Research Article

ADVANCED IMAGING

Real-time 3D imaging based on ROI fringe projection and a lightweight phase-estimation network

Yueyang Li, Junfei Shen, Zhoujie Wu, Yajun Wang, and Qican Zhang*
College of Electronics and Information Engineering, Sichuan University, Chengdu, China

Abstract. Realizing real-time and highly accurate three-dimensional (3D) imaging of dynamic scenes presents a fundamental challenge across various fields, including online monitoring and augmented reality. Currently, traditional phase-shifting profilometry (PSP) and Fourier transform profilometry (FTP) methods struggle to balance accuracy and measurement efficiency simultaneously, while deep-learning-based 3D imaging approaches lack in terms of speed and flexibility. To address these challenges, we proposed a real-time method of 3D imaging based on region of interest (ROI) fringe projection and a lightweight phase-estimation network, in which an ROI fringe projection strategy was adopted to increase the fringe period on the tested surface. A phase-estimation network (PE-Net) assisted by phase estimation was presented to ensure both phase accuracy and inference speed, and a modified heterodyne phase unwrapping method (MHPU) was used to enable flexible phase unwrapping for the final 3D imaging outputs. The experimental results demonstrate that the proposed workflow achieves 3D imaging with a speed of 100 frame/s and a root mean square (RMS) error of less than 0.031 mm, providing a real-time solution with high accuracy, efficiency, and flexibility.

Keywords: 3D imaging; deep learning; fringe analysis; phase demodulation; real time.

Received Jun. 11, 2024; revised manuscript received Aug. 5, 2024; accepted Sep. 2, 2024; published online Sep. 23, 2024.

[DOI: 10.3788/AI.2024.10008]

1. Introduction

The technique of structured-light three-dimensional (3D) imaging based on fringe projection, also known as fringe projection profilometry (FPP)[1–4], has long been a focal point in both academic research and industrial applications due to its efficiency, precision, and cost-effectiveness on converting data into 3D models. Through the processing of 3D imaging data, precise measurement and analysis of object dimensions, shapes, and surface features have been well achieved in static scenes[5–7]. However, various limiting factors exist in the 3D imaging of dynamic scenes, such as inter-frame movements of objects and the low efficiency of pattern encoding, preventing FPP from achieving the same level of precision as in static scenes. Furthermore, in scenarios such as augmented reality and online monitoring, it is crucial to perform high-precision 3D imaging in real-time, which brings challenges to FPP algorithms for their complexity and robustness.

FPP-based 3D reconstruction consists of four pivotal stages: fringe pattern projection and modulation, phase demodulation, phase unwrapping, and 3D calibration and reconstruction. To solve the problems of dynamic scene measurement, these stages have been optimized by many researchers over the past decade to enhance the adaptability of the algorithms.

One intuitive way is to reduce the number of required projected patterns using composite fringes. Some researchers have proposed different ways to embed additional encoding information into the fringes, such as binary coding patterns[8], speckle patterns[9,10], triangular waves[11], and coded phases[12]. Through corresponding decoding algorithms, the composite fringes are separated into sinusoidal intensity and another portion used to compute fringe orders. This approach employs encoded information to assist with phase unwrapping, obviating the necessity for directly projecting additional images.

*Address all correspondence to Qican Zhang, zqc@scu.edu.cn

Besides, the reuse of fringes or redundant information can also reduce the number of images required for phase unwrapping. Wu *et al*. proposed cyclic complementary Gray-code[13] and shifting Gray-code[14] by re-encoding Gray-code sequences in both temporal and spatial domains, which circumvent edge errors while simultaneously reducing the number of fringes by one. Zuo *et al*.[15] assumed that the background light intensity of the fringes remains constant throughout the measurement process and utilized the same background light intensity for two adjacent sets of phase-shifting fringes to compute phase values, while his alternative method (2 + 2)[16] employed ramp patterns with a linear intensity variation to compute phase. The previously mentioned methods have been proven to be beneficial in improving measurement efficiency. However, due to the utilization of multiple fringes, the motion-induced error will be introduced, leading to periodic error distributions.

An alternative approach involves modulating multiple fringe patterns with different carrier frequencies along orthogonal directions and combining them into a single pattern image[17,18]. The wrapped phases corresponding to these patterns can be separated from the Fourier spectrum of the composite fringe pattern and then unwrapped using temporal phase unwrapping (TPU)[19,20]. While frequency multiplexing can enhance the efficiency of FPP, careful selection of frequencies is necessary to prevent spectrum aliasing. Additionally, the filtering operations involved in Fourier fringe analysis can pose challenges for achieving high-accuracy measurements on complex surfaces.

Recently, numerous deep learning-based FPP methods have been developed and have demonstrated outstanding performance in various aspects, including fringe enhancement[21], fringe analysis[22,23], and high-dynamic-range measurement[24,25]. Among these, some methods are particularly suitable for dynamic scene measurement due to their reduced requirement for the number of fringes. Sam Van der Jeught *et al*.[26] designed a convolutional neural network (CNN) to directly predict depth from a single fringe pattern without any additional intermediate processing steps, showcasing the potent mapping capability of deep learning technology. However, such direct mapping approaches often struggle to accurately learn the physical parameters during the calibration process, resulting in reduced precision when dealing with complex surfaces.

Many methods have instead adopted multi-stage processing pipelines, wherein deep learning is employed to obtain accurate fringe patterns or retrieve phase information, followed by the use of traditional methods to convert phase into 3D shapes—for instance, separating different grayscale phase-shifted fringes from a single-frame color composite fringe pattern through CNN[27] and then using traditional phase-shifting profilometry (PSP)[28,29] and TPU methods to compute the 3D results. Alternatively, the method of extracting phase information and coarse unwrapped phase from a frequency multiplexing fringe[30], followed by utilizing TPU to obtain the unwrapped phase, can be used. Benefiting from the support of traditional methods, such approaches can achieve greater accuracy when dealing with complex surfaces, but the large number of parameters of neural networks and the complexity of network structures make it hard to achieve real-time performance.

To address this issue, Li *et al.*[31] utilized network architecture search (NAS) to automatically design a lightweight phase demodulation network and combined it with a modified heterodyne phase unwrapping (MHPU) method, successfully achieving real-time 3D imaging at 58 frame/s. As the speed increases, the accuracy of this method tends to decrease compared to models with larger parameter numbers. Yin *et al*.[32] introduced a physics-informed deep learning method for phase retrieval, which replaces the filtering process of Fourier fringe analysis with two learnable filters. By combining this approach with a lightweight network, they achieved high-precision measurements in low latency. However, this method requires prior knowledge similar to Fourier transform profilometry (FTP) to determine the positions, sizes, and initial weights of the learnable filters, which may limit its flexibility when faced with different frequencies.

In order to strike a better balance between the measuring accuracy, speed, and flexibility, we proposed a real-time 3D imaging method based on region of interest (ROI) projection and a phase-estimation network. For fringe projection, by determining an appropriate projection resolution for fringes, we effectively increased the number of fringe periods within the tested target (i.e., ROI) and thereby improved phase accuracy. For phase demodulation, a phase estimation module was designed to provide reliable initial phases for the subsequent lightweight neural network, resulting in the phase-estimation network (PE-Net). In formulating the loss function, the similarity of information in both spatial and frequency domains was considered for effectively improving the accuracy and speed of the phase prediction. For phase unwrapping, theoretical analysis was employed for the MHPU method from the perspective of depth constraint, resulting in a more rational selection of frequencies. Experimental results validate that the proposed method, by jointly improving the three stages mentioned above, can achieve real-time 3D imaging with an error of less than 0.031 mm at a resolution of over a million points. Moreover, the imaging speed exceeds 100 frame/s, which provides an efficient lightweight solution for deep-learning-assisted 3D shape measurement in dynamic scenes.

2. Principle

2.1. Basic principle of fringe projection profilometry

A typical monocular FPP system consists of three main components: a projector, a camera, and a computational unit. The projection device projects one or more encoded fringe patterns onto the target object. The imaging device captures the deformed fringes reflected from the object's surface. Finally, the computational unit decodes the deformed fringes, providing the 3D shape information of the tested object. In N-step PSP, the captured fringe patterns can be represented as

$$I_n = A + B\cos(\phi - \delta_n), \qquad \delta_n = 2\pi(n-1)/N, \quad n = 1, 2, \ldots, N. \tag{1}$$

Here, (x, y) donates the pixel coordinate, A is background intensity, B is intensity modulation, and ϕ is the desired phase map. According to the least squares algorithm, the phase map ϕ and intensity modulation can be calculated as

$$\phi = \arctan\frac{M}{D} = \arctan\frac{\sum_{n=1}^{N} I_n \sin\delta_n}{\sum_{n=1}^{N} I_n \cos\delta_n}, \tag{2}$$

$$B = \frac{2}{N}\sqrt{M^2 + D^2}. \tag{3}$$

Here, $M(x, y)$ and $D(x, y)$ represent the numerator and denominator of the arctangent function, respectively. Due to the nature of the distribution of values computed by the arctangent function, the phase obtained will be truncated within $(-\pi, \pi]$ and needs to be unwrapped. By employing an appropriate phase unwrapping method, the wrapped phase ϕ can be mapped to its unwrapped counterpart:

$$\Phi = \phi + 2\pi k. \tag{4}$$

Here, k is the integer fringe order. Finally, after obtaining the unwrapped phase containing object information, further conversion of the unwrapped phase from pixel coordinates to real-world 3D coordinates can be achieved through calibration models such as the phase-height mapping model[33,34] and triangular stereo model[35,36]. This completes the overall process of 3D imaging. The triangular stereo model was employed in the experimental section.

