



Reconstruction of transparent objects using phase shifting profilometry based on diffusion models

QINGHUI ZHANG,¹ FENG LIU,¹ LEI LU,^{1,6} ZHILONG SU,^{2,3,7} 
WEI PAN,⁴  AND XIANGJUN DAI⁵

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

²Shanghai 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

³Shaoxing Research Institute of Shanghai University, Shaoxing 312074, China

⁴Department of R & D, OPT Machine Vision Tech Co., Ltd, Dongguan 523860, China

⁵School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255049, China

⁶lulei@haut.edu.cn

⁷szloong@shu.edu.cn

Abstract: Phase shifting profilometry is an important technique for reconstructing the three-dimensional (3D) geometry of objects with purely diffuse surfaces. However, it is challenging to measure the transparent objects due to the pattern aliasing caused by light refraction and multiple reflections inside the object. In this work, we analyze the aliasing fringe pattern formation for transparent objects and then, propose to learn the front surface light intensity distribution based on the formation principle by using the diffusion models for generating the non-aliased fringe patterns reflected from the front surface only. With the generated fringe patterns, the 3D shape of the transparent objects can be reconstructed via the conventional structured light. We show the feasibility and performance of the proposed method on the data of purely transparent objects that are not seen in the training stage. Moreover, we found it could be generalized to other cases with local-transparent and translucent objects, showing the potential capability of the diffusion based learnable framework in tackling the problems of transparent object reconstruction.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Phase shifting profilometry (PSP) is one of the most popular approaches for three-dimensional (3D) reconstruction [1–3]. The camera and projector are employed in the typical system. The projector projects a set of sinusoidal fringe patterns to the object surface and the camera captures the reflected image from another angle. The captured fringe patterns are distorted because of the height of the object [4]. Based on the intensity value of the captured fringe patterns, the phase information is retrieved and the object is reconstructed based on calibration parameters. As PSP relies on the correct intensity value of the captured fringe patterns, it has good performance for the reconstruction of object with diffuse reflection surface. However, when the transparent object such as glass is reconstructed, the projected fringe pattern is reflected by the front and back surfaces simultaneously [5–10], leading to intensity aliasing in the captured fringe pattern.

Recently, transparent object reconstruction has attracted intensive attentions by the researchers. The existing algorithms can be divided into three categories: reflection-based method, refraction-based method and, deep learning-based method.

1.1. Reflection-based methods

The reflection-based method reconstructs the object based on the signal reflected by the object surface. Eren et al. [11] proposed to reconstruct the transparent object based on infrared camera and heating projecting. A laser projector is employed to heating the object by projecting laser line to the object surface. Then an infrared camera captures the image and retrieves the position of temperature peak of the line. Finally, the object is reconstructed based on triangulation relationship between the camera, projector, and object. This method has high time cost as line scanning is employed and the object needs to be heated. Instead of projecting lines, Wiedenman et al. [12] proposed to project sinusoidal fringe pattern to heat the object. However, it still requires waiting time to reduce the temperature during the projection of multiple fringe patterns. Landmann et al. [13] developed a two-step heating technique that does not require the waiting time for temperature decrease. Instead of using the temperature information, He et al. [6] suggested to reconstruct the transparent object based on the distortion information of the line-scan technology. A line laser is used to scan the transparent object and the camera captures the image aliasing from the front and back surface. The LTFtF algorithm is proposed to separate the aliasing image and the shape information of the transparent object is reconstructed.

1.2. Refraction-based methods

The refraction-based method reconstructs transparent objects based on tracking the path of the light when passing through the object. Guo et al. [14] proposed to reconstruct the transparent object by analyzing the distorted image capturing behind the object. A cosine fringe pattern is employed and the local change in the period is retrieved. The surface profile is simulated and optimized to achieve minimal error with experimental data. Loper et al. [15] employed an approximate differentiable renderer (DR) that models the relationship between changes in model parameters and image observations when the transparent object is reconstructed. Lyu et al. [16] reconstructed the transparent object by establishing a mapping between the camera view rays and locations on the background firstly. Then, the transparent object was placed on the workbench, and the camera records the background pattern distorted by a transparent object. The 3D information of the object was obtained by tracing these light rays from the background and previously established mappings.

1.3. Deep learning-based methods

Recently, the deep learning method achieved quality results in 3D reconstruction. Li et al. [17] proposed a NeTO method to get the information of 3D geometry of solid transparent object from a 2D image. The surface of object is expressed by implicit Signed Distance Function (SDF) firstly; then, the SDF field is optimized via volume rendering with a self-occlusion aware refractive ray tracing. The implicit representation enables us to reconstruct the object with high quality even with a limited set of images and the self-occlusion aware strategy makes it possible to accurately reconstruct the self-occluded regions. Mathai et al. [18] combined compressive sensing (CS) and super-resolution convolutional neural network (SRCNN) to reconstruct the surface of the transparent objects. The transparent object's details are extracted by a single pixel detector during the surface reconstruction. The SRCNN could avoid the influence of low-quality images owing to speckles and deformations and the CS could reduce the size of the sample required for training. Guo et al. [19] proposed to split a scene into transmitted and reflected components; then, model the two components with separate neural radiance fields. The geometric priors are exploited and training strategies are designed to achieve reasonable decomposition results; at last, the NeRFReN is proposed to reconstruct the transparent object with complex reflection. Shao et al. [20] proposed a method to get the information of the surface of the transparent objects by using polarimetric cues. The object's geometry is represented as a neural network, while the polarization render is capable of rendering the object's polarization images from the given shape

and illumination configuration. The reflection percentage is calculated by a ray tracer and then used for weighting the polarization loss. The object is reconstructed based on the polarization dataset with multi-view transparent shapes. Sha et al. [10] introduced the TransNeXt network to enhance the features of the sample of the transparent objects. First, the transparent shape is initialized with a visual hull reconstructed with the contours obtained by the TOM-Net. Then, the normal reconstruction network is constructed to estimate the normal values. Finally, the transparent object is reconstructed using the TransNeXt network.

