I_f	Direct transmission intensity, backward-scattering intensity, and forward-scattering intensity
a, b_f, b_b	Absorption coefficients, forward coefficients, and backward scattering coefficients
h_{PSF}	Scattering-induced point spread function
X_0, X_t	Clean and diffused fringe images
$\varepsilon, \varepsilon_\theta$	True and predicted diffusion noise
α_t	Diffusion noise scheduling coefficient
$D(\cdot)$	Simplified physical degradation operator

The key idea of FP-DiffNet is to progressively emulate underwater degradation during forward diffusion and to gradually recover the spatial structure and phase information during reverse diffusion via a neural network. In the forward diffusion process, given a fringe sample X_0, Gaussian noise is gradually injected from $t=0$ to $t=T$, so that the sample is degraded toward a near-Gaussian distribution. Unlike the standard DDPM, the proposed formulation is tailored to underwater fringe degradation. Specifically, the injected noise not only represents random Gaussian perturbations but also implicitly accounts for physical degradations in underwater imaging, including scattering-induced blur, ambient-light interference, and local contrast reduction. Therefore, at each forward step, the noise term can be interpreted as the progressive superposition of forward- and backward-scattering effects, and the statistical distribution of the intermediate state X_t is driven to approximate the distribution of fringe images observed in real turbid water. The forward diffusion distribution $q(X_t|X_{t-1})$ is given by

$$q(X_t | X_0) = N\left(X_t; \sqrt{\overline{\alpha_t}} X_0, (1-\overline{\alpha_t}) I\right), \quad (7)$$

where $\overline{\alpha_t}$ is a predefined or adaptive coefficient. Accordingly, direct sampling of X_t at an arbitrary diffusion step can be written as

$$X_t = \sqrt{\overline{\alpha_t}} X_0 + \sqrt{1-\overline{\alpha_t}} \varepsilon, \quad (8)$$

where ε corresponds to an approximation of scattering noise affecting underwater fringe images under different depths and turbidity levels.

The reverse diffusion process starts from standard Gaussian noise and progressively removes the noise components associated with underwater scattering and attenuation, thereby

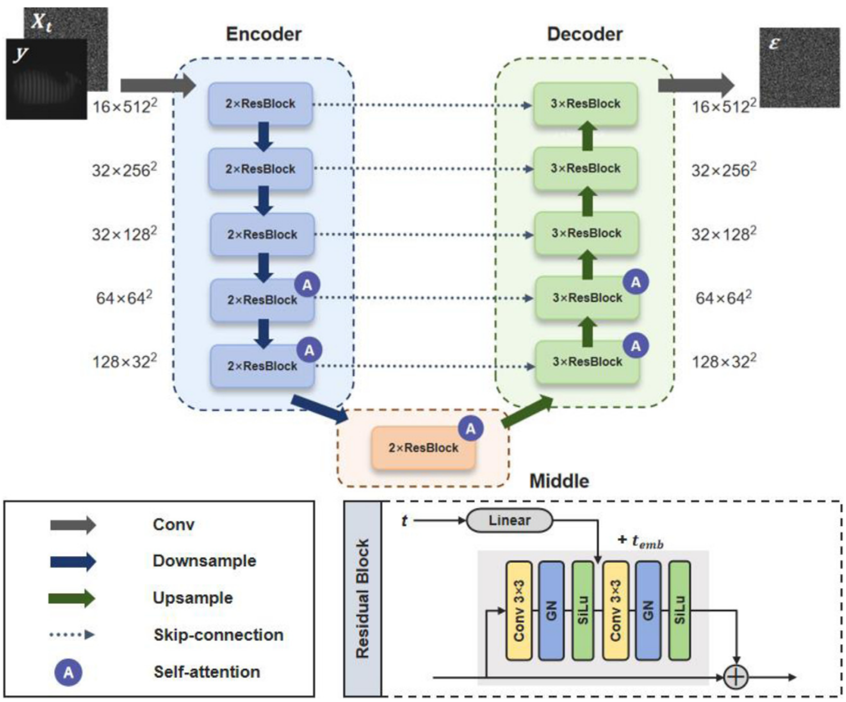

Fig. 2. Schematic representation of the proposed network architecture.

generating a high-quality approximation of structured-light fringe patterns. The reverse Markov chain is formulated as

$$p_\theta(X_{t-1} \mid X_t) = N(X_{t-1}; \mu_\theta\,(X_t, t)\,,\, \Sigma_t), \tag{9}$$

where the mean term μ_θ is obtained by predicting the noise $\varepsilon_\theta(X_t, t)$ using a neural network. In accordance with the degradation characteristics of underwater fringe images, the network is trained to capture not only generic texture perturbations but also scattering-induced fringe blurring, phase warping, and spatially varying illumination. Consequently, the periodic structure and contrast of fringe patterns are progressively recovered during reverse diffusion.

In this study, to implement the above denoising diffusion process, a U-Net-based deep neural network is employed to represent ε_θ, as illustrated in Fig. 2. The encoder extracts multi-scale features of underwater fringe images, including amplitude variations, contrast attenuation, and detail blurring. The decoder progressively recovers high-frequency fringe structures during upsampling and compensates for local degradation induced by scattering. With multi-scale feature fusion and skip connections, the network captures both the periodic characteristics of fringes and the spatial distribution of underwater noise.

