



Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement

YIXUAN LI,^{1,2}  JIAMING QIAN,^{1,2}  SHIJIE FENG,^{1,2,3}  QIAN CHEN,^{2,4}  AND CHAO ZUO^{1,2,*} 

¹Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

²Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China

³shijiefeng@njust.edu.cn

⁴chenqian@njust.edu.cn

*zuochao@njust.edu.cn

Abstract: Single-shot fringe projection profilometry (FPP) is essential for retrieving the absolute depth information of the objects in high-speed dynamic scenes. High-precision 3D reconstruction using only one single pattern has become the ultimate goal in FPP. The frequency-multiplexing (FM) method is a promising strategy for realizing single-shot absolute 3D measurement by compounding multi-frequency fringe information for phase unwrapping. In order to solve the problem of serious spectrum aliasing caused by multiplexing schemes that cannot be removed by traditional spectrum analysis algorithms, we apply deep learning to frequency multiplexing composite fringe projection and propose a composite fringe projection deep learning profilometry (CDLP). By combining physical model and data-driven approaches, we demonstrate that the model generated by training an improved deep convolutional neural network can directly perform high-precision and unambiguous phase retrieval on a single-shot spatial frequency multiplexing composite fringe image. Experiments on both static and dynamic scenes demonstrate that our method can retrieve robust and unambiguous phases information while avoiding spectrum aliasing and reconstruct high-quality absolute 3D surfaces of objects only by projecting a single composite fringe image.

© 2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Structured light (SL) projection is one of the most representative 3D optical imaging technologies for macroscopic objects due to its non-contact, high-resolution, and easy-to-implement measurement capabilities [1–4]. Among them, fringe projection profilometry (FPP), has become one of the most prevalent SL methods with the advantages of full-field scanning and high-precision measurement [5–7], which has been widely applied in multiple fields, such as intelligent manufacturing [8] and reverse engineering [9]. For FPP, the projector projects a series of fringe patterns onto the target object, and then the camera captures these images modulated and deformed by the objects. The measured objects' absolute phases and related depth information can be retrieved with the captured fringe patterns by processing the following three steps: fringe analysis, phase unwrapping, and phase-to-height mapping. With the rapid development of optoelectronic information technology [10–12], people subsequently set higher expectations on FPP, requiring both higher precision and higher speed. However, these two aspects seem contradictory in nature. Due to the increasing demand for dynamic scene measurement (such as online industrial inspection, stress deformation analysis, fast reverse modeling, etc. [13]), “speed” has gradually become a fundamental factor that must be taken into account when using FPP. There are two factors