Traditional FPP techniques are highly effective for measuring static scenes, but they often encounter limitations in achieving high-precision and real-time processing in dynamic scenes. Therefore, many recent works have attempted to integrate deep learning into FPP to overcome the efficiency bottlenecks in measurements.

2.2. Proposed real-time 3D imaging based on ROI projection and phase estimation

2.2.1. Framework of the proposed real-time 3D imaging method

Currently, most deep-learning-based 3D imaging methods adopt end-to-end CNNs, leveraging a large dataset to establish an accurate mapping relationship between input fringes and ground truth. UNet[37] and its derivatives are the most popular networks used for fringe analysis. While such methods can achieve high precision in phase retrieval, the introduction of numerous convolutional layers or complex modules makes it challenging to meet the real-time measurement demands of dynamic scenes. Although some methods have proposed lightweight solutions to solve this problem, they often sacrifice accuracy or flexibility. Therefore, achieving improved inference speed without compromising the phase precision and method flexibility remains a pressing issue.

To address this issue, a real-time 3D imaging method based on ROI projection and phase estimation is proposed, as illustrated in Fig. 1. First, the ROI fringe projection strategy is employed to precisely target the area of interest on the tested object, narrowing down the ROI for fringe projection and imaging, thus reducing the fringe period and enhancing phase accuracy. Subsequently, a lightweight network architecture based on phase estimation is designed for real-time and high-precision phase demodulation. Finally, an improved dual-frequency phase unwrapping method is utilized to achieve robust and efficient single-frame 3D imaging. Since the *a priori* knowledge about phase unwrapping is required in ROI projection, the details about each module in the workflow will be introduced, respectively, in the order of phase demodulation, phase unwrapping, and fringe projection within Secs. 2.2.2 to 2.2.4.

2.2.2. Lightweight phase-estimation neural network

The proposed PE-Net, as shown in Fig. 2(a) and inspired by the work of Yin *et al.*[32], utilizes a phase estimation module to pre-estimate initial phase information from a single-frame fringe pattern. Subsequently, this initial phase and the input fringe pattern are jointly used as inputs to the lightweight network, yielding the output phase information. Two learnable filters are constructed for estimating the initial phase, each aimed at

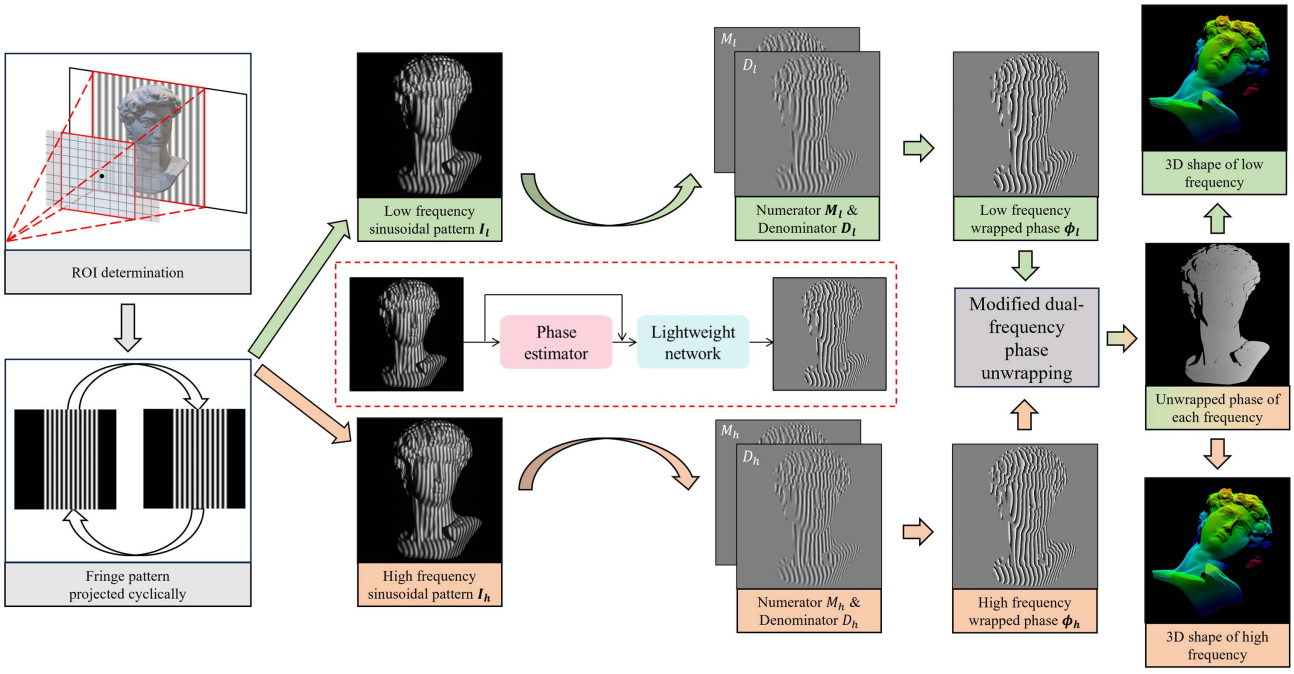

Fig. 1 Flowchart of real-time 3D imaging based on ROI projection and phase estimation. The green and orange arrows indicate the processing flow for low- and high-frequency fringe patterns, respectively.

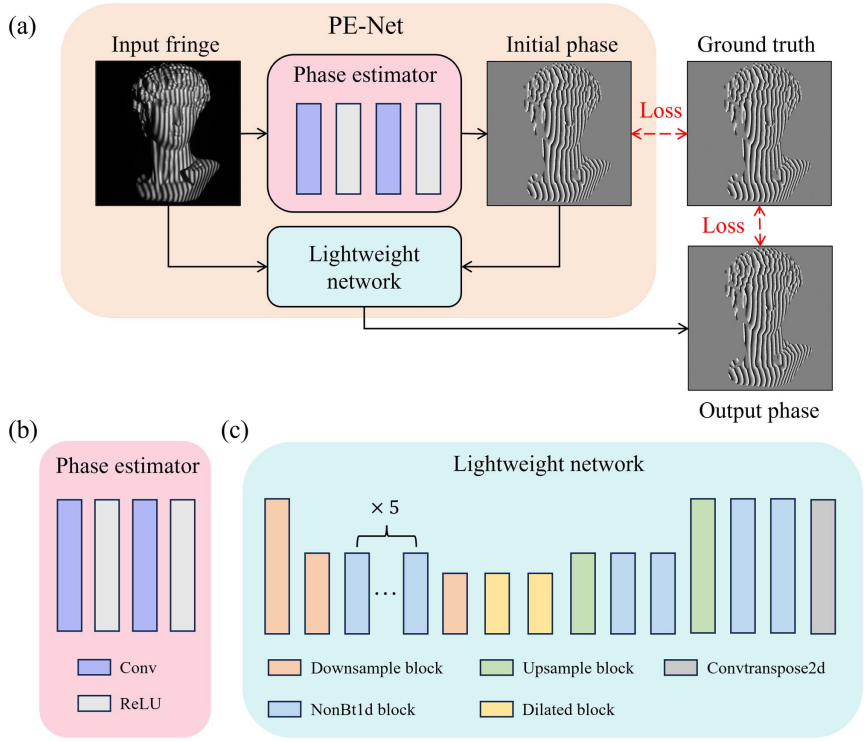

Fig. 2 Real-time phase estimation network structure. (a) Outer structure of PE-Net, loss function design, and training process; (b) internal structure of the phase estimation module; (c) internal structure of the lightweight network.

removing zero frequency and filtering out the fundamental frequency in the FTP method. However, the sizes and center points of these learnable filters still need to be empirically chosen and cannot be flexibly applied to different numbers of fringe periods. Furthermore, existing deep learning frameworks still lack complete support for 2D fast Fourier transform operations. For example, in PyTorch[38], FFT cannot be converted to open neural network exchange (ONNX), and its absence in the model may necessitate custom implementation or integration with specialized libraries for FFT operations, which could impact overall speed depending on the implementation efficiency. Therefore, learnable convolution operations in the spatial domain are employed to replace frequency-domain filtering operations, as shown in Fig. 2(b).

The phase estimation module utilizes two convolutional layers combined with ReLU[39] activation functions as the main components. It extracts the fundamental frequency phase information from the input image through spatial convolution. The first convolutional layer has 64 channels, responsible for converting the fringe pattern into multidimensional features, while the second one has 2 channels, responsible for learning the arctan numerator and denominator terms used to compute the initial phase from the multidimensional features. This module is relatively simple and does not require prior knowledge about FTP, making it flexible for various fringe analysis tasks.

The initial phase information predicted by the phase estimation module is concatenated with the input fringe pattern along the channel dimension and input into the lightweight network to obtain the final phase. The specific design of the lightweight network structure and the internal structures of each module are illustrated in Figs. 2(b), 2(c), and 3, respectively.