After encoding, the input tensor is first mapped to a base number of channels via a 3×3 convolution and then fused with the time-step embedding and the conditional observed-fringe image, so that each diffusion stage is guided by the real underwater observation. Subsequently, hierarchical downsampling modules compress the spatial resolution and extract blur-related features caused by scattering. During decoding, skip connections and local convolutional refinement are jointly used to restore fringe details and phase-related structures, ultimately producing a high-quality reconstructed fringe image with the same spatial resolution as the input.

To incorporate physics-consistent constraints into the diffusion process, the imaging degradation is modeled using a differentiable simplification based on the two dominant radiometric effects in the J–M model described in Section 1.A, namely forward scattering and backward scattering. Because the full physical model [Eqs. (1)–(6)] is complex and not easily embedded into the stepwise diffusion framework, the most influential factors on fringes, scattering blur and background noise, are approximated and unified with the diffusion-noise mechanism. Accordingly, a simplified physical degradation operator is defined as

$$D(I) = \left(I \otimes h_g\right) + n_g, \tag{10}$$

where the Gaussian kernel h_g approximates the point spread function associated with forward scattering and the noise term n_g characterizes the stochastic component of backward scattering. The geometric displacement induced by refraction is a spatial deformation rather than a purely radiometric degradation. Therefore, it is not explicitly modeled in $D(\cdot)$ and is instead implicitly captured by the spatial feature learning capability of the U-Net. Based on the above simplification, a physics-consistency constraint is formulated as

$$L_{phy} = \| D(X_{t-1}) - D(X_0) \|_1\,. \tag{11}$$

To train the model shown in Fig. 2, an underwater structured-light system is used to acquire a dataset of degraded fringe images $\{y_i\}_{i=1}^N \sim q(S)$ as conditional inputs, which are affected by scattering and refraction in turbid water. High-SNR fringe images captured in air, $X_0 \sim q(F \mid y)$, are simultaneously collected as the target distribution. In the forward diffusion process

[Eq. (7)], a Markov chain consistent with underwater optical transport is constructed. By progressively injecting Gaussian perturbations that conform to underwater degradation mechanisms, the intermediate states are driven to approximate the spatial-frequency characteristics of real underwater fringe observations.

Under the physical constraints for cross-medium fringe enhancement, an optimization objective is constructed by quantifying the time-varying discrepancy between the degradation noise ε_t injected at diffusion step t and the network-predicted noise $\varepsilon_\theta(X_t, t, y)$. The resulting loss function is defined as

$$L = \lambda_1 \mathrm{SSIM}(X, \hat{X}) + \lambda_2 \mathrm{E}\left[|\varepsilon_t - \varepsilon_\theta(X_t, t, y)|\right] + \lambda_3 \|\theta\|_1 + \lambda_4 \left\| D(X_{t-1}) - D(X_0) \right\|_1, \tag{12}$$

where L jointly accounts for perceptual structural similarity, latent-variable reconstruction error, model sparsity, and physics consistency. By combining the SSIM term, L1 regularization, and the physics-residual penalty, the objective achieves a balanced optimization across image-structure fidelity, latent modeling capability, model-complexity control, and physical plausibility.

The hyperparameters λ_i are used to adjust the relative weights of each term and are configured for the turbid underwater fringe-imaging scenario. In our implementation, $\lambda_1 = 0.7$ is assigned to enforce the structural similarity constraint, which is designed to ensure that the recovered fringe patterns maintain a high degree of resemblance to the original clear fringe patterns, thereby preserving fringe integrity in turbid media; $\lambda_2 = 0.8$ is adopted to moderately regulate the weight of the pixel-wise fidelity term, which promotes the accuracy of intensity restoration and suppresses outliers, thus avoiding excessive amplification of backscatter noise under high turbidity and enhancing the robustness under low-SNR underwater conditions; $\lambda_3 = 0.02$ imposes relatively strong L1 regularization to suppress overfitting to high-noise data and to improve generalization under low-SNR conditions; $\lambda_4 = 0.5$ introduces a moderate physics-consistency constraint, ensuring that the generated results conform to the fundamental physics of underwater light transport while maintaining a balance between network learning capacity and physical priors. Experimental results indicate that this parameter setting effectively alleviates fringe blurring and distortion in turbid water and substantially improves cross-medium optical imaging quality.

The training and sampling procedures are illustrated in Fig. 3. During training, a noise vector ε is first randomly sampled from a standard Gaussian distribution. The target image X_0 and its corresponding degraded observation y are prepared as the basic inputs for each iteration. Next, a diffusion time step t is uniformly sampled from the interval $[1, T]$. Using the reparameterization of the forward diffusion process, the intermediate state X_t is obtained by linearly combining the sampled noise ε and the clean sample X_0 with the predefined noise-scheduling coefficients $\sqrt{\bar{\alpha}_t}$ and $\sqrt{1-\bar{\alpha}_t}$. The intermediate state X_t, together with the condition y and the time embedding t, is then fed into the training network p_θ, which outputs the predicted noise component ε^*. Finally, the customized loss function L is used to quantify the discrepancy between the ground-truth noise ε and the predicted noise ε^*. Model parameters

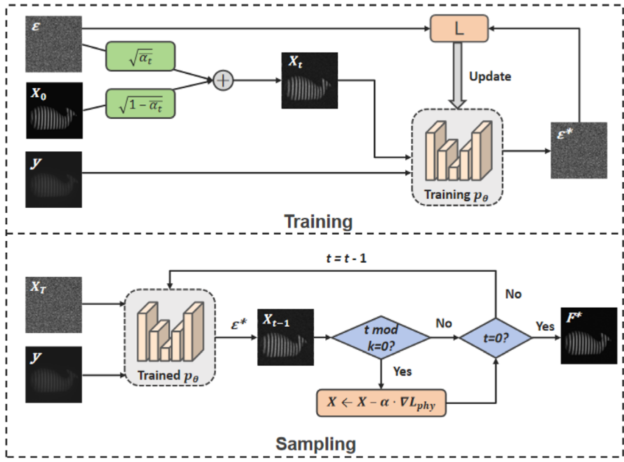

Fig. 3. Training and sampling pipeline.

are updated via backpropagation using a stochastic gradient descent optimizer, and the iterative optimization continues until convergence.

