ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

Research on Underwater Structured Light Three-Dimensional Reconstruction Based on Deep Learning

Yuheng Wang[1, a], Haojun Zhang[1, b*], Lei Lu[2, c*], Zhilong Su[3, d], and Wei Pan[4, e]

[1] College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China;

[2] Institute for Complexity Science, 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 R&D, OPT Machine Vision Tech Co., Ltd, Dongguan 523860, China.

[a] 16638247200@163.com, [b] zhj@haut.edu.cn, [c] ielulei@163.com, [d] szloong@shu.edu.cn,

[e] vpan@foxmail.com

Abstract. Underwater structured light three-dimensional (3D) imaging is essential for applications in underwater exploration and inspection. However, it suffers from stripe brightness attenuation, contrast reduction, and random noise introduction due to absorption, scattering, and refraction across media interfaces, leading to degraded stripe decoding and reconstruction accuracy. This paper presents a diffusion model-based approach for underwater structured light 3D reconstruction. Without altering the core framework of structured light imaging, the method employs the diffusion model to learn the degradation process of underwater stripes and effectively restore them, thereby enhancing 3D reconstruction quality. The approach integrates an encoder-decoder denoising network in the reverse diffusion stage, taking noisy images, time-step embeddings, and turbidity conditions as input to progressively predict and eliminate noise, ensuring incremental detail recovery and stability. Time-step embeddings are incorporated to adaptively adjust the denoising strategy at varying degradation stages, maintaining stripe structure consistency. Experiments conducted on an underwater structured light platform, with turbidity levels controlled by milk injection, demonstrate the method's performance using multi-frequency phase-shifted fringes and geometric targets. Results show that, under the highest turbidity, the PSNR reaches 42.42 dB, SSIM is 0.97, and the 3D reconstruction RMSE remains below 0.1 mm, validating the method’s robustness and high-precision reconstruction capability in complex aquatic environments.

Keywords: Underwater structured light; 3D reconstruction; diffusion model; deep learning.

1. Introduction

With the development of marine resource exploitation, underwater engineering, and underwater robotics, the demand for high-precision 3D reconstruction of underwater targets has been steadily increasing. Structured light 3D imaging has gained significant attention for its advantages, including active illumination, high precision, and low dependence on surface texture, making it widely used in close-range, high-precision measurements [1,2]. However, existing structured light theories and systems are primarily based on air medium environments. When applied underwater, the presence of multi-medium interfaces, such as refraction between water, glass, and air, as well as scattering and absorption by the water, causes stripe brightness attenuation, contrast reduction, and noise introduction, significantly degrading stripe decoding and 3D reconstruction accuracy [3,4].

To address these challenges in underwater structured light imaging, previous studies have primarily adopted cross-medium refraction models for calibration and compensation. While these methods can enhance reconstruction accuracy, they typically require high-precision calibration and are sensitive to environmental variations, limiting their robustness in practical applications. Traditional approaches for mitigating underwater image degradation can be divided into restoration methods relying on prior modeling (e.g., polarization imaging [5], stereo vision [6]) and

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91
enhancement methods focusing on post-processing (e.g., histogram equalization [7], color correction [8]). However, the former is constrained by the difficulty of obtaining water optical parameters, while the latter often introduces color distortion and noise, making them ineffective in addressing imaging degradation in complex aquatic environments.

In recent years, deep learning has emerged as a powerful tool for underwater vision tasks due to its strong feature learning capabilities, providing new opportunities for underwater image enhancement and object detection [9]. For underwater 3D reconstruction, existing learning frameworks are mainly divided into two categories: model-driven methods that integrate physical models, and data-driven end-to-end methods [10]. For example, Li et al. [11] used WaterGAN combined with a light transport model for unsupervised underwater image restoration and depth estimation, while Mousavi et al. [12] proposed iDehaze, which generates data through physical simulation and employs a two-step framework for dehazing and color correction. However, the explicit constraints imposed by physical models may limit their adaptability in real-world, complex scenarios. In contrast, data-driven end-to-end methods exhibit stronger generalization potential. For instance, Xu et al. [13] designed FPENet to directly optimize low-quality fringe images for improved phase measurement accuracy, and Feng et al. [14] used a lightweight AL-GAN to effectively restore image details. However, such methods often lack physical consistency constraints, which may result in distorted fringe patterns, thus introducing errors in high-precision phase reconstruction. Overall, effectively modeling fringe degradation in complex aquatic environments, while balancing stripe restoration quality and reconstruction stability, remains a topic for further exploration.