The overall structure adopts an encoder-decoder architecture, incorporating modules from ERFNet[40] to achieve a balance between high-quality phase and low computational requirements. The encoder is responsible for capturing hierarchical features from the input fringe images. It reduces the spatial resolution of the feature maps and increases the number of channels through three downsampling steps. The initial number of channels is 64, and the number of channels doubles with each subsequent downsampling step.

The downsampling module adopts the design of ENet[41], which combines 2×2 max pooling and 3×3 convolution with a stride of 2. This combination extracts richer fringe feature information compared to a single convolution operation. The features obtained from downsampling are further processed by the non-bottleneck (NonBt1d) module. This module decomposes the standard convolution in the residual module into two convolutions of sizes 3×1 and 1×3, ensuring approximate accuracy while reducing the required parameters, which is also known as depthwise separable convolution. The last downsampling module utilizes NonBt1d modules with different dilation rates for dilated convolutions to achieve a larger receptive field, which is advantageous for handling high-resolution fringe patterns.

In the decoder part, a symmetrical structure to the encoder is adopted. The upsampling module in the decoder uses 3×3 transposed convolution with a stride of 2 to adjust the spatial resolution of the feature maps. After three upsampling steps, the output is used to predict the numerator and denominator terms for phase computation, consistent with the M and D in Eq. (2).

The training process of PE-Net is illustrated in Fig. 2(a). The loss function evaluates the network's output from both spatial

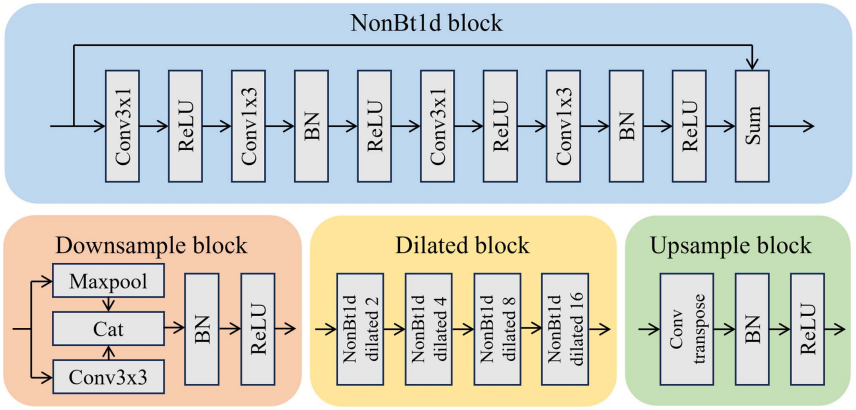

Fig. 3 NonBt1d module, downsampling module, dilated convolution module, and upsampling module within the lightweight network.

and frequency domains. The spatial loss function measures the mean squared error (MSE) among the initial phase information $\hat{y}_{\text{init}}$ obtained by the phase estimation module, the phase $\hat{y}$ obtained by the lightweight network, and the ground truth phase y:

$$L_{\text{spatial}} = \alpha_1[\hat{y}(\mathbf{w}) - y]^2 + \alpha_2[N\hat{y}_{\text{init}}(\mathbf{w}) - y]^2, \tag{5}$$

where $\mathbf{w}$ represents the neural network parameters; the tensor sizes of $\hat{y}_{\text{init}}$, $\hat{y}$, and y are $H \times W \times 2$, where the first channel contains the numerator terms required for phase computation; and the second channel contains the corresponding denominator terms. The frequency domain loss function is given by

$$L_{\text{Fourier-mse}} = \beta_1[\hat{G}(\mathbf{w}) - G]^2 + \beta_2[\hat{G}_{\text{init}}(\mathbf{w}) - G]^2, \tag{6}$$

$$L_{\text{Fourier-ssi}} = \gamma_1 \frac{\sum|\hat{G}(\mathbf{w}) - G|}{\sum|\hat{G}(\mathbf{w}) + G|} + \gamma_2 \frac{\sum|\hat{G}_{\text{init}}(\mathbf{w}) - G|}{\sum|\hat{G}_{\text{init}}(\mathbf{w}) + G|}, \tag{7}$$

where $\hat{G}_{\text{init}}$, $\hat{G}$, and G are the Fourier transforms of $\hat{y}_{\text{init}}$, $\hat{y}$, and y, respectively. $L_{\text{Fourier-mse}}$ measures the pixel-wise difference between the network output and the ground truth from the perspective of spectral amplitude, while $L_{\text{Fourier-ssi}}$ evaluates the structural difference between the network output and the ground truth from the perspective of spectral similarity. The parameters α_1, α_2, β_1, β_2, γ_1, and γ_2 are hyperparameters that adjust the importance of spatial and frequency domain components. The total loss function is expressed as

$$L = L_{\text{spatial}} + L_{\text{Fourier-mse}} + L_{\text{Fourier-ssi}}. \tag{8}$$

The proposed PE-Net includes an additional phase estimation module, which provides more effective input features for the lightweight network. In terms of network structure, spatial separable convolutions are utilized to reduce computational complexity, and multiple layers of dilated convolution NonBt1d modules are stacked to increase the receptive field, facilitating the handling of high-resolution fringe patterns.

2.2.3. *Theoretical analysis of the modified heterodyne phase unwrapping method*

The aforementioned PE-Net achieves high-precision phase retrieval from fringe patterns captured by ROI projection. The MHPU, as reported in our previous work[31], is adopted to unwrap the phase for a wider selection range of the frequencies, as shown in Fig. 4. The unwrapping process begins with taking the phase differences between the phases of two frequencies:

$$\phi_{\text{eq}} = \phi_h - \phi_l, \tag{9}$$

where ϕ_l, ϕ_h, and ϕ_{eq} represent the low frequency, high frequency, and equivalent phase, respectively. The corresponding beat frequency $f_{\text{eq}} = f_h f_l/(f_h - f_l)$. Different from the traditional dual-frequency heterodyne phase unwrapping method, the MHPU method relaxes the limitation on the number of fringe periods, allowing the beat frequency to be greater than 1. This means that higher numbers of fringe periods can be used to modulate the information of the object, thereby improving phase accuracy. Then an unwrapped phase ϕ_r of the reference plane is employed to eliminate the ambiguity of ϕ_{eq}:

$$\Phi_{\text{eq}} = \phi_{\text{eq}} + 2\pi \cdot \text{Round}\left[\frac{\Phi_r - \phi_{\text{eq}}}{2\pi}\right]. \tag{10}$$

Furthermore, the wrapped phases of low frequency and high frequency can be unwrapped by

$$\begin{cases} \Phi_l = \phi_l + 2\pi \cdot \text{Round}\left[\frac{(f_l/f_{\text{eq}})\Phi_{\text{eq}} - \phi_l}{2\pi}\right] \\ \Phi_h = \phi_h + 2\pi \cdot \text{Round}\left[\frac{(f_h/f_{\text{eq}})\Phi_{\text{eq}} - \phi_h}{2\pi}\right] \end{cases}. \tag{11}$$

Utilizing the pre-computed phase of the reference plane can significantly expand the selectable range of frequency for heterodyning, enhancing accuracy and aiding the network's feature extraction from input fringes. However, the introduction of a reference plane may lead to a reduction in the measurement range. Therefore, theoretical analysis of MHPU is necessary to ensure a reasonable measurement range while avoiding errors in phase unwrapping.

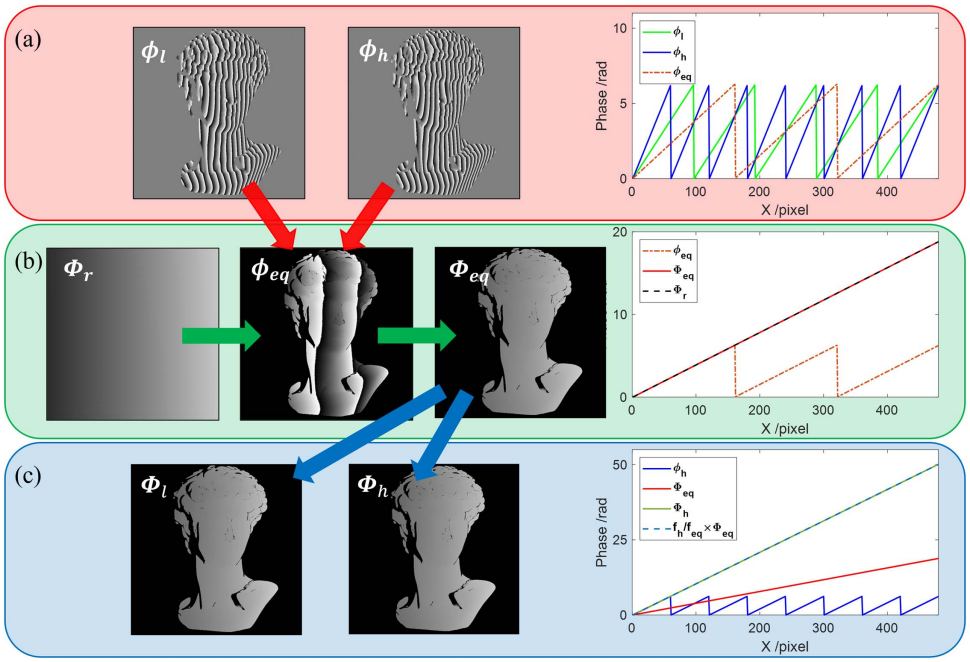

Fig. 4 Flowchart of the MHPU method. (a) Equivalent phase with the phase difference of low- and high-frequency fringe; (b) unwrapped equivalent phase assisted by a reference plane; (c) unwrapped phases of low frequency and high frequency.