During sampling, the procedure starts from the terminal time step T. The initial state X_t is set to a random noise sample drawn from a standard normal distribution, and the time index is initialized to T. At each reverse step t, the noisy state X_t, the corresponding time encoding t, and the conditional observation y are jointly fed into the trained denoising network p_θ, which produces an estimate ε^* of the noise component at time t. Based on ε^* and the reverse-diffusion formulation, the latent state at the previous step, X_{t-1}, is then computed.

To further improve physical consistency, a periodic gradient correction mechanism is introduced. When the time-step index t is divisible by the physical correction period k, i.e., $t \bmod k = 0$, a physics-residual-based gradient correction is applied to the current state:

$$X_{t-1} = \begin{cases} X_{t-1} - \alpha \cdot \nabla_X \left\| I_{obs} - G_{simple}(X) \right\|_2 & \text{if } t \bmod k = 0 \\ X_{t-1} & \text{else} \end{cases}, \tag{13}$$

where $\alpha = 0.01$ is the correction strength, $k = 50$ is the correction period, optimally set to balance the strength of physical consistency constraints and computational efficiency, and $G_{simple}(X)$ denotes the simplified physical degradation function. By injecting physical priors at selected diffusion stages, this correction strategy achieves a balance between the strength of physical constraints and computational efficiency. Consequently, the generated results better conform to underwater light-transport physics, while the additional computational overhead remains acceptable. By progressively decreasing the time index and repeating the above inference steps until $t = 0$, a fully denoised, high-quality reconstructed fringe image F^* is finally obtained.

C. Phase Retrieval and 3D Reconstruction

The output F^* is essentially a restored fringe-pattern sequence on the object surface, whose mathematical form is consistent with that of the standard N-step phase-shifting profilometry

(PSP). To further obtain surface phase information and perform 3D reconstruction, each restored fringe frame in F^* is substituted into the standard N-step PSP formulation.

In this method, the fringe intensity on an arbitrary surface S can be uniformly expressed as

$$I_n^S(x, y) = A + B\cos\left(\phi^S(x, y) + \frac{2\pi(n-1)}{N}\right), \quad S \in \{R, O\}, \tag{14}$$

where $n = 1, 2, \cdots, N$ and $I_n^S(x, y)$ denotes the intensity distribution of the nth fringe image. When $S = O$, it corresponds to the object-surface fringe pattern in F^*, and when $S = R$, it corresponds to the reference-plane fringe pattern. The term A represents the ambient-light intensity, B is the modulation amplitude of the sinusoidal fringe, and $\phi^S(x, y)$ is the phase distribution associated with surface S.

When the number of acquired fringe images satisfies $N \geq 3$, the phase $\phi^S(x, y)$ can be computed using a unified arctangent formulation:

$$\phi^S(x, y) = -\arctan\frac{\sum_{n=1}^{N} I_n^S(x, y)\sin 2\pi(n-1)/N}{\sum_{n=1}^{N} I_n^S(x, y)\cos 2\pi(n-1)/N}, \quad S \in \{R, O\}, \tag{15}$$

where the resulting $\phi^R(x, y)$ and $\phi^O(x, y)$ are wrapped phases, exhibiting periodic discontinuities over $[-\pi, \pi]$. A phase-unwrapping algorithm is then applied to obtain the absolute phases $\Phi^{\mathrm{R}}(x, y)$ and $\Phi^{\mathrm{O}}(x, y)$, which vary monotonically. The phase difference is subsequently computed as

$$\Phi(x, y) = \Phi^{\mathrm{O}}(x, y) - \Phi^{\mathrm{R}}(x, y), \tag{16}$$

yielding the equivalent phase shift $\Phi(x, y)$ induced by object height modulation. With the calibrated geometric parameters of the projector–camera system, $\Phi(x, y)$ can be converted into 3D point-cloud coordinates, enabling high-accuracy reconstruction of the object-surface geometry [37].

3. EXPERIMENTS AND EVALUATIONS

To validate the accuracy and effectiveness of the trained FP-DiffNet in practical measurements, an underwater 3D measurement system based on structured-light principles was constructed. As shown in Fig. 4(a), the hardware setup consists of a high-resolution industrial camera (OPT-CC1-048A-RM-G, 1800 × 1600 pixels) and a digital projector (TJ-S50, 1280 × 1024 pixels). Model training was performed on a computing platform equipped with an Intel Xeon Silver 4215R CPU (3.20 GHz), 36 GB memory, and an NVIDIA GeForce RTX 4090 GPU.

A. Dataset

In our dataset, fringe patterns were generated and projected using a three-frequency four-step phase-shifting scheme. Specifically, the projector sequentially projected fringe patterns with spatial frequencies $f_1 = 1/59$, $f_2 = 1/64$, and $f_3 = 1/70$. During acquisition, an effective region of 512 × 512 pixels was cropped to construct the dataset. The object under test was placed in a transparent acrylic tank with dimensions 0.4 m × 0.4 m × 0.3 m. The measurement apparatus was mounted directly above the tank, and an approximate working distance of 0.6 m was maintained to cover the target measurement area. To minimize the influence of external ambient light, all images were captured in a controlled laboratory environment with minimized ambient illumination. The structured-light projector served as the only active light source during data acquisition.