In response to these challenges, this paper proposes a novel method for underwater structured light 3D reconstruction based on a diffusion model. The proposed method utilizes a diffusion model to perform data-driven modeling of the underwater stripe degradation process, enabling the progressive restoration of degraded stripes. Additionally, we introduce a denoising network with time-step embeddings and turbidity conditions to enhance the stability and consistency of stripe recovery at varying levels of degradation. Experimental validation is carried out under multiple turbidity levels, multi-frequency phase-shifted fringes, and different geometric target conditions, with quantitative results provided for 2D recovery, phase accuracy, and 3D reconstruction.

2. Proposed Method

2.1 Principle of Underwater Structured-Light Imaging

In underwater environments, the structured light imaging process is affected by multiple factors, including absorption, scattering, and cross-medium refraction. According to the classical Jaffe-McGlamery theory [16], the structured light stripes during propagation can be decomposed into three components: direct transmission light, forward scattering light, and backward scattering light. The underwater physical propagation of these components is illustrated in Figure 1.

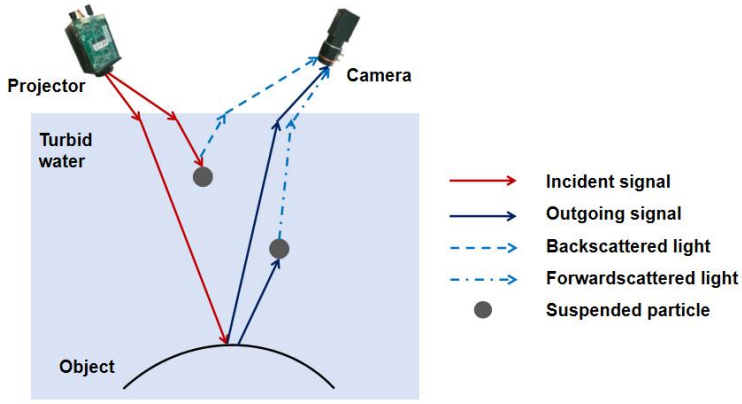


Fig. 1 Schematic illustration of light propagation in turbid underwater environments

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

Direct transmission light travels through the medium without undergoing scattering, maintaining its original characteristics. However, its intensity undergoes attenuation as it propagates through water, which can be described by the modified Beer-Lambert law [16]. This attenuation is primarily due to light absorption by the water, and the corresponding mathematical expression is given in Equation (1). Forward scattering light is caused by small-angle scattering from suspended particles, bubbles, and other impurities in the water. These particles scatter the light in the forward direction, which causes the edges of the stripes to blur and reduces the contrast of the overall pattern. The spatial degradation characteristics of this scattered light are typically modeled by the point spread function, as shown in Equation (2). Finally, backward scattering light occurs when light is scattered in the reverse direction, often by larger particles or air bubbles. This component primarily manifests as background noise and contributes to the degradation of the image quality. The statistical properties of this noise can be described by a random noise model, and the corresponding mathematical expression is provided in Equation (3).