2. Methodology

2.1. Principle of traditional PSP

The captured fringe patterns for N-step PSP can be expressed as follows:

$$S_i(x, y) = a + b\cos(\varphi(x, y) + 2\pi(i - 1)/N), \quad (1)$$

where $i = 1, \dots, N$, $S_i(x, y)$ is the intensity distribution of the i -th fringe pattern, a is the ambient light intensity, b is the amplitude of the fringe pattern intensity, and $\varphi(x, y)$ is the phase distribution.

The phase map $\varphi(x, y)$ can be obtained by

$$\varphi(x, y) = \arctan \frac{-\sum_{i=1}^N S_i(x, y)\sin 2\pi(i - 1)/N}{\sum_{i=1}^N S_i(x, y)\cos 2\pi(i - 1)/N}. \quad (2)$$

The range of phase map is wrapped into $-\pi$ to π , which leading to ambiguity among different fringes. The unwrapped phase map with monotonous value is obtained by phase unwrapping and the phase discontinuities are removed. At last, the object is reconstructed based on the phase information and calibration parameters.

2.2. Analysis of fringe pattern aliasing

Equation (1) describes the fringe pattern reflected by the object with diffuse surface. When the transparent object is reconstructed, the illumination is reflected and refracted by the front and back surface of the object as shown in Fig. 1. Assume F is the light reflected by the front surface and captured by the camera directly. When F pass through the front surface and reflected by the back surface, the captured light can be noted as B . M is the light which is multiple reflected between the front and back surface and finally captured by the camera. Normally, F and B have the strongest intensity while the component M has the weakest intensity.

By applying the above reflection and refraction model to PSP, the captured fringe pattern is composed by four sources at time t , i.e., the fringe pattern reflected by the front surface $f(t)$, the fringe pattern reflected by the back surface with the single reflection $b(t)$, the multiple reflection $m(t)$ and the background light captured by the camera a . Figure 2 shows the procedure schematically. According to the fundamental imaging principle, the captured fringe pattern of the transparent object can be expressed as the summation of the four components above at the moment t by

$$s(t) = a + m(t) + b(t) + f(t). \quad (3)$$

We suppose that the projector and the camera work synchronously and the projector switches the projected pattern at the moment t_i corresponding to the start of the camera exposure. Consequently, the camera captures the fringe pattern image at this moment, denoted by $s_i(t)$. Let $t_d = t_{i+1} - t_i$ be the exposure time of the camera, the i -th fringe pattern frame S_i captured by the

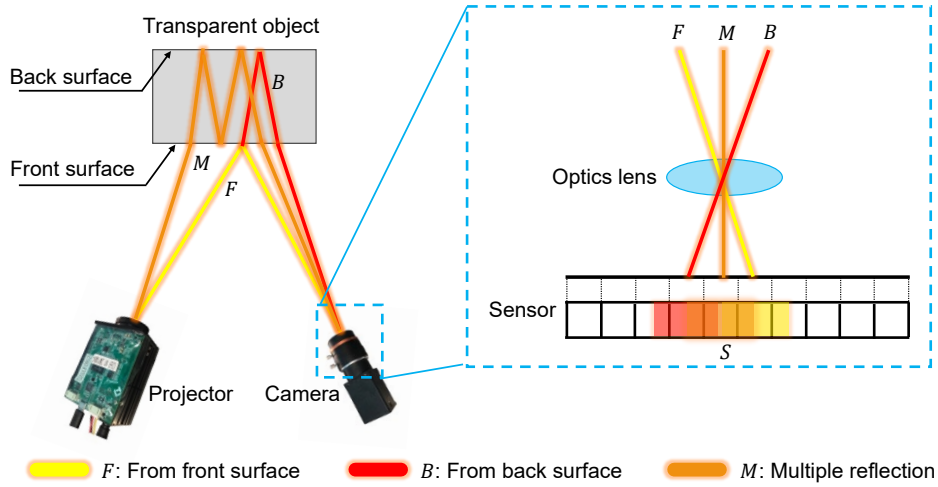


Fig. 1. Aliasing of observed fringe pattern for transparent object.

camera is given by

$$S_i = \int_{t_i}^{t_{i+1}} s_i(t) dt = k_f \int_{t_i}^{t_{i+1}} f_i(t) dt + k_b \int_{t_i}^{t_{i+1}} b_i(t) dt + k_m \int_{t_i}^{t_{i+1}} m_i(t) dt + at_d, \quad (4)$$

where the factors k_f , k_b and k_m are used to balance the light intensity contribution of each component, considering the differences in the action of light on different transparent objects and the errors generated by the camera during the actual sampling. For simplicity, we denote each of the integral terms on the right-hand side of Eq. (4) by a symbol, i.e., $F_i = k_f \int_{t_i}^{t_{i+1}} f_i(t) dt$, $B_i = k_b \int_{t_i}^{t_{i+1}} b_i(t) dt$, and $M_i = k_m \int_{t_i}^{t_{i+1}} m_i(t) dt$, and omit the subscription, obtaining

$$S = F + B + M + A, \quad (5)$$

with $A = at_d$. In this way, the fringe pattern component F , together with the background light, that can be used to reconstruct the front surface and can be express as $F = S - B - M$. With this baseline, it is possible to estimate the fringe pattern of the front surface from the aliasing fringe pattern. We next train a network p within the diffusion model deep learning framework for sampling the fringe pattern F as

$$F^* \sim p(S; \theta), \quad (6)$$

where F^* is the estimated fringe pattern and θ is the network parameters to be learned.

2.3. Diffusion models

Diffusion models [21] have gained increasingly attention due to its excellent performance to generate seemingly realistic images. They are proved to accurately learn the true distribution of data from some given priori samples with improved sample quality and more stable training protocols compared with some other deep networks such as generative adversarial nets [22], leading to excellent results in work related to estimating sample distributions.