determining the 3D measurement speed of FPP: (1) the speed of hardware devices: optoelectronic devices (e.g., digital light projectors, spatial light modulators, and high-speed image sensors) and digital signal processing units (e.g., high-performance computers and embedded processors); and (2) the number of patterns required per 3D reconstruction of the software algorithms. From the perspective of hardware, Lei and Zhang *et al.* [14,15] has achieved speed breakthroughs by developing the binary defocusing techniques, where the binary defocusing methods coincide with the inherent operation mechanism of the digital-light-processing (DLP) technology, and permit tens of kHz fringe projection speed by using a digital micromirror array device (DMD). Heist *et al.* [16] introduced a GOBO projector that projects aperiodic sinusoidal fringe patterns with high frame rates and high radiant flux, which can generate more than 1,000 independent point clouds per second. In addition, Zuo *et al.* [17] proposed micro Fourier profilometry (μ FTP), which used high-speed fringe projection hardware as well as the number of patterns reduction strategy to achieve 3D shapes reconstruction at 10,000 fps. From the perspective of algorithms, several composite phase-shifting methods have been proposed to reduce the number of projected patterns required per unambiguous 3D reconstruction [8,17–24]. Liu *et al.* [18] proposed a dual-frequency pattern strategy that embedded low and high frequency components into a single pattern, at least five patterns were required to reconstruct the 3D point cloud. Zuo *et al.* [20] employed two $\pi/2$ phase-shifting sinusoidal patterns and two linear increasing/decreasing ramp patterns to reduce the number of patterns required per 3D reconstruction from five to four. Zhang *et al.* [21] embedded the speckle-like signal in three sinusoidal phase-shifted fringe patterns for absolute depth recovery, which can eliminate the phase ambiguity without reducing the fringe amplitude or frequency. Feng *et al.* [22] presented a two-frame fringe projection technique for real-time 3D measurement, using a speckle image and a speckle-embedded fringe image. Tao *et al.* [23] used three composite fringe patterns embedded with the triangular wave into a multi-view system to strengthen the robustness of phase unwrapping. Qian *et al.* [8] further established a complete multi-view fringe projection system, which can achieve real-time high-precision 360-degree 3D model measurement with only three high-frequency fringe patterns. Nevertheless, high-precision 3D reconstruction using only one single pattern is a considerable challenge and has been the ultimate goal of structured light 3D imaging in perpetual pursuit. In 1983, Takeda *et al.* [25,26] proposed Fourier transform profilometry (FTP), which decoded the wrapped phase by Fourier filtering in the spatial frequency domain and achieved the phase demodulation from a single fringe pattern. Afterwards, a series of influential and improved single-shot fringe analysis methods were proposed, such as windowed Fourier transform (WFT) [27–30] and wavelet transform (WT) methods [31]. Particularly, Su *et al.* [32] applied a single high-frequency fringe projection image for drumhead vibration at a speed of kHz level. However, the key to the success of FTP is that the high-frequency fringe information modulated by the object surface can be well separated from the background intensity in the frequency domain. As a result, the FTP technique [33] is limited to measuring smooth surfaces with limited height variations. Besides, the phase distribution retrieved by FTP, ranging between $-\pi$ to π , suffers from 2π periodic ambiguity. Consequently, the wrapped phases require phase unwrapping algorithms to further obtain the absolute phase distribution [34].

To achieve single-shot phase unwrapping, Takeda *et al.* [35] further introduced frequency multiplexing (FM) to FPP to encode two fringe patterns with different spatial carriers into a single snapshot measurement. The projected composite fringe pattern and its spectrum and intensity calculation function are illustrated in Fig. 1(a). After performing the Fourier transform on the FM composite fringe pattern, the spatial frequencies in two orthogonal directions can be extracted from the spectrum simultaneously, with which the periodic phase ambiguity can be removed. Although this method solved the spectrum aliasing problem to some extent and can measure 3D objects with discontinuous and isolated surfaces, the residual phase errors still lead to phase unwrapping errors. Guan *et al.* [36] used four high-frequency carrier information to convolve

the single-frequency phase-shifting fringe patterns to different positions of the Fourier spectrum (Fig. 1(b)), by which the absolute phase maps can be directly recovered from these modulated single-frequency signals through the temporal phase-shifting algorithm [37]. However, due to the weak anti-noise ability of low-frequency image and residual spectrum aliasing, this method cannot be applied to highly precision measurement fields. Yue *et al.* [38] designed another composite structured light pattern formed by modulating two fringe patterns with π phase difference along the orthogonal direction of two distinct carrier frequencies (Fig. 1(c)). Lu *et al.* [39] proposed a fast modulation measuring profilometry based on a single-shot cross grating projection to reconstruct the 3D shape of the objects (Fig. 1(d)). In addition some other single-shot composite coding strategies, such as spatial neighborhood coding schemes [40] (e.g., De Bruijn sequences), color channels multiplexing methods [41], and directly coding methods [42] can also solve the problem of motion. Although the above methods achieve high measurement efficiency, they suffer from compromised measurement accuracy due to the spectrum aliasing problem of FTP. In recent years, deep learning technology has been applied to FPP as a new tool to solve the measurement efficiency and phase/or depth retrieval accuracy [43,44], such as fringe analysis [44–46], fringe enhancement [47], phase demodulation [48,49], phase unwrapping [50–52], and 3D data acquisition [53–55]. These studies have given us an inspiration-whether it is possible to combine fringe projection profilometry with deep learning techniques to achieve higher precision and more robust phase retrieval and 3D reconstruction from only a single composite fringe image.