The limitations of MHPU are similar to those of absolute phase unwrapping methods based on geometric constraints[42], as shown in Fig. 5. The angle between the projector and camera optical axes is θ, and the physical size of the fringe period is Δx. Points P and Q represent the intersection of the same ray projected by the projector with planes at two different depths, $z_{\min}$ and $z_{\max}$. The maximum depth range of the measurement is $[z_{\min}, z_{\max}]$, and $\Delta z_{\max} = z_{\max} - z_{\min}$. To avoid errors in phase unwrapping, the maximum depth range $\Delta z_{\max}$ needs to satisfy

$$\Delta z_{\max} = \frac{\Delta x}{\tan\theta}. \tag{12}$$

Assume the projector's field of view (FOV) is W_{FOV} (mm). When the camera can capture the entire projected fringe pattern, and the depth range of the object under test is Δz_{obj}. Then the following inequality holds:

$$\frac{W_{\mathrm{FOV}}}{\tan\theta\sqrt{f_h}} \geq \Delta z_{\mathrm{obj}}. \tag{13}$$

The number of fringe periods can be obtained by solving the above inequality:

$$f_h \leq \left(\frac{W_{\mathrm{FOV}}}{\Delta z_{\mathrm{obj}}\tan\theta}\right)^2. \tag{14}$$

In conclusion, the selection of the low-frequency f_l and high-frequency f_h in the improved dual-frequency phase unwrapping strategy depends on the robustness requirements of the phase unwrapping and the desired range of measurement depth. To ensure good noise immunity in phase unwrapping, the low and high frequencies must satisfy $f_l = f_h - \sqrt{f_h}$. Meanwhile, to ensure the maximum measurable depth range, the high-frequency f_h needs to satisfy Eq. (14).

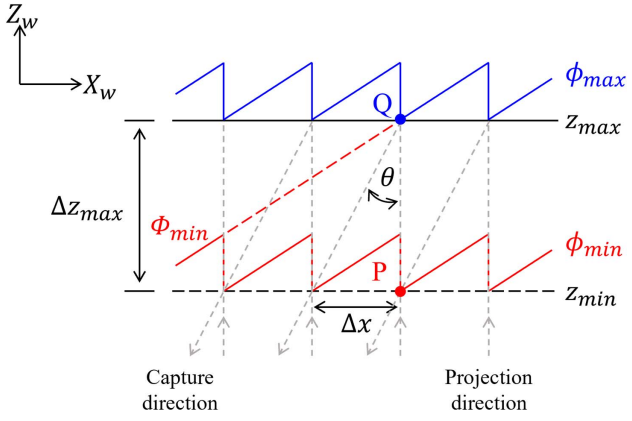

Fig. 5 Depth limitation of MHPU.

2.2.4. ROI fringe projection strategy

For a fixed FPP system, reducing camera resolution to decrease the number of pixels is a common approach to meet the real-time demands of dynamic scenes. However, this approach will result in the camera's FOV being smaller than that of the projector, causing a reduction in the effective number of fringe periods from the camera's perspective and subsequently decreasing phase accuracy. Additionally, the size of the target object is usually smaller than the projector's FOV. If the generated fringes are projected according to the full resolution of the projector, it will lead to a limited number of fringe periods effectively modulated by the object within the FOV. To address these two

problems, a fringe projection strategy based on the ROI is proposed.

Equation (14) can be rewritten as

$$p_h \geq W_p \left(\frac{\Delta z_{\text{obj}} \tan \theta}{W_{\text{FOV}}} \right)^2, \tag{15}$$

where p_h represents the width of the high-frequency fringe period and W_p represents the width of the generated fringe (pixels). It is evident that appropriately reducing W_p while ensuring that the projected fringe can cover the ROI can effectively decrease the lower bound of p_h, which can increase the corresponding number of fringe periods.

The normal fringe projection strategy, as shown in Fig. 6(a), utilizes the entire resolution $W_p \times H_p$ pixels of the projector to project fringes. For areas outside of the tested objects in the measurement scene, where the fringe does not contain valid information, the corresponding fringe pattern pixels under the projector's perspective are not involved in calculations. Therefore, during fringe generation, actively constraining and reducing the area can maintain the same number of fringe periods while decreasing the fringe period, thereby enhancing measurement accuracy.

As shown in Fig. 6(b), the resolution of the ROI fringe pattern $W'_p \times H_p$ is determined based on the actual occupied area of the object in the measurement FOV. The origin of the ROI has shifted by ΔW relative to the original image in the x-axis direction. The phase distribution of normal and ROI fringe projection can be expressed as

$$\begin{cases} \Phi = \frac{2\pi f x}{W_p} \\ \Phi' = \frac{2\pi f}{W'_p}(x - \Delta W'_p) \end{cases}. \tag{16}$$

The phase relationship before and after changing the projection strategy is

$$\Phi = \frac{W'_p}{W_p}\Phi' + \frac{2\pi f \Delta W'_p}{W_p}. \tag{17}$$

In the triangular stereo model, the calculation of pixel coordinates is based on the origin of the phase in the normal projection strategy. Therefore, when using the ROI projection strategy, the obtained phase needs to be transformed according to Eq. (17) to obtain the correct pixel coordinates, ensuring the phase starts at the appropriate zero point.

In conclusion, the ROI fringe projection makes it more effective to utilize the measurement system's FOV. In practical applications, only the approximate FOV of the measurement scene and the angle between the projector and camera optical axes are required. This allows for increasing the fringe frequency while keeping the number of fringe periods constant, thereby enhancing the precision of 3D measurements.

2.3. *Performance analysis of the proposed real-time 3D imaging method*

The proposed real-time 3D imaging method introduces several innovations. In the aspect of phase demodulation, a neural network-driven approach is adopted, using more flexible, learnable spatial convolution layers to replace manually designed operations in the spectral domain. This provides reliable initial phase estimation for the subsequent lightweight network, allowing the network to learn accurate mapping relationships with fewer parameters without sacrificing phase accuracy. In the loss function of PE-Net, the network training is constrained in both spatial and frequency domains with the consideration of pixel-level differences and structural similarity differences, resulting in more accurate output phase information.

In the aspect of phase unwrapping, the conditions required for the MHPU method to successfully unwrap the phase are analyzed from the perspective of depth constraint. The selectable range of low frequency and high frequency is determined to ensure that the correct 3D shape can be obtained while the highest frequency can be used to modulate the object information.

In the aspect of fringe projection, based on the conclusions drawn from analyzing the MHPU method, the region of actually projected fringes is reduced to within the ROI, decreasing the number of fringe periods without losing valid information. Additionally, the phase relationship before and after using ROI fringe projection is derived to provide the correct phase when performing 3D reconstruction with the fixed calibration parameters.

Through the analysis and improvements in ROI fringe projection, PE-Net phase demodulation, and MHPU phase unwrapping, the entire 3D imaging process gains the capability to efficiently process fringe images while maintaining high precision and flexibility.

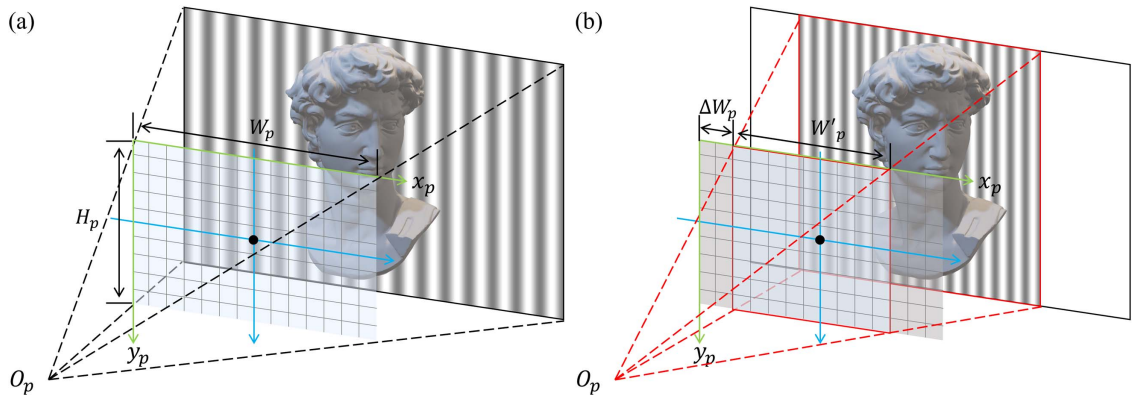

Fig. 6 Two fringe projection strategies. (a) Normal fringe projection. (b) ROI fringe projection.

3. Experiment

3.1. ROI determination and dataset establishment

The experimental setup of the FPP system consists of a camera (Baumer, VCXU-31M) with a resolution of 1280 pixel × 800 pixel and a projector (TI, DLP4500) with a resolution of 912 pixel × 1140 pixel. The focal length of the camera lens is 12 mm. Both the camera and projector have an exposure time set to 10 ms. The system's FOV is approximately 300 mm × 150 mm, with a working distance of around 600 mm.