When it comes to our problem, diffusion models are trained to reverse a stochastic forward process, as shown in Fig. 3, that iteratively transfers a 2D data of fringe patterns, denoted by $\mathbf{x}_0 \in \mathbb{R}^{H \times W}$, drawn from the underlying distribution $\mathbf{x}_0 \sim q(F)$ to a full prior Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{I})$ in T steps. As the process acts as gradually adding Gaussian noise to the fringe pattern \mathbf{x}_0 for obtaining an approximate posterior, we can represent it as a fixed Markov chain with simple Gaussian

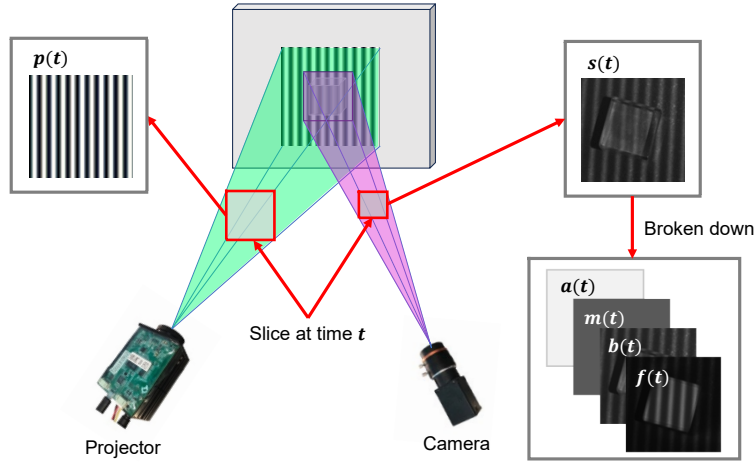


Fig. 2. Schematic diagram of light intensity distribution of transparent objects.

transitions parameterized by a given variance weighting protocol $\beta_t \in (0, 1)$, $t = 1, \dots, T$ as follows:

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \prod_{t=1}^T \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}). \quad (7)$$

Equation (7) shows that the weight β_t controls the added Gaussian noise in the diffusion step t , by denoting $\alpha_t = 1 - \beta_t$, allowing us to sample fringe pattern \mathbf{x}_t at diffusion step t through $\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\mathbf{e}$ with $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\bar{\alpha} = \prod_{k=1}^t \alpha_k$.

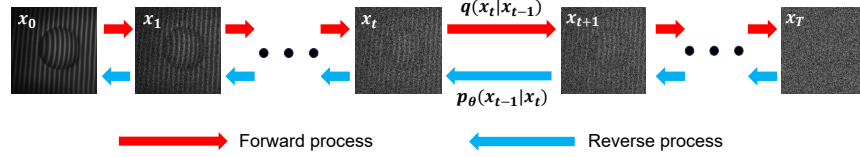


Fig. 3. Forward reverse diffusion processes of our fringe pattern.

With the diffusion model learning, the reverse process $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ can be represented by the neural network $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ with parameters θ to generate new samples $\mathbf{x}^* \sim q(F^*)$ given that \mathbf{x}_t . This rises a corresponding reverse Markov chain to solve the original distribution via

$$p_\theta(\mathbf{x}_{0:T}) = p_\theta(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad (8)$$

with $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}(\mathbf{x}_t, t))$, where $\boldsymbol{\mu}_\theta$ is the inferred mean values and $\boldsymbol{\Sigma}$ the covariance that could be set to as $\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t \mathbf{I}$ in purely time dependent formation. As shown in Eq. (8), the reverse process is refined into a series of processes to find \mathbf{x}_{t-1} from \mathbf{x}_t . Actually, the \mathbf{x}_t is not produced by the forward process but by sampling from the standard normal distribution, i.e. $\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Hence, $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ in the reverse process can be simplified as follows:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}). \quad (9)$$

In this work, we propose to design the network p_θ based on the U-Net architecture with residual block extension, as show in Fig. 4. The network is built with a conditional control consisting of down-sampling and up-sampling, which we improved according to the feature of the aliased fringe pattern by merging the observed fringe pattern sample, denoted by \mathbf{y} . In the

down-sampling stage, the number of channels in the image gradually increases while the size of the image decreases, which is a critical step in the process of information extraction. Conversely, in the up-sampling stage, the size of the image gradually increases, whereas the number of channels decreases, which is the process of information recovery. A certain layer of the network in the up-sampling phase inputs the layer output data below it along with the output data from the layer of the network in the down-sampling phase that corresponds to it.

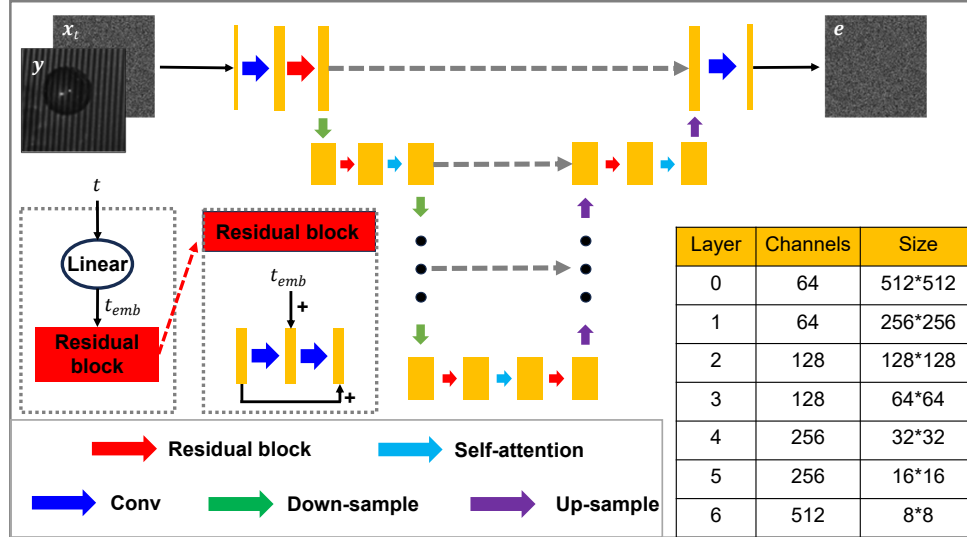


Fig. 4. Architecture of the U-Net based network p_θ .