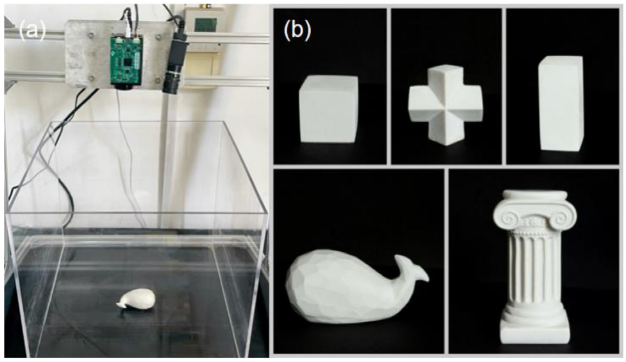

Fig. 4. (a) Experimental setup and (b) target objects used for dataset construction.

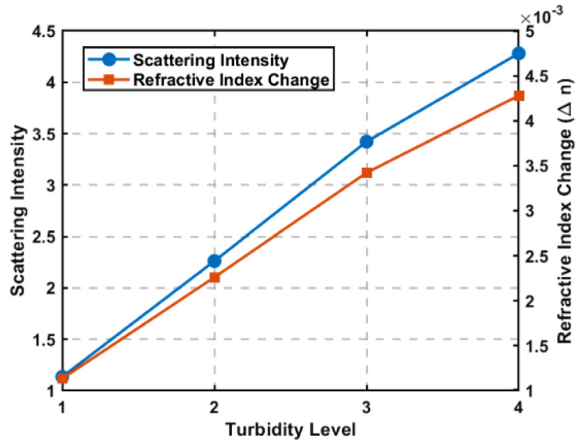

Fig. 5. Refractive-index variation and scattering intensity at different turbidity levels.

As shown in Fig. 4(b), the dataset contains five target objects with different geometries. For each turbidity level, samples were split into training and testing subsets at a ratio of 4:1, yielding 7200 images for training and 1800 images for testing. The training set was used to learn network parameters, whereas the test set was used for final performance evaluation. All experimental results reported in this work are strictly based on the test set. To increase data diversity, all objects were captured under randomly varied spatial poses. To highlight the influence of the acquisition environment on object-signal modulation, background interference was removed when visualizing fringe images, phase maps, and 3D reconstruction results, and only the effective regions corresponding to the target objects were retained.

A graded turbid-water environment was constructed by injecting milk into the tank to emulate variations in water turbidity. To ensure consistency, it should be noted that during image acquisition, the water surface was kept still. The target object was first fixed at the bottom of the tank, and ground-truth data were acquired by directly projecting fringe patterns. Based on 13 L of clear water, 4, 5, 6, and 7 mL of milk were sequentially added following a gradient-injection protocol, yielding four increasing turbidity levels (T1–T4) with gradually elevated refractive index and scattering intensity. To facilitate practical interpretation, the four turbidity levels qualitatively correspond to progressively increasing real-world underwater scattering conditions. Specifically, T1 represents relatively clear waters with a low suspended particle concentration, comparable to favorable coastal or offshore environments; T2 corresponds to mildly turbid nearshore waters with moderate particulate content; T3 simulates highly turbid estuarine or sediment-rich coastal regions with pronounced scattering and attenuation; and T4 represents extremely turbid conditions with strong backscattering effects, similar to harbor waters or inland waters containing dense suspended particles. Fringe images were synchronously captured under each turbidity condition. The scattering intensity and refractive-index variation associated with the four turbidity levels are shown in Fig. 5.

Notably, the performance of the trained FP-DiffNet was evaluated across these four turbidity levels. At the highest tested turbidity (T4), the network still restores fringe patterns effectively. However, if turbidity increases further, restoration is expected to fail because scattering becomes dominant, the direct transmission component approaches zero, and the fringe contrast drops below the noise floor. Specifically, according to the J–M model, when the attenuation coefficient increases to a level at which the direct-transmission intensity $I_d(z)$ falls below the noise baseline, fringe information is completely lost and recovery becomes infeasible. In the present setup, T4 represents an extreme condition. The network maintains sub-millimeter reconstruction accuracy at T4, indicating strong robustness, whereas performance is expected to degrade progressively when turbidity exceeds T4.

Figure 6(a) shows the fringe pattern acquired in air, and Fig. 6(b) presents the fringe images captured under the four increasing turbidity levels in the experiments. For the phase-shifting fringe images acquired in the above turbid-water conditions, a diffusion-model-based optimization is applied in this study.

B. Training Details

During model training, the key parameters were systematically configured and optimized. The main hyperparameter of the diffusion process, the number of diffusion time steps, was empirically set to $T = 1000$ to balance restoration quality and computational cost. A linear variance schedule was adopted, where β_t increases from 0.0001 to 0.03 ($\beta_t \in [0.0001, 0.03]$) for stable and smooth noise injection and removal in forward and reverse processes. Consistency of the noise variance was strictly maintained between the two processes. The model was optimized using AdamW with an initial learning rate of 1.0×10^{-5} to ensure stable convergence. A cosine annealing learning-rate schedule was applied with $T_{max} = 2500$ epochs for gradual solution refinement, and training was conducted for a total of 2500 epochs. Due to computational constraints, the batch size was set to 16.