During the measurement, the projector's FOV exceeds that of the camera, allowing for further reduction in the fringe period using ROI fringe projection. First, a white pattern was projected onto the surface of the target object. Next, the valid area of the object within the image captured by the camera can be determined. Finally, the resolution of the ROI under the projector's perspective was calculated. An additional redundancy is added to the maximum size of the object area to calculate the ROI. The actual required fringe width W'_p for projection is 500 pixel, with an offset of 256 pixel in the x-axis direction.

After determining the ROI for projecting fringes, it is necessary to calculate the suitable number of fringe periods f_l and f_h required in MHPU. The projected fringes occupy the entire FOV of the camera, with the measured W_{FOV} approximately 300 mm. The depth range Δz_{obj} of the object under test was approximately 120 mm. The angle θ between the optical axes of the camera and projector was obtained through calibration as 16.32°. Substituting $W_{\mathrm{FOV}} = 300$, $\Delta z_{\mathrm{obj}} = 120$, and $\theta = 16.32°$ into Eq. (14) yields

$$f_h \le \left(\frac{300}{120 \times \tan 16.32}\right)^2 \approx 72.9. \tag{18}$$

The maximum selectable fringe period number is 72.9. Therefore, f_h was selected as 72. The maximum measurement depth range can be driven by Eq. (14), $\Delta z_{\max} = 120.7$ mm, which is quite reasonable in many application scenarios. Consequently, f_l was selected as $f_l = 72 - \sqrt{72} \approx 64$ to ensure better noise resistance, which can be derived from the traditional three-frequency phase unwrapping method.

12-step phase-shifting fringe patterns with $f_l = 64$ and $f_h = 72$ were projected to calculate the ground truth of the numerator and denominator terms in the phase demodulation task. A total of 100 sets of data have been collected under different scenes and divided into training and validation sets at a ratio of 9:1. A mask was obtained by setting the average intensity threshold to 10, which is then applied to the fringe pixels to remove unreliable regions with low-intensity modulation. To further improve the model's generalization and robustness, crop transformation was adopted to input images during the training stage. Specifically, a random region with a width of 64 to 256 pixel was selected from the image, and the pixel values within that region were set to zero.

3.2. Quantitative evaluation of ROI fringe projection

In this section, we compared the accuracy variation between the normal and the ROI projection strategy, using the three-step phase shift method to calculate the wrapped phase and MHPU for phase unwrapping.

Figure 7 shows a comparison between two projection strategies, using a number of fringe periods of 72 for both strategies. It can be seen that, after determining the appropriate ROI, the fringe period width is significantly reduced, indicating that more phase values are used for modulation of the target object in the ROI, resulting in higher measurement accuracy.

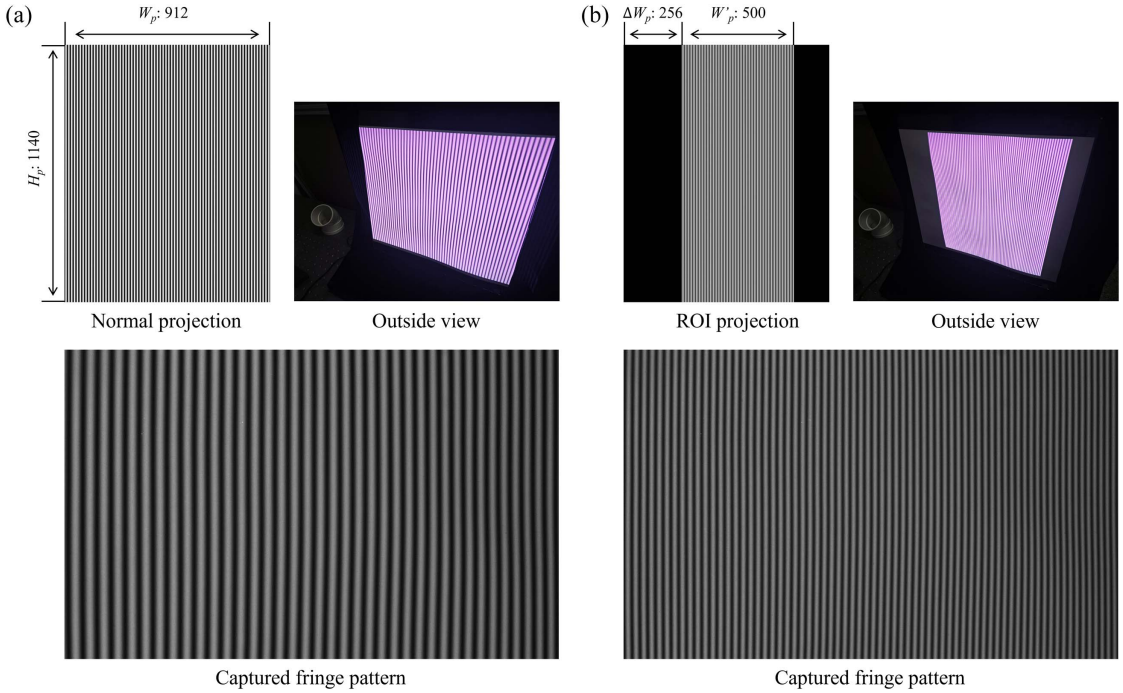

Fig. 7 Comparison of two fringe projection strategies. (a) Generated fringe pattern (top left) and perspective view (top right) of normal fringe projection. (b) Corresponding results of ROI fringe projection.

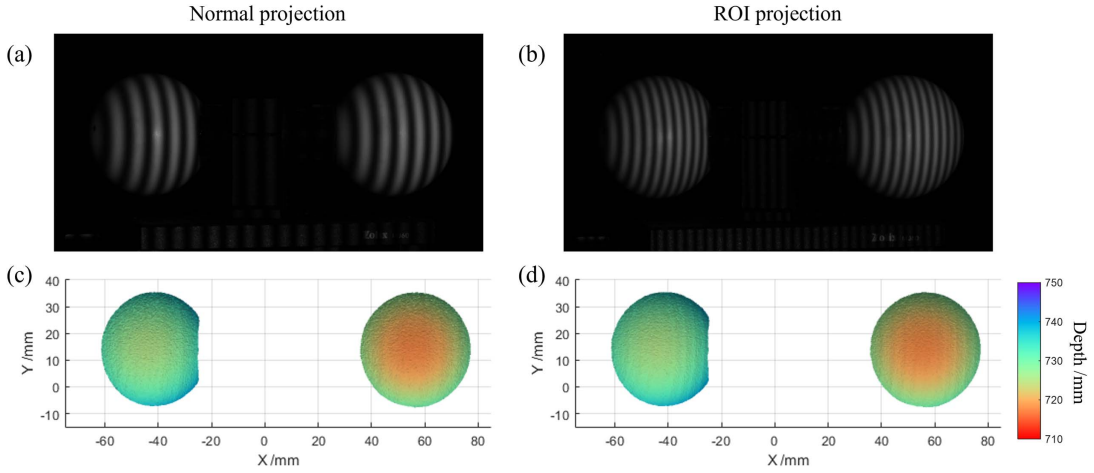

Fig. 8 Measurement results of standard spheres using three-step PSP and two projection strategies. (a), (b) Fringe patterns of two projection strategies. (c),(d) 3D reconstruction results of two projection strategies.

Two standard spheres were measured to quantitatively evaluate the accuracy of the ROI fringe projection strategy. The diameters of the two standard spheres are $D_A = 50.7991$ mm (left sphere A) and $D_B = 50.7970$ mm (right sphere B), with a distance between their centers D_C of 100.2537 mm. 3D reconstruction results of the two strategies are shown in Fig. 8, and the corresponding values are given in Table 1. After using ROI projection, the fringe period width on the surface of the standard sphere was significantly reduced, and the MAE_A and RMS_A errors decreased by 0.0225 and 0.0137 rad, respectively. The MAE_B and RMS_B decreased by 0.028 and 0.0243 rad, respectively, proving the effectiveness of the proposed ROI projection strategy.

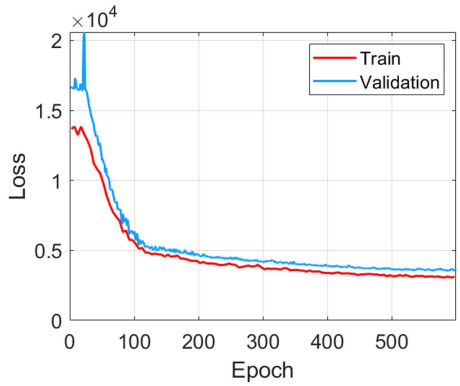

Fig. 9 Train and validation loss curves of PE-Net.

3.3. Training and validation of PE-Net

PE-Net was implemented by PyTorch 1.13.1. The training and validation process was completed in a hardware environment with an Intel Xeon Gold 622 R 2.90 GHz CPU and one 40 GB NVIDIA A40 GPU. The inference time measurement was conducted with FP16 precision on a laptop with an Intel i9-13800HX 2.20 GHz and 8 GB NVIDIA GeForce RTX 4070 Laptop GPU through TensorRT. An AdamW[43] optimizer with an initial learning rate of 10^{-3} was used to minimize the joint loss function in Eq. (9). During the training stage. The learning rate was adjusted by the cosine annealing strategy within the range of 10^{-3} to 10^{-5}. The hyperparameters of the loss function were set to $\alpha_1 = 1$, $\beta_1 = 0.5$, and $\gamma_1 = \gamma_2 = 1000$. The initial values of α_2 and β_2 were set to 0.1 and gradually decreased using the step decay strategy to 0.02 as the epoch increased. The batch size was set to 4. The train and validation loss curves are shown in Fig. 9.