2.4. Training and sampling

To train the model in Fig. 4, we obtain the sample x_0 of the fringe pattern distribution F reflected from the front surface by painting a developer onto it. At the same time, an aliasing fringe pattern y of the corresponding distribution S can be obtained as a conditional for training, so that $y \sim q(S)$ and $x_0 \sim q(F|y)$. The underlying distribution the pattern F represented by the sample x_0 is gradually disturbed to close to the state of the pattern y by following the forward diffusion process in Eq. (7). With the corresponding relationship, we can train the network to learn the diffusion process in Eq. (9) via the following loss function:

$$L = \left\| e - p_\theta \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} e, t, y \right) \right\|^2. \quad (10)$$

Then, the trained network can be used to generate the samples of fringe pattern F by following the reverse process in Eq. (8).

The training and sampling pipelines are schematically shown in Fig. 5. In the training stage, we take the Gaussian random noise e , the original sample x_0 , and the observed sample y as inputs in each epoch and then, loop over the randomly and uniformly sampled time step $t \in [0, T]$. For each step t , the computed x_t (from e and x_0) and y are fed into the network to evaluate the forward pass, resulting in the noise map e^* . By comparing e^* with the given noise e according to the loss function in Eq. (10), the network parameters are optimized by the stochastic gradient descent based back-propagation until convergence. In our training practice, the total number of diffusion model steps $T = 2600$; β_t increase linearly with the time step t , where $\beta_1 = 0.0001$ and $\beta_T = 0.002$. The loss curve of the training process is shown in Fig. 6. We can see that, even if

the loss values change violated, but they show a health convergence behavior by following an exponential falling paradigm to a lower and stable level.

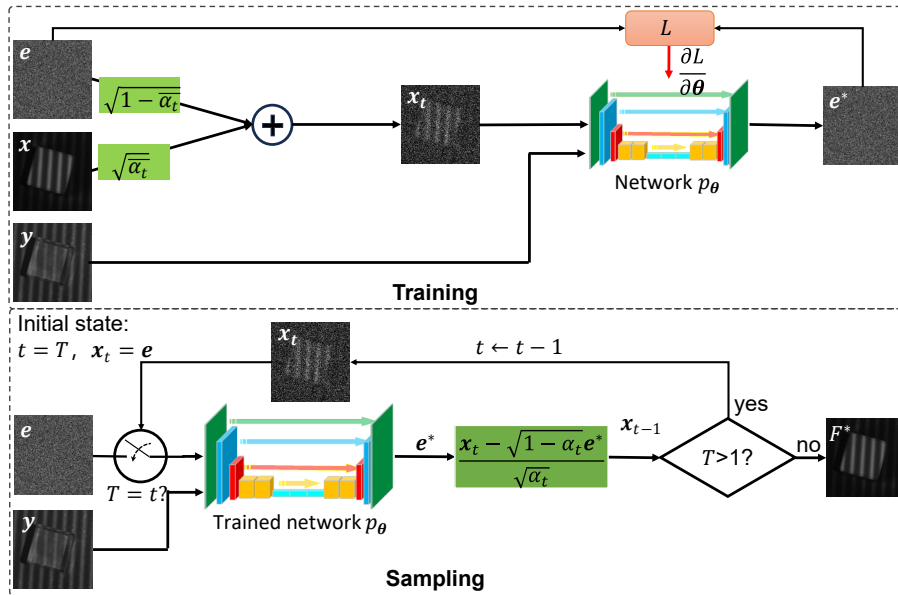


Fig. 5. Network training with the forward diffusion process and reverse sampling to recover the fringe pattern F of the surface to be measured.

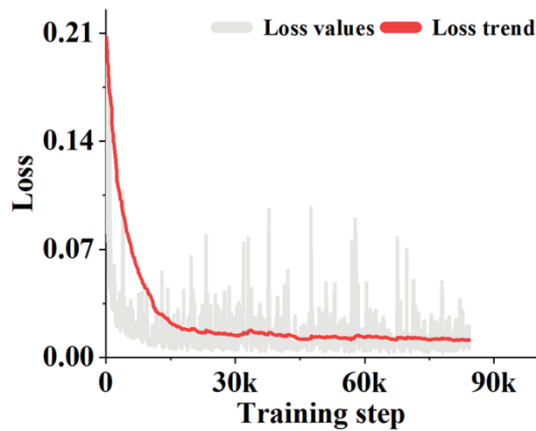


Fig. 6. Loss curve of the model training.

In the sampling stage, x_t is initialized to as the Gaussian noise e and t is initialized to the last step T . That relevant information is fed into the trained network p_θ to inference the noise state and then, to estimate x_{t-1} by incorporating its previous version x_t . This process corresponds to the reverse diffusion given in Eq. (8) and is performed iteratively until $t = 0$, yielding an estimated fringe pattern F^* of the front surface intensity distribution.

3. Results and discussion

3.1. Production of datasets

The experimental equipment is shown in Fig. 7. The camera (S2 Camera TJ1300UM with resolution 2048×1536 pixels) and DLP projector with resolution 1280×1024 pixels are employed, and the 4-step PSP algorithm is used to perform the reconstruction. In order to fully verify the effectiveness of the proposed method, four objects with different types are employed in the experiments: a transparent cube, a transparent hemisphere, a translucent object and a partially-transparent object. The computational platform for the model training is configured by an Intel Xeon E5-2620 CPU (2.10 GHZ), 72 GB RAM and a GPU GeForce GTX 3060Ti.

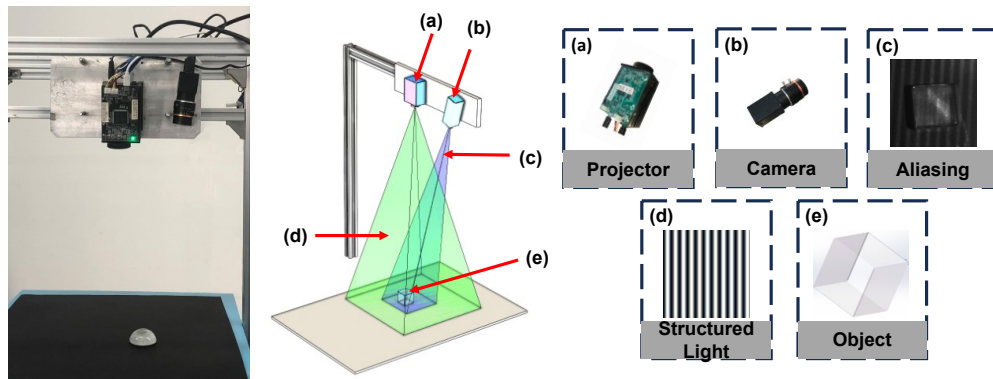


Fig. 7. Schematic diagram of the experimental equipment.