The inference speed of FP-DiffNet is quantified on the experimental computing platform equipped with an NVIDIA GeForce RTX 4090 GPU to assess its practical feasibility for near-real-time underwater inspection. For the core task of denoising and refraction correction on 512×512 pixel fringe images, the average single-image inference time of FP-DiffNet is approximately 430 ms, a performance demonstrating promising potential for applications in near-real-time underwater 3D imaging scenarios.

C. Performance Validation and Evaluation

To evaluate the performance of the trained FP-DiffNet model under different turbidity conditions, assessments were conducted from three aspects: 2D image analysis, phase-accuracy evaluation, and 3D reconstruction performance. A progressive validation strategy was adopted. Experiments were first carried out on simple objects and were then extended to two new scenarios: a composite scene containing three geometric primitives and a complex-object scene.

1. 2D Image Analysis

Figures 7(a) and 7(b) present qualitative comparisons before and after processing under four increasing turbidity levels. To evaluate the quality of the generated fringe patterns, the absolute error distributions between the ground-truth images and the network-predicted fringes were computed for the four turbidity conditions, as shown in Figs. 8(a)–8(d). Visual inspection indicates no noticeable differences between the two sets of images, and the maximum absolute error for the highest-turbidity sample is only 0.225 [Fig. 8(d)]. To quantify the discrepancy between the ground truth and the network output and to objectively assess fringe restoration quality, PSNR (peak signal-to-noise ratio), SSIM (structural similarity index), and MS-SSIM (multi-scale structural similarity index) were adopted as evaluation metrics, which quantify fringe restoration performance in terms of pixel fidelity and structural preservation. As reported in Table 2, under extremely high turbidity, the

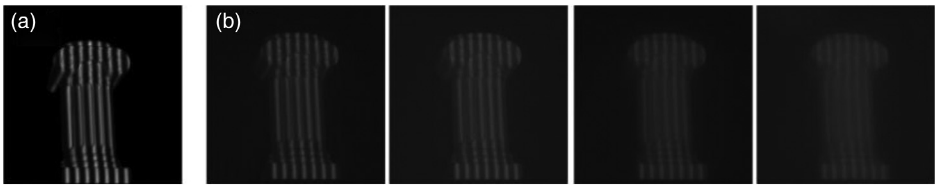

Fig. 6. Input images to the model: (a) fringe pattern acquired in air and (b) fringe patterns captured under four turbidity levels.

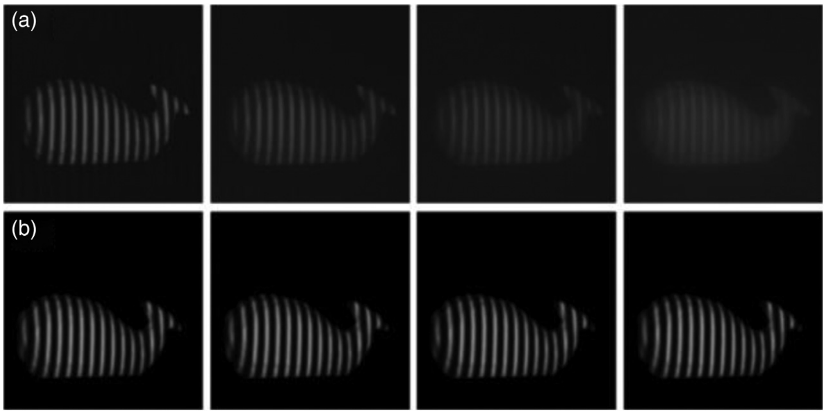

Fig. 7. Fringe images under different turbidity levels: (a) before processing and (b) after processing.

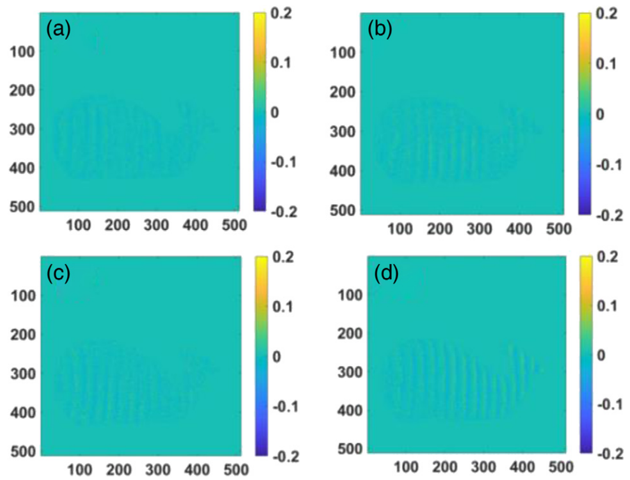

Fig. 8. (a)–(d) Absolute error between the ground-truth and restored fringe images for each turbidity level.

Table 2. Errors between the Network Predictions and the Ground-Truth Images at Each Turbidity Level

	T1	T2	T3	T4
PSNR (dB)	43.0212	42.8786	42.5056	42.4192
SSIM	0.9849	0.9825	0.9752	0.9750
MS-SSIM	0.9982	0.9979	0.9975	0.9974

model achieves the best performance on the simple-object test set, with a PSNR of 42.42 dB and SSIM (0.97) and MS-SSIM (0.99) values close to their theoretical upper bounds, demonstrating the expected fringe restoration capability in turbid environments.