The phase error of PE-Net relative to the 12-step PSP was given in Table 2, and Feng's method[22], UNet method[23], and NAS method[31] were trained and validated on the same dataset for comparison. The inference time was calculated by averaging the execution time of each algorithm, obtained from running it 1000 times in FP16. It can be seen that the MAE of the proposed method is 0.02979 rad, which is the smallest among the four methods, and the corresponding RMS is 0.05163 rad, only 0.0016 rad larger than that of UNet with the largest parameter number. The mean inference time of PE-Net is 6.01 ms (about 166 frame/s), which is the fastest among the four methods. The inference time of Feng's method, UNet method, and NAS

Table 1 Quantitative Results of Standard Sphere Measurement Using Three-Step PSP and Two Projection Strategies

	D_A	D_B	D_C	MAE_A	MAE_B	RMS_A	RMS_B
Normal	50.8002	50.8158	100.1140	0.0995	0.0855	0.1207	0.1100
ROI	50.8000	50.8062	100.1362	0.0770	0.0718	0.0927	0.0857

Table 2 MAE and RMS of the Phase Error, Inference Time, and Model Size of Four Methods

Method	MAE (rad)	RMS (rad)	Inference time (ms)	Model size (M)
Feng	0.03417	0.05811	67.70	0.70
UNet	0.03103	**0.05007**	41.02	31.00
NAS	0.03728	0.06364	17.02	**0.13**
PE-Net	**0.02979**	0.05163	**6.01**	2.10

method are approximately 11.26 times, 6.103 times, and 2.83 times that of PE-Net, respectively. The experimental results preliminarily demonstrate the superiority of the proposed method in terms of speed and accuracy.

3.4. Quantitative evaluation of the proposed method

To verify the effectiveness of the proposed method in complex scenarios, we first measured a sculpture of David. The 3D reconstruction results obtained from the 12-step PSP (ground truth), Feng's method, UNet method, NAS method, and PE-Net are shown in Fig. 10. Additionally, Figure 11 illustrates

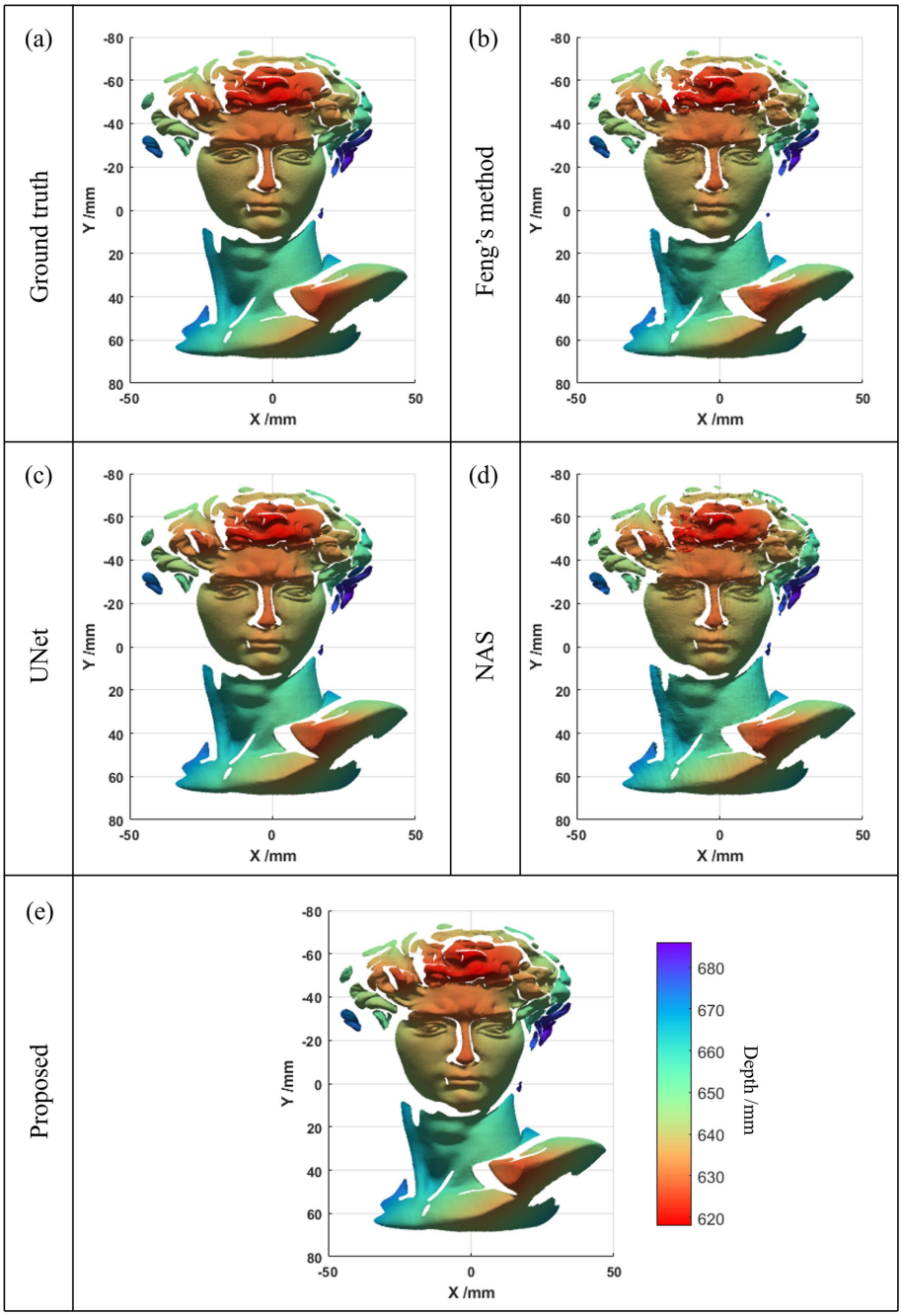

Fig. 10 3D reconstruction results of the David sculpture. (a) Ground truth. (b) Feng's method; (c) UNet method; (d) NAS method; (e) proposed PE-Net.

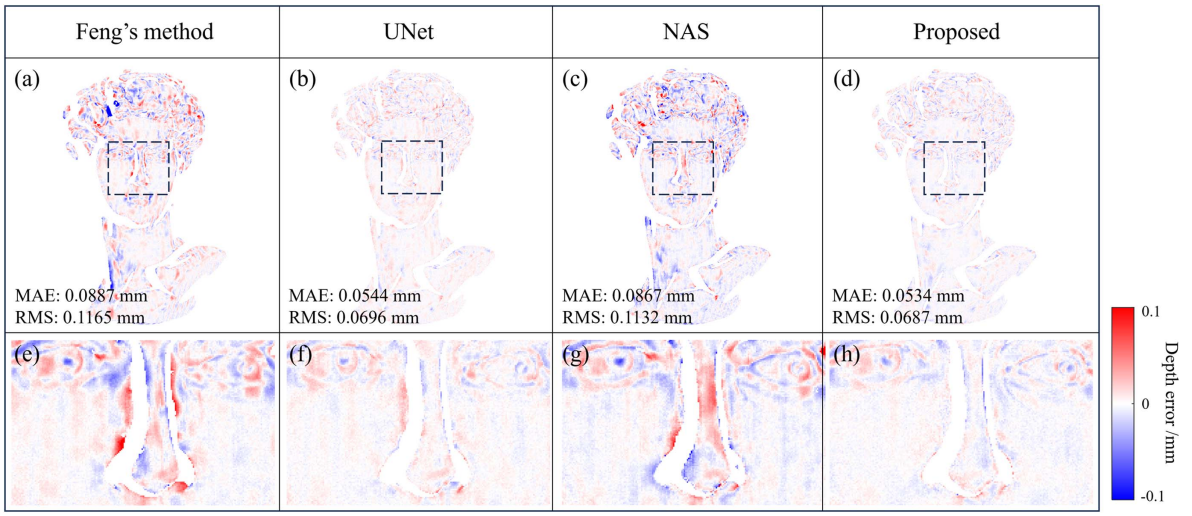

Fig. 11 Error distributions of the David sculpture. (a) Feng's method; (b) UNet method; (c) NAS method; (d) proposed PE-Net. (e)–(h) Localized enlarged images of the dashed-boxed regions corresponding to each method.

the depth error distributions relative to the 12-step PSP, along with the corresponding MAE and RMS values. It can be seen that near the area around David's nose, the errors are more pronounced for Feng's method and the NAS method, whereas UNet and the proposed method exhibit smaller errors. Benefiting from the phase estimation module providing initial phase information to the network, the proposed PE-Net achieves lower MAE and RMS errors compared to the UNet method by 0.001 and 0.0009 mm, respectively, while maintaining relatively fast inference speed and lower computational complexity. The results indicate that the proposed method can obtain accurate 3D reconstruction results when dealing with complex objects.