It is difficult to acquire pure fringe patterns on the front surface of transparent objects. We tried the simulation method, the surface sticker method, the spray developer method. For the simulation method, the differences between the images obtained by simulation and the experimental are significant. The surface sticker method can obtain correct fringe pattern for the front surface by pasting cardboard on the transparent object, but it is difficult to operate in practice especially for the non-planar objects. In the end, we selected the spray developer method to build the dataset.

The pure fringe pattern of the front surface and its corresponding aliasing fringe pattern are captured. For each object, 15,000 image pairs are employed for training and 1500 image pairs for the test. The positions and poses of the objects are randomly placed.

3.2. Reconstruction results

The four different types of objects reconstructed in the experiments are shown in Fig. 8. The objects in Fig. 8(a) and (c) is made of acrylic and Fig. 8(b) and (d) are made of glass. After completing the training of the diffusion model, the aliasing fringe patterns are sent into the diffusion model and the corresponding fringe pattern of the front surface is estimated, following by the 3D reconstruction with PSP. For each object, we analyzed the performance of the proposed method from three aspects: fringe pattern intensity error, phase error and reconstruction error.

Transparent objects. Firstly, we conduct the reconstruction testing by using the transparent objects with those shown in Fig. 8(a) and (b). The side length and the radius of the transparent cube and hemisphere are 25 mm and 20 mm , respectively. A set of aliasing fringe patterns are obtained as shown in Figs. 9(a) and 10(a), respectively. Then, the trained model is applied to estimate the fringe patterns of the surfaces being measured, as shown in Figs. 9(b) and 10(b). It is apparent that the fringe patterns reflected by both front surfaces are recovered from the aliasing image. For the convenience of performance comparison, the fringe pattern after spray developer is also captured as shown in Figs. 9(c) and 10(c), respectively. Then, the difference between

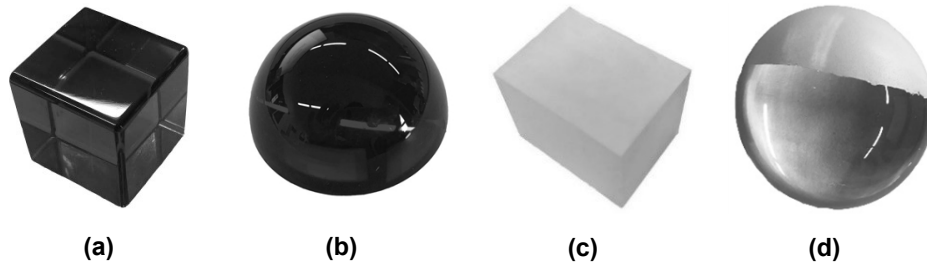


Fig. 8. Four types of testing objects: (a) transparent cube, (b) transparent hemisphere, (c) translucent object, and (d) partially transparent object.

the estimated fringe pattern and the one after spray developer is calculated in a normalization manner. The results are shown in Figs. 9(d)-(e) and 10(d)-(e). By line sampling (the red dot lines) comparison, it can be found that although the intensity values are not identical strictly, the estimated fringe pattern has the same characteristic as the ones after spray developer.

Then, the wrapped phase maps are calculated by 4-step PSP, resulting in the front surface reconstruction, as shown in Figs. 9(f) and 10(f). It can be seen that the inferred fringe patterns are capable of reconstructing the measured surface with good quality. As the reconstruction accuracy is dependent on the phase computing, we also investigate the phase recover performance by given the inferred fringe patterns and the directly measured fringe patterns in (c). The resulting phase maps are shown in Figs. 9(g)-(h) and 10(g)-(h). To closely investigate the quality, we compare the phase distributions along the line samplers in Figs. 9(i) and 10(i), respectively. Obviously, the phase values in the object regions present a close correlation.

Translucent objects. For such cases, the refraction-based method does not work as the light cannot pass through the object. The captured aliasing fringe pattern, the estimated fringe pattern and captured fringe pattern after spray developer are given in Fig. 11(a)-(c). The quality estimation of the inferred fringe pattern is shown in Fig. 11(e). The corresponding reconstruction result and the phase maps are presented in Fig. 11(f)-(i).

Partially-transparent objects. For such cases, most of the existing reconstruction methods for this purpose have difficulty in reconstructing the entire object completely as the transparent parts and the solid parts require different reconstruction algorithms. However, the proposed method can effectively cope with this situation as we take the light that comes from non-front surface as noise, and they are removed by the trained diffusion model. This is demonstrated in the experiment results as shown in Fig. 12.

Finally, we calculate the RMS (root mean square) error of the reconstruction for each object, listed in Table 1. The results show that the proposed method achieves quality reconstruction as we expected.

Table 1. Reconstruction errors for the tested objects

Object	Transparent cube	Transparent hemisphere	Translucent object	Partially-transparent object
RMS(mm)	0.0420	0.0423	0.0437	0.0388