Table 3. Errors between the Wrapped Phases Computed from the Network Predictions and the Ground-Truth Wrapped Phases at Each Turbidity Level

	T1	T2	T3	T4
MAE (rad)	0.0260	0.0290	0.0301	0.0354
Cosine similarity	0.9962	0.9959	0.9957	0.9951

2. *Phase-Accuracy Evaluation*

For all conditions, phase computation was performed using the standard four-step phase-shifting algorithm. Phase accuracy was assessed by comparing the model-inferred images with the reference images acquired in the air. Using a representative multi-object pose configuration from the test set as an example, Figs. 9(a)–9(d) show the wrapped-phase error distributions under the four turbidity levels. Table 3 summarizes the wrapped-phase mean absolute error (MAE) and cosine similarity for each turbidity condition, which jointly characterize phase accuracy from both error magnitude and overall phase-map consistency. The results indicate that, after network-based optimization, phase errors are mainly concentrated around object boundaries, shadowed regions, and phase-discontinuity locations. Overall, phase accuracy is substantially improved.

For quantitative analysis, fringe images acquired under high turbidity were selected as representative samples. Figures 10(a) and 10(c) show the ground-truth wrapped-phase maps, whereas Figs. 10(b) and 10(d) show the wrapped-phase maps computed from the fringe patterns restored by the network. Line profiles were extracted along the reference rows $y = 310$ and $y = 230$ pixels, and the corresponding phase curves are plotted in Figs. 10(e) and 10(g), respectively. The zoomed-in views in

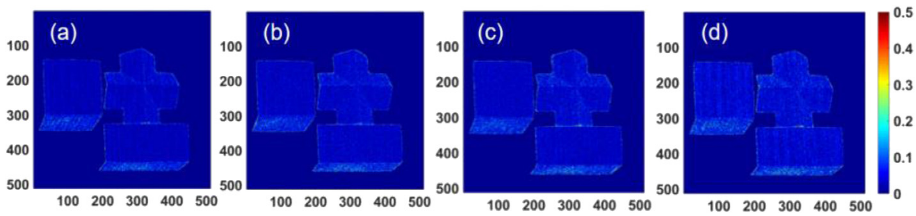

Fig. 9. Wrapped-phase error maps under different turbidity levels.

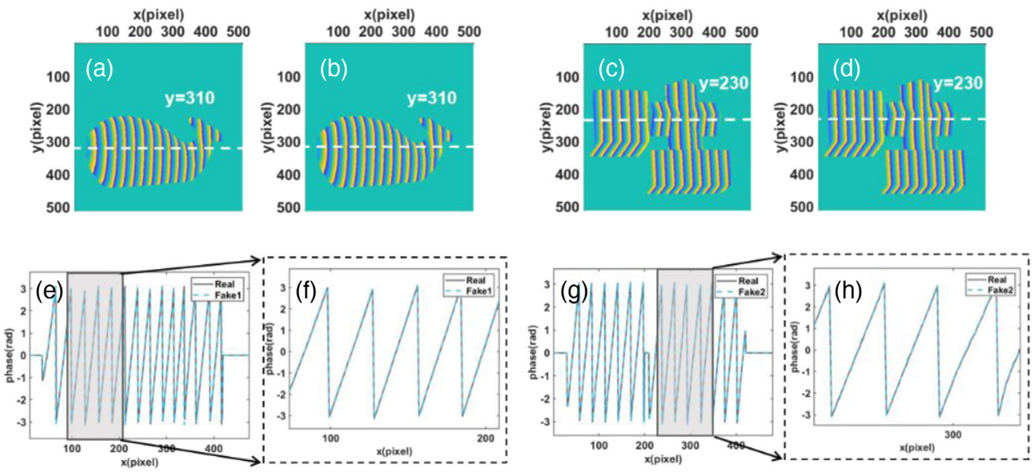

Fig. 10. Comparison of wrapped phases from the ground truth and those computed from the network predictions: (a), (c) ground-truth wrapped-phase maps; (b), (d) wrapped-phase maps computed from the restored fringes; (e) line profiles at $y = 310$ for (a), (b); (g) line profiles at $y = 230$ for (c), (d); (f) and (h) zoomed-in views of (e) and (g), respectively.

Figs. 10(f) and 10(h) further demonstrate that the two-phase curves are highly consistent.

3. 3D Reconstruction Performance

In the test cases, 3D geometric reconstruction was performed on complex workpieces to comprehensively evaluate the feasibility and performance of the proposed method. Reconstruction accuracy was quantitatively assessed by computing the root mean square error (RMSE) between the reference reconstruction and the model-optimized reconstruction under each turbidity level. The RMSE values are reported in Table 4.

The 3D reconstruction comparisons are shown in Fig. 11. Specifically, Figs. 11(a)–11(d) present the original reconstruction results under different turbidity levels, and Figs. 11(e)–11(h) show the reconstructions obtained after model-based optimization. Visual inspection indicates that turbid water causes pronounced degradation of 3D geometric features, whereas the optimized results exhibit geometric fidelity and detail recovery that are close to the reference reconstruction. These results demonstrate that (1) the proposed model effectively mitigates fringe-information loss induced by refraction and scattering; and (2) sub-millimeter reconstruction accuracy is maintained even under extremely turbid conditions (RMSE < 0.1 mm), with markedly improved reconstruction completeness and feature preservation compared with conventional methods.