Two standard spheres were measured to quantitatively evaluate the four methods, with parameters identical to those in Sec. 3.2. The 3D reconstruction results and error values of the four methods are shown in Fig. 12. From the depth error distributions, it can be observed that the proposed method exhibits smaller residual periodic errors compared to the other three methods. This is reflected in the corresponding MAE and RMS errors, all of which are less than 0.031 mm for the proposed method, which is attributed to the PE module proposed in the method, which simulates the FTP method's extraction of coarse phase, allowing the overall network to achieve higher accuracy with fewer parameters and faster inference speed. These results demonstrate that, by combining ROI projection, PE-Net phase demodulation, and MHPU phase unwrapping, high-precision real-time 3D imaging can be achieved through the entire pipeline, balancing the speed, accuracy, and flexibility of the algorithm.

3.5. Real-time 3D imaging of the dynamic scene

To validate the performance of the proposed method in dynamic scene measurement, we developed a real-time 3D imaging system on the laptop with an Intel i9-13800HX 2.20 GHz and 8 GB NVIDIA GeForce RTX 4070 Laptop GPU, which is mentioned in Sec. 3.2. Before the measurement process, it is necessary to generate ROI patterns based on parameters such as FOV and working distance and upload them to the projector firmware. Additionally, the network model and pre-calibrated parameters are stored in the GPU. During the measurement process, the projector triggered the camera to capture images via hardware signals. The captured image stream was transferred sequentially from the CPU to the GPU. Following normalization, PE-Net execution, phase calculation, MHPU phase unwrapping, and 3D reconstruction, point cloud results were obtained. Finally, the results can be visualized through OpenGL in real-time.

As shown in Figs. 13(a) and 13(b), the measurement scene contained a static David and a rotating Nike sculpture. Two fringe patterns with 64 and 72 fringe periods were projected cyclically. Although the inference time of PE-Net is 6.01 ms (166 frame/s), consideration must be given to the extra time consumed by data transmission between the CPU and GPU, point cloud rendering time, and GPU synchronization waiting time. Therefore, the projection period of the projector was set to 10 ms (100 frame/s) to prevent potential data transmission bottlenecks and thread blocking, thereby ensuring real-time data processing efficiency. Since the entire 3D imaging can be completed within the projection interval, the speed of the entire system is consistent with the projection speed of the projector, both being 100 frame/s. The real-time measurement result is shown in Visualization 1.

Figures 13(c)–13(f) present the point cloud results of different frames during the imaging process. It is evident that intricate details such as the wings and waist of the rotating statue of Nike, as well as complex variations in the position of details like the hair and neck of David, are accurately reconstructed. The experimental results demonstrate that PE-Net can accurately recover the phase information of objects with complex surfaces. Combined with the ROI fringe projection strategy and MHPU, high-accuracy real-time 3D imaging has been achieved with a speed of 100 frame/s and a resolution of 1280×800.

4. Conclusion and Discussion

In this paper, the spatial and spectral information of the phase is utilized to provide a reasonable initial phase for the neural network. The lightweight network, combined with the phase

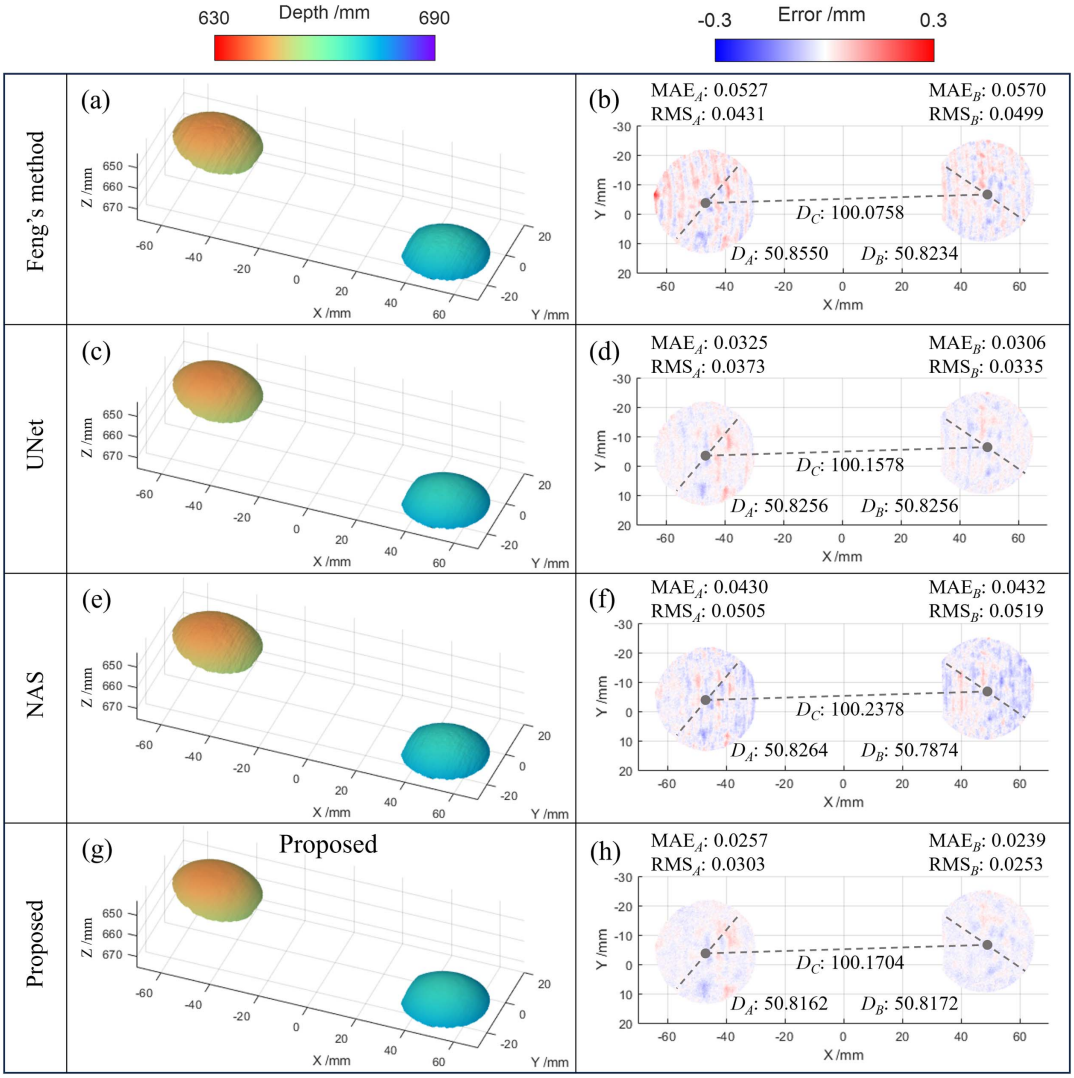

Fig. 12 3D reconstruction results and error values of the standard spheres. (a) Feng's method; (c) UNet method; (e) NAS method; (g) proposed PE-Net. (b), (d), (f), (h) Corresponding depth error distributions and error values.

estimation module, can predict high-precision phase results in real time. Additionally, theoretical analysis is conducted on the optimal selection of fringe cycle numbers in the MHPU method considering the FOV, measurement depth, and projector resolution. Based on this analysis, adjustments are made to the ROI of the fringe patterns from the perspective of the projector, further enhancing the phase accuracy. Our main contributions can be summarized in the following three points:

1. **Efficient phase demodulation with PE-Net**. The phase estimation module is designed to provide a reliable initial phase, enabling the lightweight neural network to achieve faster inference speeds without sacrificing the final accuracy. In the design of the loss function, both spatial and spectral information similarity are simultaneously considered.
2. **ROI fringe projection strategy**. To address mismatches among the projector's FOV, camera's FOV, and target object size, appropriate ROI and resolution for actually projected fringes are carefully selected based on system parameters and theoretical analysis, which enables the reduction of fringe period width and thus improvement of the phase accuracy.
3. **Real-time 3D imaging that achieves a balance among speed, accuracy, and flexibility**. Optimization is conducted at different stages of the 3D imaging process, including ROI fringe projection strategy, PE-Net phase demodulation, and MHPU phase unwrapping, which allows for one 3D reconstruction result for each newly captured image. The entire pipeline achieves real-time 3D imaging with an RMS error of less than 0.031 mm, a resolution of 1280×800, and a speed over 100 frame/s, providing a more efficient lightweight solution for dynamic scene measurement.

Several aspects require further improvement in future investigations. First, with the increase in imaging speed, the data transmission bottleneck becomes the primary constraint on the algorithm performance. The transmission time of images between CPU and GPU already accounts for approximately 30% (1.79/6.01) of the entire algorithmic process time. Hence, exploring methods to reduce the data transmission time is crucial. This could involve enhancing data compression algorithms or computing pixel values solely for regions experiencing scene changes, thereby refining the algorithm's performance.