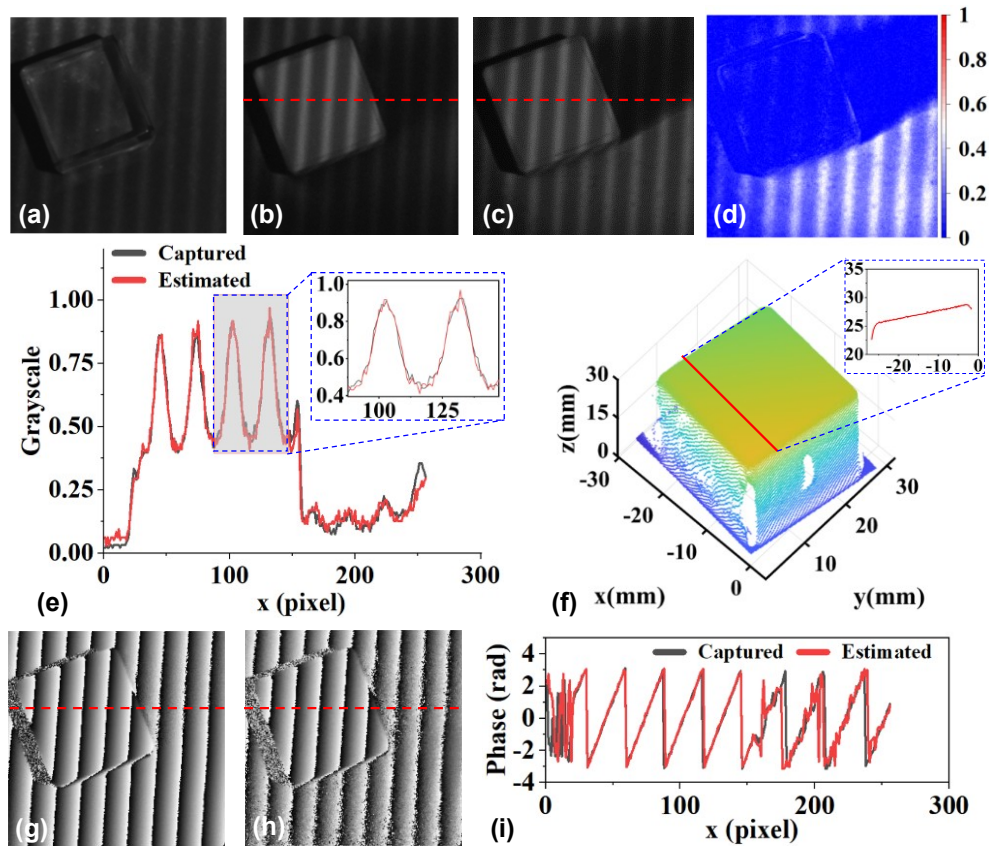


Fig. 9. Demonstration of transparent cube reconstruction. (a) Captured aliasing fringe pattern, (b) inferred fringe pattern of the front surface by the proposed method, (c) captured fringe pattern after spray developer, (d) the difference between (b) and (c); (e) comparison of the sampled intensity values along the lines in both (b) and (c); (f) the reconstructed result by the proposed method, (g) and (h) show the wrapped phase maps by the fringe patterns after spray developer and by the estimated fringe patterns respectively, and (i) the comparison of the wrapped phase maps between (g) and (h).

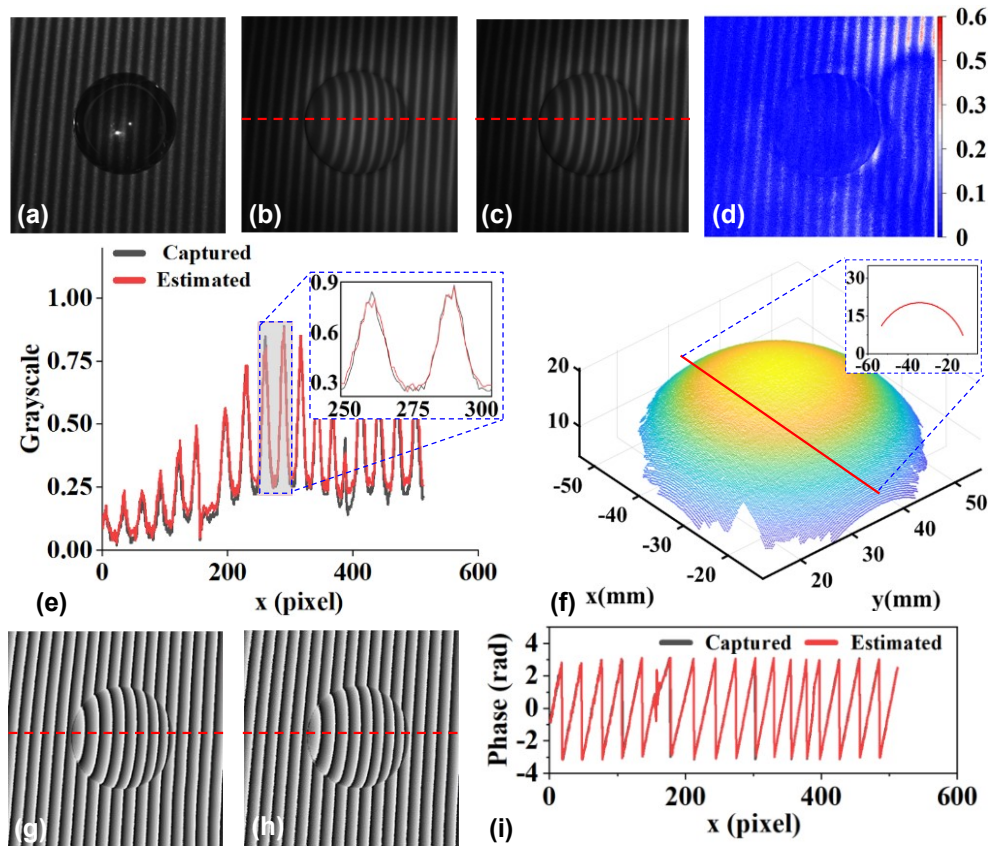


Fig. 10. Demonstration of transparent hemisphere reconstruction. (a) Captured aliasing fringe pattern, (b) inferred fringe pattern of the front surface by the proposed method, (c) captured fringe pattern after spray developer, (d) the difference between (b) and (c); (e) comparison of the sampled intensity values along the lines in both (b) and (c); (f) the reconstructed result by the proposed method, (g) and (h) show the wrapped phase maps by the fringe patterns after spray developer and by the estimated fringe patterns respectively, and (i) the comparison of the wrapped phase maps between (g) and (h).