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

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

$$I_b \sim \mathcal{P}\left(kI_0 e^{-b_b z}\right) + \mathcal{N}\left(0, \sigma_{read}^2\right). \tag{3}$$

Additionally, the scattering directionality of suspended particles in water can be characterized by the Henyey–Greenstein phase function [17], with its mathematical form shown in Equation (4). This effect further weakens the high-frequency components of the structured light stripes, causing the image to exhibit a noticeable low-contrast characteristic overall. In addition to scattering and absorption effects, underwater structured light imaging is also significantly influenced by cross-medium refraction. The propagation relationship of light at the water-air interface can be described by Snell's Law [18], with the corresponding mathematical expression shown in Equation (5). The refraction effect causes a shift in the light path, introducing depth-related system errors during the 3D reconstruction process.

$$p(\theta) = \frac{1}{4\pi} \frac{1-g^2}{(1+g^2-2g\cos\theta)^{3/2}}. \tag{4}$$

$$n_w \sin\theta_i = n_a \sin\theta_t. \tag{5}$$

Considering all these factors, the observation process of underwater structured light stripes can be expressed as the result of coupling ideal stripes with various degradation factors. The overall imaging model can be summarized by Equation (6).

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

2.2 Diffusion Model

To address the complex and difficult-to-explicitly model degradation effects in underwater structured light imaging, this paper introduces a diffusion model to model the generation and restoration processes of structured light stripes. The diffusion model works by progressively adding noise to the data and learning the corresponding reverse denoising process, effectively capturing the complex distribution of the data [19,20]. The overall process is illustrated in Figure 2, which clearly depicts the complete diffusion process from data degradation to high-quality stripe restoration. This provides an intuitive reference for understanding the subsequent forward and reverse diffusion stages. In contrast to traditional discriminative network structures, this generative model has the

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91
inherent advantage of restoring image details in a gradual manner, making it especially effective in handling situations with strong noise and low-contrast stripes.

Furthermore, the diffusion model offers significant advantages in preserving fine details during the restoration process. Unlike traditional methods that may fail to recover high-frequency components or risk introducing artifacts, the diffusion model enables a smooth and progressive recovery of information by iteratively refining the image. This makes the model particularly suited for applications in challenging environments, such as underwater imaging, where noise levels are high and image contrast is often low. The adaptability of the model to varying levels of degradation also allows it to generalize well to a wide range of imaging conditions, ensuring robust and reliable performance in real-world underwater scenarios.

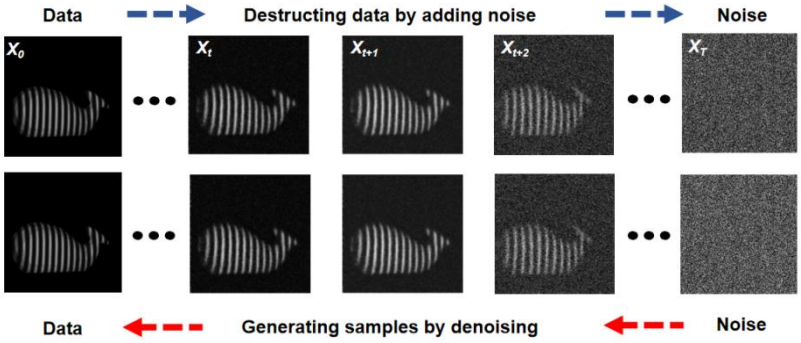


Fig. 2 Diffusion model process

In the forward diffusion stage, ideal structured light stripes X_0 progressively accumulate random noise over T time steps, with the evolution process described by a Markov chain. At each step, the image X_{t-1} is perturbed toward the noise space according to predefined noise scheduling parameters, as shown in Equation (7). By introducing noise perturbations of varying intensities across multiple time steps, this process effectively simulates the degradation effects caused by scattering, absorption, and random environmental disturbances in the underwater imaging process, allowing the model to fully capture the distribution characteristics from clear stripes X_0 to pure noise X_t during the training phase. This enables the model to learn how real-world disturbances affect the imaging process, laying the foundation for stripe recovery in the subsequent reverse diffusion stage.

$$q(X_t|X_{t-1}) = N\left(X_t; \sqrt{1-\beta_t}\,X_{t-1}, \beta_t I\right). \tag{7}$$

In the reverse diffusion stage, the model takes the noisy image X_t and time step t as input, progressively predicting and removing the noise components ε introduced at each step, with the conditional probability expression shown in Equation (8). By learning the parameters of the conditional distribution, the model gradually recovers the original structured light stripes X_0, mitigating the noise influence. To achieve this reverse denoising process, this paper designs a denoising network based on an encoder-decoder structure, with the detailed process and network structure illustrated in Figure 3.