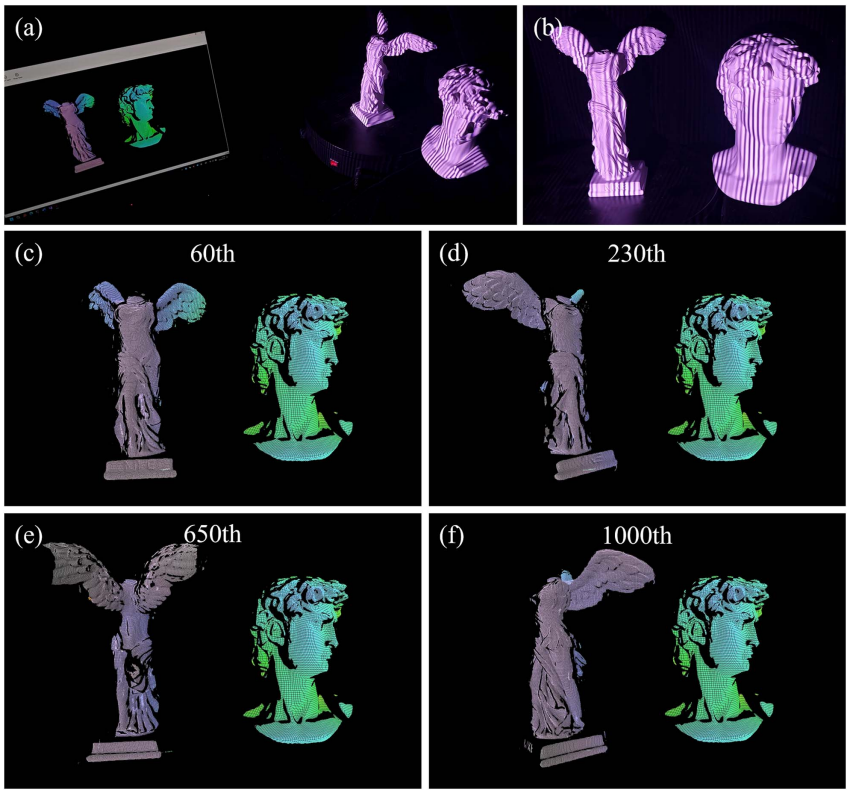

Fig. 13 Real-time 3D reconstruction results (Visualization 1). (a), (b) Measurement scene comprising a stationary David and a rotating Nike sculpture. (c)–(f) 3D point cloud of different frames.

Second, when dealing with objects with large-scale motion, it is challenging to flexibly determine suitable ROI projection strategies based on the current object pose in real-time. This limitation hampers the effective enhancement of measurement accuracy. Utilizing common techniques in deep learning such as object detection and semantic segmentation, it is possible to predict the spatial positions of the objects of interest in the current frame based on the captured fringe sequence. Continuously adjusting the ROI of fringe projection accordingly represents a promising avenue for future research in achieving high-precision real-time 3D imaging.

Lastly, recollecting data is necessary if the system parameters change. The sensitivity of the model to changes in specific system parameters (e.g., fringe frequency selection based on optical setup) plays a crucial role in the performance of the proposed method. Techniques like fine-tuning or transfer learning can enhance the model's adaptability. If system parameters change within certain bounds or patterns, fine-tuning the pre-trained model on new data reflecting these changes can update the model without starting from scratch. This approach can reduce the need for extensive data recollection.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62075143 and 62205226). The authors declare no conflicts of interest.

References

1. S. S. Gorthi and P. Rastogi, "Fringe projection techniques: whither we are?" *Opt. Lasers Eng.* **48**, 133 (2010).
2. S. Van der Jeught and J. J. J. Dirckx, "Real-time structured light profilometry: a review," *Opt. Lasers Eng.* **87**, 18 (2016).
3. Z. Wu *et al.*, "High-speed and high-efficiency three-dimensional shape measurement based on Gray-coded light," *Photonics Res.* **8**, 819 (2020).
4. Z. Wu *et al.*, "Dynamic 3D shape reconstruction under complex reflection and transmission conditions using multi-scale parallel single-pixel imaging," *Light Adv. Manuf.* **5**, 34 (2024).
5. J. Xu and S. Zhang, "Status, challenges, and future perspectives of fringe projection profilometry," *Opt. Lasers Eng.* **135**, 106193 (2020).
6. K. Harding, "Engineering precision," *Nat. Photonics* **2**, 667 (2008).
7. J. Geng, "Structured-light 3D surface imaging: a tutorial," *Adv. Opt. Photonics* **3**, 128 (2011).
8. P. Wissmann, R. Schmitt, and F. Forster, "Fast and accurate 3D scanning using coded phase shifting and high speed pattern projection," in *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission* (2011), p. 108.
9. Y. Zhang, Z. Xiong, and F. Wu, "Unambiguous 3D measurement from speckle-embedded fringe," *Appl. Opt.* **52**, 7797 (2013).
10. S. Feng, Q. Chen, and C. Zuo, "Graphics processing unit-assisted real-time three-dimensional measurement using speckle-embedded fringe," *Appl. Opt.* **54**, 6865 (2015).
11. T. Tao *et al.*, "Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system," *Opt. Express* **24**, 20253 (2016).
12. Y. Wang *et al.*, "Period coded phase shifting strategy for real-time 3-D structured light illumination," *IEEE Trans. Image Process.* **20**, 3001 (2011).
13. Z. Wu *et al.*, "High-speed three-dimensional shape measurement based on cyclic complementary Gray-code light," *Opt. Express* **27**, 1283 (2019).

14. Z. Wu, W. Guo, and Q. Zhang, "High-speed three-dimensional shape measurement based on shifting Gray-code light," *Opt. Express* **27**, 22631 (2019).
15. C. Zuo *et al.*, "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," *Opt. Lasers Eng.* **51**, 953 (2013).
16. C. Zuo *et al.*, "High-speed three-dimensional profilometry for multiple objects with complex shapes," *Opt. Express* **20**, 19493 (2012).
17. M. Takeda *et al.*, "Frequency-multiplex Fourier-transform profilometry: a single-shot three-dimensional shape measurement of objects with large height discontinuities and/or surface isolations," *Appl. Opt.* **36**, 5347 (1997).
18. H.-M. Yue, X.-Y. Su, and Y.-Z. Liu, "Fourier transform profilometry based on composite structured light pattern," *Opt. Laser Technol.* **39**, 1170 (2007).
19. H. O. Saldner and J. M. Huntley, "Temporal phase unwrapping: application to surface profiling of discontinuous objects," *Appl. Opt.* **36**, 2770 (1997).
20. C. Zuo *et al.*, "Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review," *Opt. Lasers Eng.* **85**, 84 (2016).
21. H. Yu *et al.*, "Deep learning-based fringe modulation-enhancing method for accurate fringe projection profilometry," *Opt. Express* **28**, 21692 (2020).
22. S. Feng *et al.*, "Fringe pattern analysis using deep learning," *Adv. Photonics* **1**, 1 (2019).
23. S. Feng *et al.*, "Generalized framework for non-sinusoidal fringe analysis using deep learning," *Photonics Res.* **9**, 1084 (2021).
24. L. Zhang *et al.*, "High-speed high dynamic range 3D shape measurement based on deep learning," *Opt. Lasers Eng.* **134**, 106245 (2020).
25. J. Zhang *et al.*, "Single-exposure optical measurement of highly reflective surfaces via deep sinusoidal prior for complex equipment production," *IEEE Trans. Ind. Inf.* **19**, 2039 (2023).
26. S. Van der Jeught and J. J. J. Dirckx, "Deep neural networks for single shot structured light profilometry," *Opt. Express* **27**, 17091 (2019).
27. H. Nguyen and Z. Wang, "Accurate 3D shape reconstruction from single structured-light image via fringe-to-fringe network," *Photonics* **8**, 459 (2021).
28. V. Srinivasan, H. C. Liu, and M. Halioua, "Automated phase-measuring profilometry of 3-D diffuse objects," *Appl. Opt.* **23**, 3105 (1984).
29. C. Zuo *et al.*, "Phase shifting algorithms for fringe projection profilometry: a review," *Opt. Lasers Eng.* **109**, 23 (2018).
30. Y. Li *et al.*, "Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement," *Opt. Express* **30**, 3424 (2022).
31. Y. Li *et al.*, "Real-time 3D shape measurement of dynamic scenes using fringe projection profilometry: lightweight NAS-optimized dual frequency deep learning approach," *Opt. Express* **31**, 40803 (2023).
32. W. Yin *et al.*, "Physics-informed deep learning for fringe pattern analysis," *Opto-Electron. Adv.* **7**, 230034 (2024).
33. W.-S. Zhou and X.-Y. Su, "A direct mapping algorithm for phase-measuring profilometry," *J. Mod. Opt.* **41**, 89 (1994).
34. H. Guo, "Least-squares calibration method for fringe projection profilometry," *Opt. Eng.* **44**, 033603 (2005).
35. B. Li, N. Karpinsky, and S. Zhang, "Novel calibration method for structured-light system with an out-of-focus projector," *Appl. Opt.* **53**, 3415 (2014).
36. Z. Li, "Accurate calibration method for a structured light system," *Opt. Eng.* **47**, 053604 (2008).
37. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, Vol. **9351** (2015), p. 234.
38. A. Paszke *et al.*, "PyTorch: an imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (2019), p. 8026.
39. V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (2010), p. 807.
40. E. Romera *et al.*, "ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transport. Syst.* **19**, 263 (2018).
41. A. Paszke *et al.*, "ENet: a deep neural network architecture for real-time semantic segmentation," arXiv:1606.02147 (2016).
42. Y. An, J.-S. Hyun, and S. Zhang, "Pixel-wise absolute phase unwrapping using geometric constraints of structured light system," *Opt. Express* **24**, 18445 (2016).
43. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv:1711.05101 (2017).