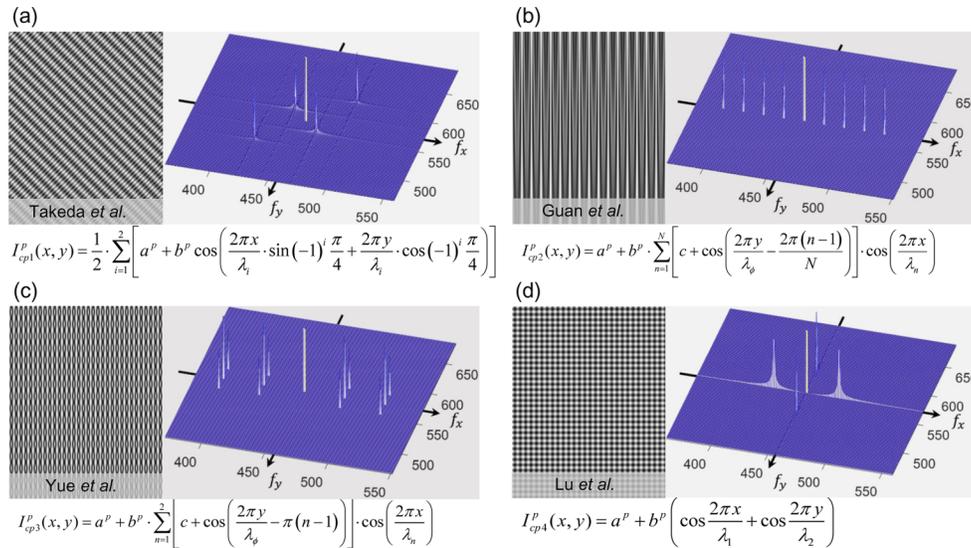


Fig. 1. Spatial frequency multiplexing composite fringe patterns and their corresponding spectrum.

In this work, we present a novel composite fringe projection deep learning profilometry (CDLP), which construct a one-to-three convolutional neural network to analyze the single-shot spatial frequency multiplexing fringe pattern, to reconstruct high-quality 3D shape information in transient scenes. The main contributions of this work are as follows. (1) Under supervised learning, the use of high-quality data sets (including input data and ground truth labels) significantly affects the quality of network model training. In this regard, we first propose a fringe encoding scheme based on spatial frequency multiplexing, which takes into account both unambiguity and multi-spatial information fusion to enhance the training abilities of the deep learning network. Meanwhile, through the N-step phase-shifting (PS) method and multi-frequency temporal phase unwrapping (TPU) combined with the projection distance minimization (PDM) algorithm [17],

we have successively obtained the high-quality wrapped phase numerator, denominator term, and unambiguous unwrapped phase. They serve as three sets of high-quality ground truth labels for our one-to-three single-shot phase retrieval network. (2) For our proposed network framework, we use an improved one-to-three deep convolution neural network to simultaneously achieve high-quality phase analysis and robust phase retrieval. The specific network architecture will be elaborated in subsequent sections. Experimental results validated that the proposed single-shot composite fringe projection deep learning profilometry can directly perform the single-shot robust and unambiguous phase retrieval process and high-quality absolute 3D surface information of the objects under fast, dynamic, and even transient scenes, with a reconstruction accuracy of 60 μm . The remainder of this paper is organized as follows. In Section 2., we start with the basic principles of composite fringe projection deep learning profilometry (CDLP). In Section 3., experimental verifications and comparison results are presented in detail. In the final Section 4., we draw conclusions.