The network takes the noisy image X_t, time embeddings t, and condition information y obtained from the forward process as input and learns to predict the noise, gradually recovering the denoised image X_{t-1}. The main structure of the network employs a multi-layer convolutional architecture, progressively down-sampling to extract multi-scale features of the stripes, and then gradually restoring the spatial resolution during the decoding phase, balancing global structural information with local stripe details.

$$p_\theta(X_{t-1}|X_t) = N(X_{t-1}; \mu_\theta(X_t, t), \Sigma_t). \tag{8}$$

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

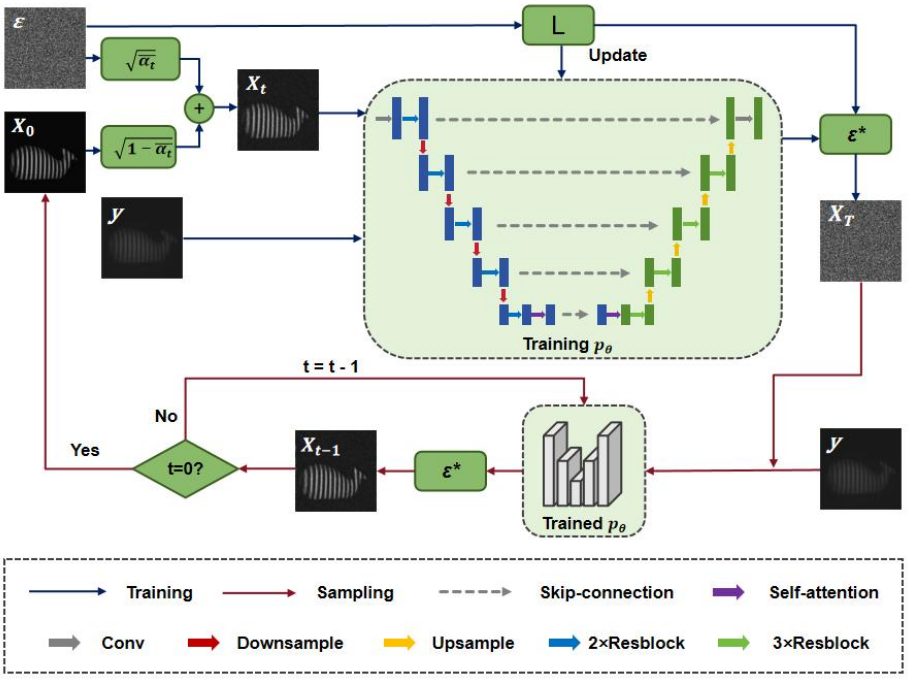


Fig. 3 Detailed workflow of the model

In the encoding stage of the network, the input fringe image first passes through multiple convolutional layers to extract shallow features, which describe the basic spatial distribution of the stripes. Then, through successive down-sampling operations, the receptive field is expanded, allowing the network to capture global degradation features caused by underwater scattering and blurring. In the decoding stage, the network adopts a symmetric structure to progressively restore the resolution of the feature maps and uses skip connections to merge the feature information from the encoding stage, preventing the loss of high-frequency stripe details in deeper layers of the network.

Additionally, to enhance the model's ability to perceive diffusion time step information, the network introduces a time-step embedding mechanism in the middle layers. This mechanism maps time information from the diffusion process to a high-dimensional feature space, which is then fused with the image features. This design enables the network to adaptively adjust its denoising strategy according to different degradation stages, thereby improving the stability and consistency of stripe recovery.