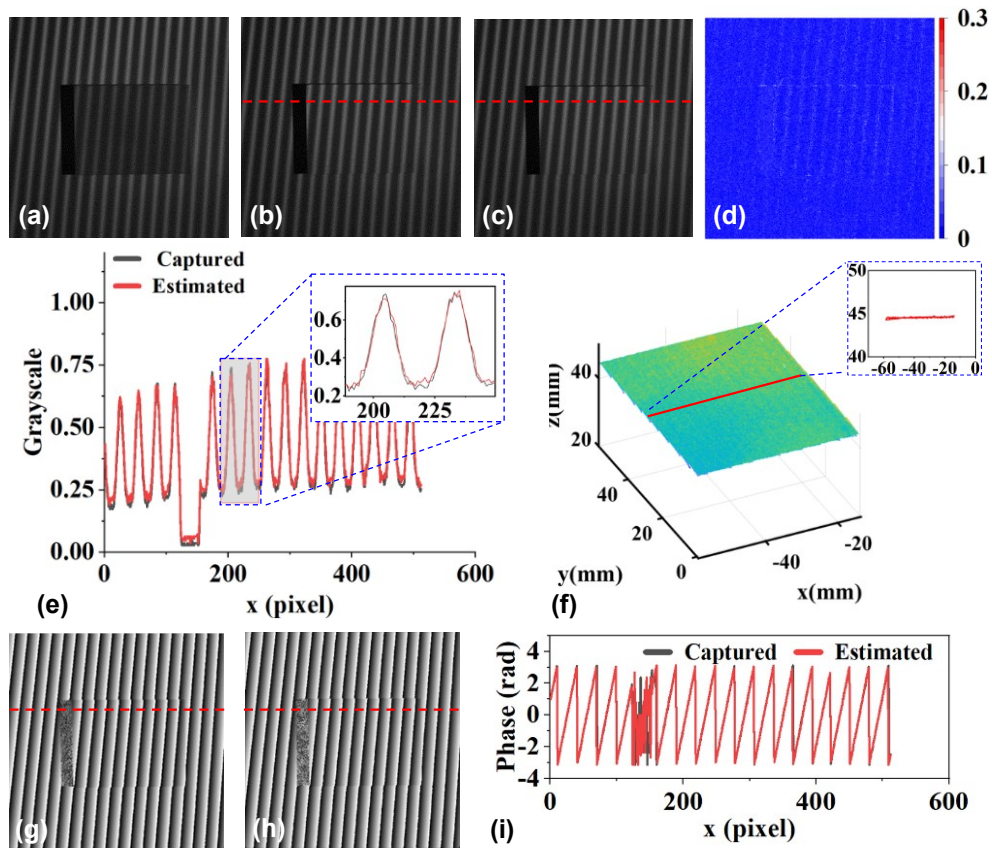


Fig. 11. Reconstruction demonstration of translucent object. (a) Captured aliasing fringe pattern, (b) inferred fringe pattern of the front surface by the proposed method, (c) captured fringe pattern after spray developer, (d) the difference between (b) and (c); (e) comparison of the sampled intensity values along the lines in both (b) and (c); (f) the reconstructed result by the proposed method, (g) and (h) show the wrapped phase maps by the fringe patterns after spray developer and by the estimated fringe patterns respectively, and (i) the comparison of the wrapped phase maps between (g) and (h).

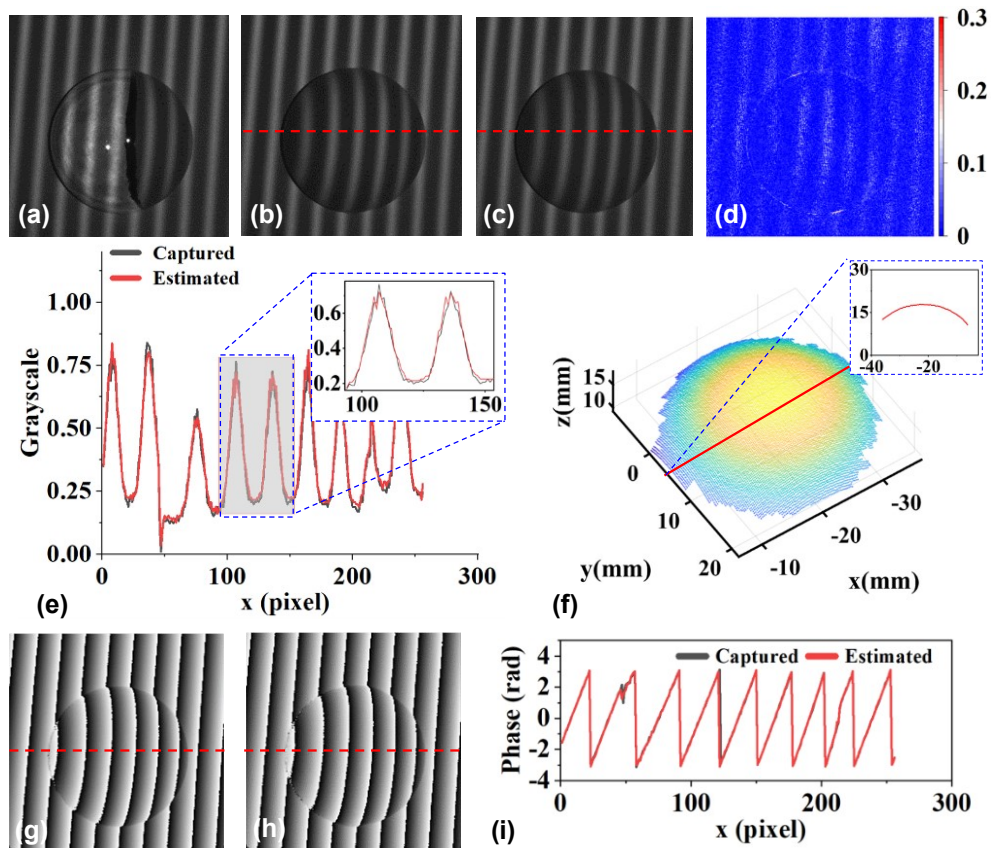


Fig. 12. Reconstruction demonstration of partially-transparent object. (a) Captured aliasing fringe pattern, (b) inferred fringe pattern of the front surface by the proposed method, (c) captured fringe pattern after spray developer, (d) the difference between (b) and (c); (e) comparison of the sampled intensity values along the lines in both (b) and (c); (f) the reconstructed result by the proposed method, (g) and (h) show the wrapped phase maps by the fringe patterns after spray developer and by the estimated fringe patterns respectively, and (i) the comparison of the wrapped phase maps between (g) and (h).