2. Principle of composite fringe projection deep learning profilometry (CDLP)

2.1. Single composite fringe pattern (CFP) coding scheme

To incorporate multiple patterns in an image as the unique input of a deep learning network to achieve a single-shot fast and robust phase retrieval, we consider the following considerations for the composite fringe coding strategy. (1) It is hard to directly use a absolute single-frequency fringe pattern as a deep learning input to predict the high-precision phase. The reason is that when the single-frequency fringe pattern is projected and captured, the camera resolution is inconsistent with the projector's resolution. Thus only one single-frequency fringe cannot obtain complete sinusoidal intensity information and accurate phase information within a sinusoidal period. (2) Although the absolute phase information of the objects can be obtained directly by using the single-frequency N-step phase-shifting fringe images without phase unwrapping [56], the phase accuracy demodulated by the single-frequency fringe is poor; besides, this strategy always needs to project N fringe images, so the speed cannot exceed the single-frame projection. (3) If we consider combining N single-frequency fringe patterns into an image with a complete sine period, we cannot directly superimpose these N single-frequency phase-shifting images into one image, because the final composite result is a white image. (4) Guan's coding strategy [36], which separated the four single-frequency phase-shifting patterns in the frequency spectrum through four high-frequency carrier frequencies, can directly retrieve the unambiguous phase distributions from a single composite image and avoid the complicated unwrapping process; however, due to the low frequencies lack of high-frequency detail information of the objects, this method cannot fulfill the requirements of high-precision 3D shape measurement. To sum up, we propose a novel three short-wavelength superimposed three carrier-frequency composite fringe pattern coding strategy.

The composite fringe pattern (CFP) and its generation process are shown in Fig. 2(a). Firstly, we generate three sets of sinusoidal fringe patterns with different short wavelengths (or high frequencies), which are recorded as fringe patterns to be modulated:

$$I_{\phi_n}^p(x^p, y^p) = a + b \cos\left(\frac{2\pi x^p}{\lambda_{\phi_n}}\right), \quad (1)$$

where (x^p, y^p) is the projector pixel coordinate. The constants a and b are the background intensity and the modulation of the short-wavelength fringe patterns, and their values should strictly meet the following constraints: on the one hand, the cos term of $I_{\phi_n}^p$ after compensation by a and b is non-negative, on the other hand, they must ensure that the patterns can be made to reach the maximum contrast ratio. The wavelengths λ_{ϕ_n} are changed in the phase direction (the x^p dimension). Constant n represents the n th short-wavelength fringe pattern index, $n = 1, 2, 3$.

Then, the short-wavelength fringe patterns $I_{\phi_n}^p$ are respectively multiplied by three standard cosine fringe patterns $I_{carrier_n}$ with different carrier frequencies along the orthogonal direction to produce three composite sub-images. By superimposing these three composite sub-images of each channel, a frequency-multiplexed CFP is generated:

$$\begin{aligned}
 I_{cp}^p(x^p, y^p) &= A + B \cdot \sum_{n=1}^3 I_{\phi_n}^p(x^p, y^p) \circ I_{carrier_n}(x^p, y^p) \\
 &= A + B \cdot \sum_{n=1}^3 \left[a + b \cos \left(\frac{2\pi x^p}{\lambda_{\phi_n}} \right) \right] \cdot \cos(2\pi f_{carrier_n} y^p),
 \end{aligned}
 \tag{2}$$

where $I_{\phi_n}^p$ is the intensity of the projected CFP, A and B are the mean intensity and the amplitude constants to make value of the 8-bit CFP $I_{\phi_n}^p$ between 0 and 255. The operator \circ represents the Hadamard product operation, which is a pixel-level calculation process, shown in Fig. 2(b). The frequencies $f_{carrier_n}$ that change in the orthogonal direction (the y^p dimension) are recorded as the carrier frequency. The designed CFP contains three short wavelengths (modulation frequencies) λ_{ϕ_n} and three carrier frequencies $f_{carrier_n}$. The directions between the short-wavelength fringes $I_{\phi_n}^p$ and the carrier-frequency fringes $I_{carrier_n}$ are orthogonal, so that the modulation frequencies corresponding to the short wavelengths can be modulated at different positions of the Fourier spectrum through different carrier frequencies. Appropriate short wavelengths and carrier frequencies have to be carefully assigned. The selection conditions of the three short-wavelength λ_{ϕ_n} we will discuss in the next section. For the selection of carrier frequency $f_{carrier_n}$ combinations, in order to expand the bandwidth of each modulation channel and minimize channel leakage, the selected carrier frequencies should be separated as much as possible and far away from zero frequency. However, limited by the spatial resolution of the projector and camera, they have to be restricted within a certain range to ensure reliable phase retrieval.