Through the above structural design, the diffusion model is able to learn the nonlinear mapping from underwater degraded stripes to ideal stripes. During training, the goal is to maximize the fidelity of the mapping. The loss function (Equation 9) consists of a structural similarity constraint, noise prediction error, and sparsity regularization term, which jointly optimize the structural reconstruction quality, noise modeling capability, and model complexity. Here, the hyperparameters are configured based on the characteristics of underwater imaging: λ_1 is set to 0.7 to preserve the integrity and distinguishability of the stripe structure; λ_2 is set to 0.8 to introduce moderate constraints in noise modeling, preventing excessive noise amplification under high turbidity conditions; and λ_3 is set to 0.02 to suppress overfitting through L1 regularization, which enhances the robustness and generalization performance of the model under conditions of low signal-to-noise ratio. As a result, the model effectively decouples the medium degradation effects, enabling high-quality stripe restoration.

$$L = \lambda_1 \cdot \mathrm{SSIM}(X, \widehat{X}) + \lambda_2 \cdot \mathrm{E}[|\varepsilon_t - \varepsilon_\theta(X_t, t, y)|] + \lambda_3 \cdot \|\theta\|_1. \tag{9}$$

In the inference stage, the input underwater structured light image is progressively denoised and restored by the diffusion model, significantly enhancing stripe clarity and contrast. The restoration

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91
result is essentially a sequence of surface stripe patterns, mathematically consistent with the fringe patterns in standard N-step phase shifting profilometry. To further obtain the surface phase information and complete the 3D reconstruction, the fringe patterns from each frame of the sequence are substituted into the phase calculation and unwrapping formulas of standard N-step phase shifting profilometry, yielding a continuous phase distribution. This provides reliable input for subsequent 3D reconstruction.

3. Experiments

To validate the effectiveness of the proposed method, an underwater structured light 3D imaging platform was constructed, with its system structure shown in Figure 4. The experimental platform consists of a high-resolution industrial camera, a digital projector, and an acrylic water tank, which can stably capture structured light image data under different turbidity conditions. The experiment uses a three-frequency four-step phase-shifting method to generate fringes with frequencies $f_1 = 1/59, f_2 = 1/64, f_3 = 1/70$. The effective imaging area of the captured images is 512×512 pixels, with the target object placed inside a 0.4 m $\times$ 0.4 m $\times$ 0.3 m water tank. The working distance of the imaging system is approximately 0.6 m.

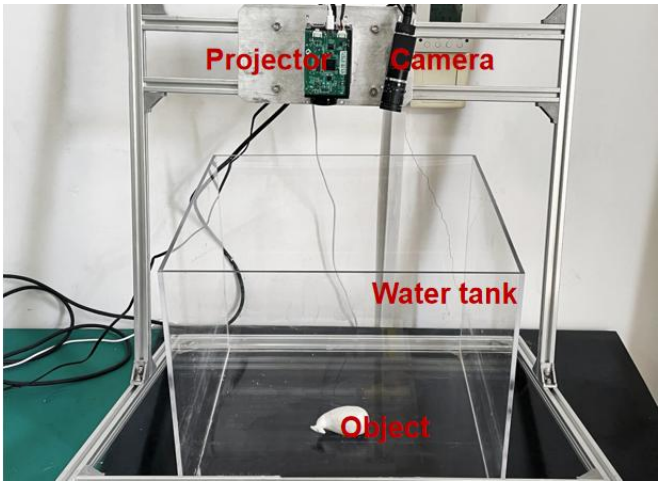


Fig. 4 Imaging experimental setup

The experimental objects include targets with various geometric shapes, used to comprehensively evaluate the adaptability of the proposed method in complex scenarios. Milk is gradually injected into the water tank to simulate variations in water turbidity, creating four increasing turbidity levels (T1–T4). Structured light fringe images are captured under each turbidity condition, with corresponding ground truth fringe images (i.e., fringe images captured in the air environment) simultaneously obtained as reference baselines to support quantitative evaluation and comparison analysis.