Table 4. Errors between the Reconstructed Results and the Reference Reconstruction at Each Turbidity Level

	T1	T2	T3	T4
RMSE (mm)	0.0317	0.0457	0.0593	0.0890

4. CONCLUSION

This work investigates the multi-physics coupled disturbances affecting structured-light 3D imaging in underwater optical environments, including optical-path deviation, scattering attenuation, and refraction-induced distortion. To address these

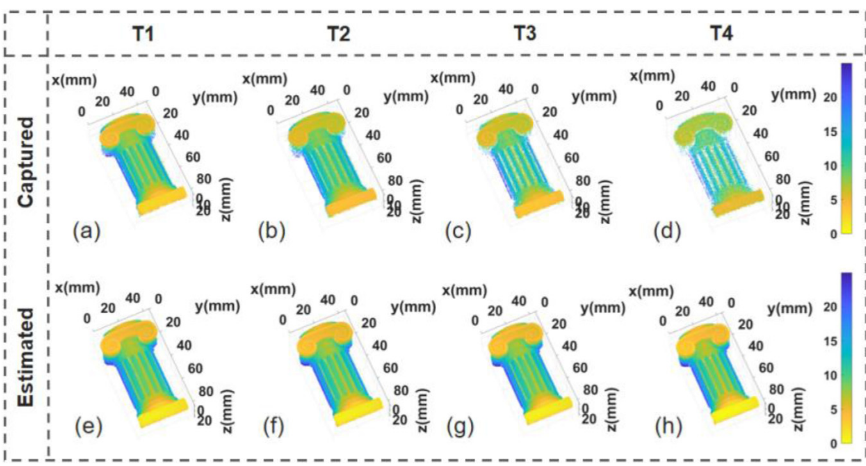

Fig. 11. Comparison of the original reconstructions and the model-optimized reconstructions at each turbidity level.

challenges, we propose FP-DiffNet, a deep diffusion-model-based fringe restoration framework that formulates underwater fringe degradation as a constrained reverse diffusion process. By integrating an adaptive noise-annealing mechanism with physics-guided constraints, FP-DiffNet incorporates priors of underwater light transport physics through a simplified degradation model and periodic gradient correction. This design enables effective decoupling of scattering-induced noise from geometric fringe structures and mitigates phase distortion caused by non-uniform illumination, overcoming key limitations of conventional underwater phase reconstruction methods.

Extensive experiments demonstrate the effectiveness, robustness, and practical value of the proposed framework. Under extremely high turbidity conditions, FP-DiffNet achieves a PSNR of 42.42 dB, with SSIM and MS-SSIM of 0.9750 and 0.9974, respectively, and reduces the wrapped-phase MAE to 0.0354 rad. Combined with dynamic compensation, the diffusion-driven reconstruction enables sub-millimeter absolute measurement in turbid water, with the 3D reconstruction RMSE consistently below 0.1 mm across graded turbidity levels. Notably, this performance is achieved without prior modeling of water optical parameters, significantly enhancing robustness and generalization. It is worth noting that our FP-DiffNet serves as a modular fringe-restoration front end compatible with the standard N-step phase-shifting structured-light framework, producing outputs consistent with the conventional PSP so that phase retrieval and 3D reconstruction can proceed unchanged. While the framework is generally applicable to sinusoidal phase-shifting patterns, adaptation to other structured-light configurations or underwater imaging conditions may require fine-tuning with configuration-specific data. These results highlight FP-DiffNet as a practical and reliable solution for underwater 3D imaging tasks such as deep-sea equipment inspection, underwater cultural heritage digitization, and marine structural monitoring.

Funding. National Natural Science Foundation of China (62375078, 62073123, 12572209); Science and Technology Innovation Project of Chinese Academy of Traditional Chinese Medicine (ZN2024A02); National Key Laboratory of Ship Structural Safety (NAKLAS2024KF016-S); Key Research Project Plan for Higher Education Institutions in Henan Province (24ZX011); Training Plan for Young Backbone Teachers in Undergraduate Universities in Henan Province (2023GGJS058); Cultivation Programme for Young Backbone Teachers in Henan University of Technology; Open Fund of Institute for Complexity Science (CSKFJJ-2024-3).

Disclosures. The authors declare no conflict of interest.