4. Conclusion

In this paper, we present a deep learning method based on the diffusion models to estimate the fringe patterns reflected by the transparent targets from the pattern distribution measurements. To this end, the formation of the aliased fringe pattern is explored by tracing the illumination light paths pass through and reflected from the front surface of the transparent objects. According to the aliasing formation, a U-Net like deep network is introduced and trained within the diffusion model learning framework for sampling the front-surface reflected fringe patterns from a given Gaussian noise map and observed fringe patterns, which are then used for 3D shape reconstruction via the conventional structured light technique. The experimental results show that the proposed method can effectively reconstruct various types of transparent objects, including transparent cubes, hemispheres, translucent objects, and partially-transparent objects, with promising accuracy and reliability. The findings suggest that the diffusion-based learnable framework holds great promise in addressing the challenges of transparent object reconstruction, potentially opening new avenues for advancements in the field of 3D imaging and object reconstruction.

It should be noted that, according to the principle of diffusion models, the learned fringe patterns come from Gaussian noise, leaving residual noise that is invisible but may reduce the phase estimation accuracy as well as the quality of 3D reconstruction. This problem is solvable by increasing the diffusion step number. However, we experimentally found the larger time step number will lead to an increase of the computational cost in both training and sampling processes. Therefore, further effort should be made to solve this problem or to find a suitable tradeoff between quality pursuing and efficiency in practice.

Funding. National Natural Science Foundation of China (12002197, 62375078); General Science Foundation of Henan Province (222300420427); Key Research and Development Program of Henan Province (231111222100); Key Research Project Plan for Higher Education Institutions in Henan Province (24ZX011); Cultivation Programme for Young Backbone Teachers in Henan University of Technology.

Disclosures. The authors declare no conflicts 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. S. Feng, Chao Zuo, Tianyang Tao, *et al.*, "Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry," *Opt. Lasers Eng.* **103**, 127–138 (2018).
2. L. Lu, Vignesh Suresh, Yi Zheng, *et al.*, "Motion induced error reduction methods for phase shifting profilometry: A review," *Opt. Lasers Eng.* **141**, 106573 (2021).
3. C. Zuo, Shijie Feng, Lei Huang, *et al.*, "Phase shifting algorithms for fringe projection profilometry: A review," *Opt. Lasers Eng.* **109**, 23–59 (2018).
4. L. Lu, Hao Liu, Hongliang Fu, *et al.*, "Kinematic target surface sensing based on improved deep optical flow tracking," *Opt. Express* **31**(23), 39007–39019 (2023).
5. K. Chen, S. Wang, B. Xia, *et al.*, "Tode-trans: Transparent object depth estimation with transformer," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, (IEEE, 2023), 4880–4886.
6. K. He, Congying Sui, Tianyu Huang, *et al.*, "3D Surface reconstruction of transparent objects using laser scanning with LTFtF method," *Opt. Lasers Eng.* **148**, 106774 (2022).
7. K. He, Congying Sui, Tianyu Huang, *et al.*, "3D surface reconstruction of transparent objects using laser scanning with a four-layers refinement process," *Opt. Express* **30**(6), 8571–8591 (2022).
8. Z. Li, Y.-Y. Yeh, M. Chandraker, *et al.*, "Through the looking glass: neural 3D reconstruction of transparent shapes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020), 1262–1271.
9. N. J. Morris and K. N. Kutulakos, "Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography," in *2007 IEEE 11th International Conference on Computer Vision*, (IEEE, 2007), 1–8.
10. X. Sha, "End-to-end three-dimensional reconstruction of transparent objects with multiple optimization strategies under limited constraints," SSRN, 10.2139/ssrn.4632280SSRN:4556675 2024.
11. G. Eren, Olivier Aubreton, Fabrice Meriaudeau, *et al.*, "Scanning from heating: 3D shape estimation of transparent objects from local surface heating," *Opt. Express* **17**(14), 11457–11468 (2009).
12. E. Wiedenmann, M. Afrough, S. Albert, *et al.*, "Long wave infrared 3D scanner," in *Dimensional Optical Metrology and Inspection for Practical Applications IV*, (SPIE, 2015), 82–92.
13. M. Landmann, Henri Speck, Patrick Dietrich, *et al.*, "High-resolution sequential thermal fringe projection technique for fast and accurate 3D shape measurement of transparent objects," *Appl. Opt.* **60**(8), 2362–2371 (2021).

14. H. Guo, Haowen Zhou, Partha P. Banerjee, *et al.*, “Use of structured light in 3D reconstruction of transparent objects,” *Appl. Opt.* **61**(5), B314–B324 (2022).
15. M. M. Loper and M. J. Black, “OpenDR: An approximate differentiable renderer,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part VII 13*, (Springer, 2014), 154–169.
16. J. Lyu, Bojian Wu, Dani Lischinski, *et al.*, “Differentiable refraction-tracing for mesh reconstruction of transparent objects,” *ACM Trans. Graph.* **39**(6), 1–13 (2020).
17. Z. Li, X. Long, Y. Wang, *et al.*, “NeTO: neural reconstruction of transparent objects with self-occlusion aware refraction-tracing,” *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 18501–18511.
18. A. Mathai, L. Mengdi, S. Lau, *et al.*, “Transparent object reconstruction based on compressive sensing and super-resolution convolutional neural network,” *Photonic Sens.* **12**(4), 220413 (2022).
19. Y.-C. Guo, D. Kang, L. Bao, *et al.*, “Nerfren: Neural radiance fields with reflections,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 18409–18418.
20. M. Shao, C. Xia, D. Duan, *et al.*, “Polarimetric Inverse Rendering for Transparent Shapes Reconstruction,” *arXiv*, arXiv:2208.11836 (2022).
21. J. Ho, A. Jain, P. Abbeel, *et al.*, “Denoising diffusion probabilistic models,” *Advances in neural information processing systems* **33**, 6840–6851 (2020).
22. P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” *Advances in neural information processing systems* **34**, 8780–8794 (2021).