For data organization, the collected samples are divided into training and testing sets in a 4:1 ratio. For each turbidity level, 7,200 images are used for training, and 1,800 images are used for testing. During the experiment, the turbidity level is gradually adjusted, allowing the system to obtain fringe data at different degradation intensities, thereby verifying the stability and 3D reconstruction performance of the proposed method under various interference levels.

To comprehensively evaluate the performance of the proposed model under different turbidity conditions, this study conducts systematic validation from three aspects: 2D fringe recovery, phase calculation accuracy, and 3D reconstruction quality.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

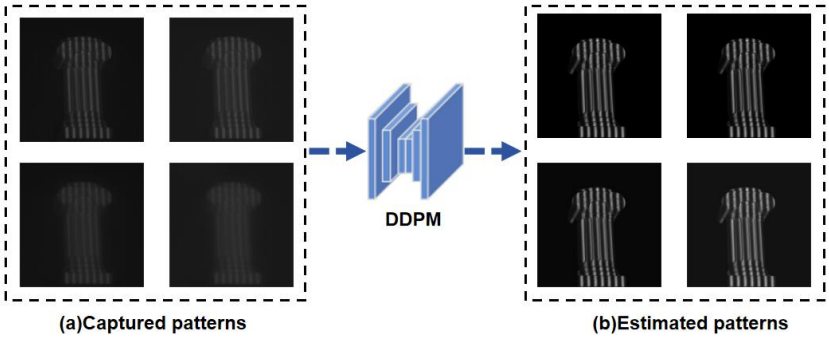


Fig.5 Model processing results(a)Captured fringe image(b)Estimated fringe image

In the 2D image analysis, the processing results under four increasing turbidity levels (T1-T4) are shown in Figure 5(a) and (b). By calculating the absolute error between the recovered fringes and the reference image in the air, the maximum error at the highest turbidity is only 0.225. The quantitative metrics show that, on the test set with extremely high turbidity, the PSNR reaches 42.42 dB, and the SSIM and MS-SSIM values are 0.97 and 0.99, respectively (Table 1), confirming the model's excellent fringe recovery capability.

Table 1. 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

Phase accuracy is evaluated by comparing the wrapped phase results obtained from the recovered fringe images and the reference fringe images in the air environment. As shown in Figure 6, (a) and (c) represent the ground truth wrapped phase results obtained from the reference fringe images in the air environment; (b) and (d) correspond to the wrapped phase results computed from the fringe images predicted by the network under two different turbidity levels. To further visually compare the phase consistency, (e) and (f) show the phase curves obtained by horizontally sampling the phase maps at y=210 pixels and y=130 pixels, respectively, with local zoom-ins. It can be seen that the network-predicted phase curves are highly consistent with the ground truth phase curves in both overall trend and local details, with the maximum phase error not exceeding 0.04 radians. The cosine similarity remains at a high level of approximately 0.995, indicating that the proposed method effectively ensures the accuracy and stability of phase computation under different turbidity conditions.

The 3D reconstruction performance is quantified by the root mean square error (RMSE) between the reconstruction results and the reference model. Quantitative results show that, under extreme turbidity conditions (T4), the RMSE remains below 0.1 mm. Visual comparisons (Figure 7) further highlight the reconstruction differences under different turbidity levels: subfigures (a)-(d) show the results directly obtained from the original degraded fringe images, while (e)-(h) present the optimized reconstruction results. Subfigure (i) shows the reference results based on ground truth fringe images (air environment reference). The comparison shows that the optimized reconstruction significantly outperforms the original degraded data in geometric fidelity and detail, and aligns better with the ground truth in both overall shape and local details, effectively reducing feature loss due to scattering and refraction.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