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

REFERENCES

1. M. Liu, J. Wei, Z. Qiao, *et al.*, "Improved shape from focus network for extended depth of field and rapid three-dimensional reconstruction biological microscopy," Opt. Laser Technol. **191**, 113345 (2025).
2. J. Shi, F. Xing, Y. Liu, *et al.*, "Three-dimensional reconstruction for the whole lung with early multiple pulmonary nodules," J. Vis. Exp. **200**, 6578637902364 (2023).
3. W. Yanzhe, P. A. N. Zemin, Y. Qian, *et al.*, "Robot trajectory planning for autonomous 3D reconstruction of cockpit in aircraft final assembly testing," Chin. J. Aeronaut. **37**, 332–350 (2024).
4. L. Lei, J. Zhaoyi, W. Ke, *et al.*, "3D reconstruction of multi-target moving objects based on phase-shifting method," Infrared Laser Eng. **49**, 0303011 (2020).
5. W. Yin, J. Zhong, S. Feng, *et al.*, "Composite deep learning framework for absolute 3D shape measurement based on single fringe phase retrieval and speckle correlation," J. Phys. Photonics **2**, 045009 (2020).
6. H. Nguyen, J. Liang, Y. Wang, *et al.*, "Accuracy assessment of fringe projection profilometry and digital image correlation techniques for three-dimensional shape measurements," J. Phys. Photonics **3**, 014004 (2021).
7. H. Cheng, S. Yu, H. Yu, *et al.*, "Bioinspired underwater navigation using polarization patterns within Snell's window," China Ocean Eng. **37**, 628–636 (2023).
8. N. Lyu, H. Yu, J. Han, *et al.*, "Structured light-based underwater 3-D reconstruction techniques: a comparative study," Opt. Lasers Eng. **161**, 107344 (2023).
9. Y. Z. Liu, K. C. Chi, and S. Lin, "Underwater image restoration based on background light estimation and transmittance optimization," Laser Optoelectron. Prog. **57**, 141010 (2020).
10. T. Treibitz and Y. Y. Schechner, "Turbid scene enhancement using multi-directional illumination fusion," IEEE Trans. Image Process. **21**, 4662–4667 (2012).
11. Y. Lee, K. B. Gibson, Z. Lee, *et al.*, "Stereo image defogging," in *IEEE International Conference on Image Processing (ICIP)* (IEEE, 2014), pp. 5427–5431.
12. X. Fu, Y. Liao, D. Zeng, *et al.*, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," IEEE Trans. Image Process. **24**, 4965–4977 (2015).
13. J. Zhou, D. Zhang, P. Zou, *et al.*, "Retinex-based Laplacian pyramid method for image defogging," IEEE Access **7**, 122459–122472 (2019).
14. J. Deng, G. Luo, and C. Zhao, "UCT-GAN: underwater image colour transfer generative adversarial network," IET Image Proc. **14**, 3613–3622 (2020).
15. L. Hong, X. Shu, Q. Wang, *et al.*, "CCM-Net: color compensation and coordinate attention guided underwater image enhancement with multi-scale feature aggregation," Opt. Lasers Eng. **184**, 108590 (2025).
16. W. Hu, S. Chen, T. Luo, *et al.*, "DATDM: dynamic attention transformer diffusion model for underwater image enhancement," Alexandria Eng. J. **126**, 591–604 (2025).
17. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).
18. M. Jie, Q. Jiarui, C. Dingfang, *et al.*, "SLAM-based underwater image enhanced visual 3D reconstruction method," China Mech. Eng. **35**, 268 (2024).
19. X. Peng, Z. Liang, J. Zhang, *et al.*, "Review of underwater image preprocessing based on deep learning," Comput. Eng. Appl. **57**, 43–54 (2021).
20. J. Li, K. A. Skinner, R. M. Eustice, *et al.*, "WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images," IEEE Rob. Autom. Lett. **3**, 387–394 (2017).
21. M. Mousavi, R. Estrada, and A. Ashok, "iDehaze: Supervised underwater image enhancement and dehazing via physically accurate photorealistic simulations," Electronics **12**, 2352 (2023).
22. S. Van Der Jeught, P. G. G. Muyshondt, and I. Lobato, "Optimized loss function in deep learning profilometry for improved prediction performance," J. Phys. Photonics **3**, 024014 (2021).
23. X. U. Xinya, L. Ü. Nenqing, Y. U. Haotian, *et al.*, "Underwater 3-D measurement using fringe projection profilometry based on deep learning," Aeronaut. Manuf. Technol. **66**, 86–95 (2023).
24. F. Jianxin, H. Yajun, P. Chengsheng, *et al.*, "AL-GAN: a lightweight GAN underwater image enhancement incorporating attention mechanisms," J. Ordnance Equipment Eng. **45**, 135–143 (2024).
25. L. Lu, C. Bu, Z. Su, *et al.*, "Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging," Adv. Photonics **6**, 046004 (2024).
26. Z. Zhu, X. Li, Q. Ma, *et al.*, "FDNet: Fourier transform guided dual-channel underwater image enhancement diffusion network," Sci. China Technol. Sci. **68**, 1100403 (2025).

27. Y. Li, Z. Mi, P. Lin, *et al.*, "Underwater optical image enhancement via an unsupervised contrastive learning-guided diffusion model," Opt. Lasers Eng. **194**, 109144 (2025).
28. W. Wu and M. Yao, "Optimization of underwater images based on gray world algorithm and Jaffe-McGlamery models," Front. Comput. Intell. Syst. **11**, 80–84 (2025).
29. Y. E. Zevatskii, S. S. Lysova, T. A. Skripnikova, *et al.*, "Bouguer–Lambert–Beer law of absorption: spectrophotometry in electrolyte solutions," Russ. J. Phys. Chem. A **98**, 242–247 (2024).
30. L. Gurdev, T. Dreischuh, I. Bliznakova, *et al.*, "Scattering of a laser beam in turbid media with forward-peaked Henyey-Greenstein indicatrices," Phys. Scr. **T149**, 014074 (2012).
31. X. Li and W. Zhang, "Application of Snell's law in reflection raytracing using the multistage fast marching method," Earthq. Res Adv. **1**, 100009 (2021).
32. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Adv. Neural Inf. Process. Syst. **33**, 6840–6851 (2020).
33. L. Lu, Y. Wang, Y. Guo, *et al.*, "Deep diffusion model-based dual structured light for large-format complex surface imaging," Meas. Sci. Technol. **37**, 065009 (2026).
34. H. Zhang, R. He, and W. Fang, "An underwater image enhancement method based on diffusion model using dual-layer attention mechanism," Water **16**, 1813 (2024).
35. J. Shao, H. Zhang, and J. Miao, "Region gradient-guided diffusion model for underwater image enhancement," Mach. Vis. Appl. **36**, 38 (2025).
36. Y. Yao, H. Wang, H. Sun, *et al.*, "UI-MSDE-diffusion: fringe pattern noise elimination based on diffusion model," Optics Laser Technol. **184**, 112435 (2025).
37. Z. Su, Y. Xu, X. Qin, *et al.*, "Automatic crack tracking with consistent stereo imaging and wavelet analysis," Int. J. Mech. Sci. **304**, 110662 (2025).