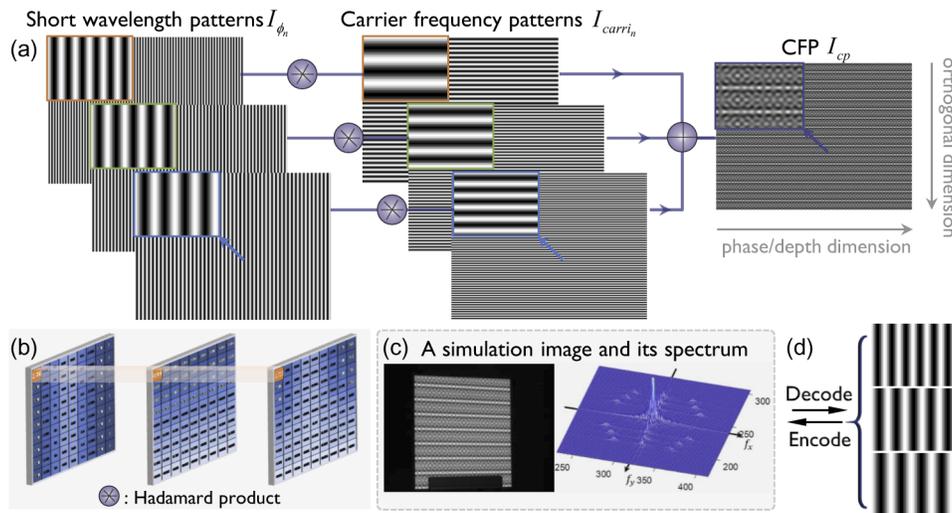


Fig. 2. The composite fringe pattern (CFP) generation process and details. (a) A CFP is formed by modulating and superimposing three short-wavelength fringe patterns with three carrier-frequency fringe patterns along with orthogonal directions. (b) Hadamard product operation between the images of the first channel. (c) A simulation composite fringe image and its spectrum. (d) Conversion between CFP and three fringe patterns.

It should be noted that the intensity range of the projected CFP should be controlled at $[0,255]$, so the intensity of the originally generated CFP needs to be normalized

$$I_{cp}^p(x^p, y^p)' = \frac{I_{cp}^p(x^p, y^p) - I_{\min}}{I_{\max} - I_{\min}} \cdot 255, \quad (3)$$

where $[I_{\min}, I_{\max}]$ is the intensity range of the original CFP. Ideally, after illuminating the object with the composite fringe pattern I_{cp}^p through a digital projector, the intensities of the captured image can be expressed as:

$$I_{cp}^c(x^c, y^c) = \alpha(x^c, y^c) \cdot \left[A + B \cdot \sum_{n=1}^3 I_{\phi_n}^c(x^c, y^c) \cdot I_{carrin}(x^c, y^c) \right], \quad (4)$$

where the fringe maps to be demodulated are

$$I_{\phi_n}^c(x^c, y^c) = a + b \cos \Phi_n(x^c, y^c), \quad (5)$$

and (x^c, y^c) is the pixel coordinate in the camera space, $\alpha(x^c, y^c)$ is the surface reflectivity of the measured object, and $\Phi_n(x^c, y^c)$ is the absolute phase. Due to perspective distortion between the projector and the camera, the actual carrier frequencies f_{carrin}^c of the captured image in the camera view may be different from f_{carrin} . Thus, the relative position of the projector and the camera should be aligned to share about the same world coordinates both in orthogonal direction and the depth direction.

From Eqs. (4) and (5), we can see that the composite fringe image contains three short-wavelength fringe maps (Fig. 2(d)): the three different wavelength fringe patterns can be encoded as one pattern, and the composite fringe pattern can also be decoded to recover these three different wavelength fringe patterns. The phase information of the three fringe maps can be demodulated separately, and then the absolute phase of the object can be retrieved by the phase retrieval algorithm. Therefore, how to accurately demodulate the phase information of one of the short-wavelength fringe images from the obtained distorted composite fringe images is one of the focuses of this work.