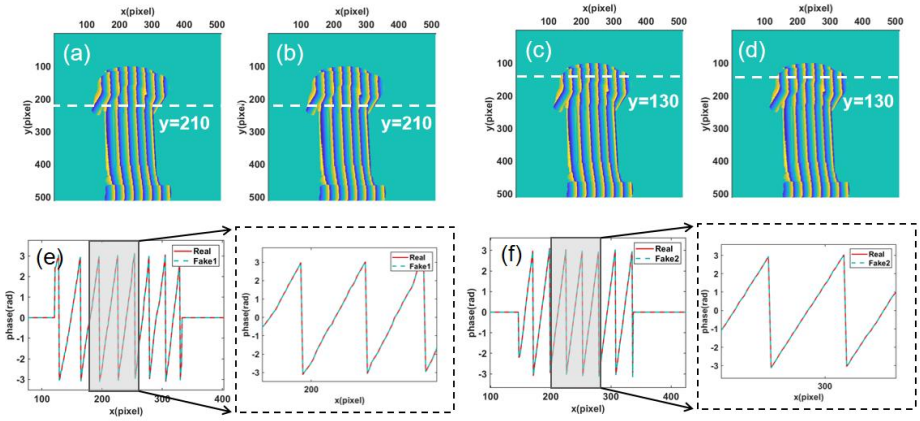


Fig.6 Comparison between the ground-truth wrapped phase and the wrapped phase computed from network-predicted fringe images

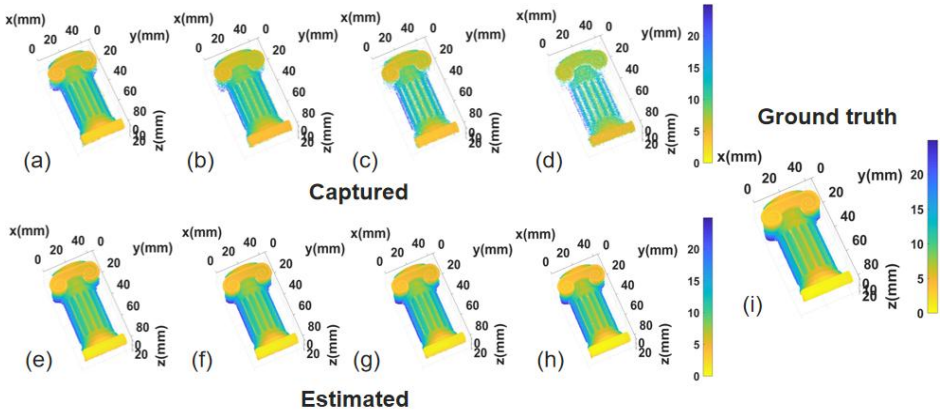


Fig.7 Comparison of the original reconstructions and the model-optimized reconstructions

4. Conclusions

This paper addresses the challenges of severe imaging degradation and complex physical modeling in underwater structured light 3D imaging by proposing a deep learning method based on a diffusion model. By introducing a data-driven strategy to model the underwater imaging degradation process, the method achieves effective recovery of structured light stripes without relying on precise physical parameter priors, thereby improving the stability and accuracy of 3D reconstruction.

Experimental results demonstrate that the proposed method exhibits good adaptability and stability under different turbidity levels and a variety of geometric target conditions. Future work will focus on expanding the data scale and scene diversity, incorporating more real-world aquatic data to enhance cross-domain generalization performance. Additionally, hybrid modeling approaches combining physical priors with deep learning models will be explored to improve the algorithm's potential for application in complex real-world aquatic environments.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [62375078], the Science and Technology Innovation Project of the Chinese Academy of Traditional Chinese Medicine [ZN2024A02], the Key Research Project Plan for Higher Education Institutions in Henan Province [24ZX011], the Training Plan for Young Backbone Teachers in Undergraduate Universities in Henan Province [2023GGJS058], and the Cultivation Programme for Young Backbone Teachers in Henan University of Technology.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P91

References

[1] Lu Lei, Jia Zhaoyi, Wu Ke, Luan Yinsen. 3D reconstruction of multi-target moving objects based on phase-shifting method[J]. Infrared and Laser Engineering, 2020, 49(3): 0303011-0303011-5.

[2] Yin W, Zhong J, Feng S, et al. Composite deep learning framework for absolute 3D shape measurement based on single fringe phase retrieval and speckle correlation[J]. Journal of Physics: Photonics, 2020, 2(4): 045009.

[3] Lyu N Q, Yu H T, Han J, et al. Structured light-based underwater 3-D reconstruction techniques: A comparative study[J]. Optics and Lasers in Engineering, 2022, 161:107344.

[4] Yuzhen Liu, Kaichen Chi, Sen Lin. Underwater Image Restoration Based on Background Light Estimation and Transmittance Optimization[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141010.

[5] Treibitz T, Schechner Y Y. Turbid scene enhancement using multi-directional illumination fusion[J]. IEEE Transactions on Image Processing, 2012, 21(11): 4662-4667.

[6] Lee Y, Gibson K B, Lee Z, et al. Stereo image defogging[C]//Proceedings of the IEEE International Conference on Image Processing, 2014: 5427-5431.

[7] Fu X Y, Liao Y H, Zeng D L, et al. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4965-4977.

[8] Zhou J C, Zhang D G, Zou P Y, et al. Retinex-based laplacian pyramid method for image defogging[J]. IEEE Access, 2019, 7: 122459-122472.

[9] Yunzhe L, Yujia X, Lei T. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media[J]. Optica, 2018, 5(10): 1181-1190.

[10] PENG Xiaohong, LIANG Zixiang, ZHANG Jun, CHEN Rongfa. Review of Underwater Image Preprocessing Based on Deep Learning[J]. Computer Engineering and Applications, 2021, 57(13): 43-54.

[11] Jie L, A. K S, M. R E, et al. WaterGAN: Unsupervised Generative Network to Enable Real-time Color Correction of Monocular Underwater Images[J]. IEEE Robotics and Automation Letters, 2017, 1-1.

[12] Mousavi M, Estrada R, Ashok A. iDehaze: Supervised Underwater Image Enhancement and Dehazing via Physically Accurate Photorealistic Simulations[J]. Electronics, 2023, 12(11): 2352.

[13] XU Xinya, LÜ Nenqing, YU Haotian, et al. Underwater 3-D measurement using fringe projection profilometry based on deep learning [J]. Aeronautical Manufacturing Technology, 2023, 66(7): 86-95.

[14] FENG Jianxin, HAN Yajun, PAN Chengsheng, et al. AL-GAN: A lightweight GAN underwater image enhancement incorporating attention mechanisms[J]. Journal of Ordnance Equipment Engineering, 2024, 45(2): 135-143.

[15] Wu W, Yao M. Optimization of Underwater Images Based on Gray World Algorithm and Jaffe-McGlamery Models[J]. Frontiers in Computing and Intelligent Systems, 2025, 11(1): 80-84.

[16] Zevatskii E Y, Lysova S S, Skripnikova A T, et al. Bouguer-Lambert-Beer Law of Absorption: Spectrophotometry in Electrolyte Solutions[J]. Russian Journal of Physical Chemistry A, 2024, 98(2): 242-247.

[17] Gurdev L, Dreischuh T, Bliznakova I, et al. Scattering of a laser beam in turbid media with forward-peaked Henyey-Greenstein indicatrices[J]. Physica Scripta, 2012, 2012(T149): 014074 (5pp).

[18] Li X, Zhang W. Application of Snell's law in reflection raytracing using the multistage fast marching method[J]. Earthquake Research Advances, 2021, 1(01): 41-48.

[19] Zhang H, He R, Fang W. An Underwater Image Enhancement Method Based on Diffusion Model Using Dual-Layer Attention Mechanism[J]. Water, 2024, 16(13): 1813-1813.

[20] Shao J, Zhang H, Miao J. Region gradient-guided diffusion model for underwater image enhancement[J]. Machine Vision and Applications, 2025, 36(2): 38-38.