2.2. Construction of high-quality network datasets

In order to keep the extracted phase information free from spectrum aliasing, we use deep learning-based methods instead of traditional Fourier transform methods to perform phase demodulation and absolute phase recovery. To begin with, we need to build a high-quality network dataset. The network model learned based on the simulation data may not realistically and comprehensively reflect the actual physical imaging process, and it may not obtain the ideal imaging results. Therefore, we collect and label actual experimental training data rather than simulation data in the deep learning task. In this work, we use a non-composite standard N -step phase-shifting algorithm (PS) for high-quality phase analysis, and use temporal phase unwrapping (TPU) combined with the projection distance minimization (PDM) method to obtain high-precision absolute phase information. The complete process of constructing a high-quality network dataset is shown in Fig. 3.

For the standard N -step phase-shifting algorithm, the fringe images captured by the camera can be expressed as:

$$I_{n(i)}^c(x^c, y^c) = A(x^c, y^c) + B(x^c, y^c) \cos(\Phi_n(x^c, y^c) + 2\pi i/N), \quad (6)$$

where $I_{n(i)}^c$ represents the $(i + 1)$ th n -frequency captured image, $i = 0, 1, \dots, N - 1$, Φ_n is the n -frequency absolute phase map, and $2\pi i/N$ is the phase shift. Then, the phase can be calculated

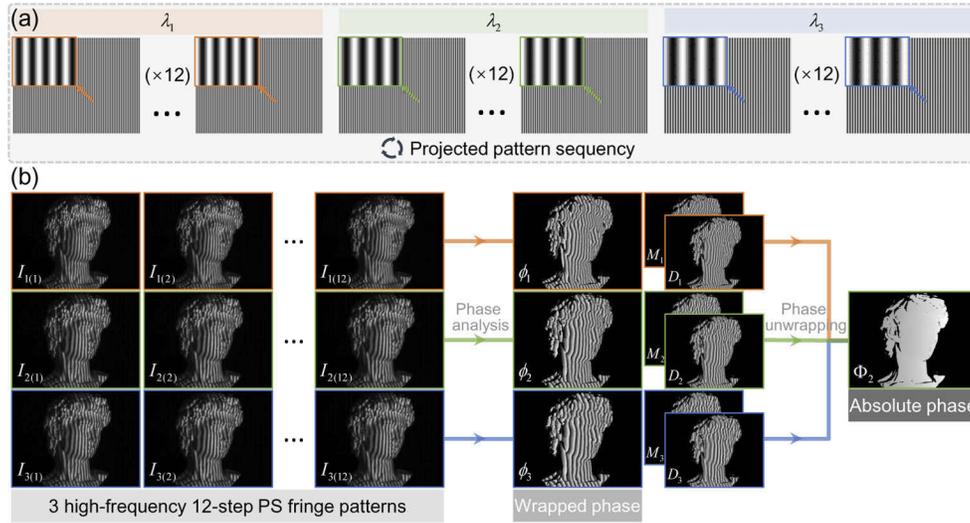


Fig. 3. The process of generating training data. (a) The projected sequence consists of three sets of 12-step phase-shifting fringe patterns with different frequencies/wavelengths. (b) The generation process includes projecting and capturing three sets of fringe images, phase analysis to obtain the wrapped phase, and phase unwrapping to retrieve the absolute phase distribution.

through the least-squares algorithm:

$$\phi = \arctan \frac{\sum_{i=0}^{N-1} I_i^c \sin(2\pi i/N)}{\sum_{i=0}^{N-1} I_i^c \cos(2\pi i/N)} = \arctan \frac{M}{D}, \quad (7)$$

where the subscripts (x^c, y^c) and n are omitted for convenience, M and D represent the numerator and denominator of the arctangent function, respectively. In addition, to improve the image quality and enhance the learning ability of the deep learning network, the image *Mask* function constructed by the modulation function B (Eq. (8)) is used to remove the invalid points of the entire captured image (Eq. (9)).

$$B(x^c, y^c) = \frac{2}{N} \sqrt{M^2 + D^2}, \quad (8)$$

$$Mask(x^c, y^c) = \begin{cases} B(x^c, y^c), & B(x^c, y^c) \geq Thr \\ 0, & B(x^c, y^c) < Thr \end{cases}. \quad (9)$$

The value of threshold Thr is set to 8, which is suitable for most of our measurement scenarios in this work. The initial phase φ we obtained is the relative/wrapped phase within $(-\pi, \pi)$ due to the truncation of the arctangent function. Thus, we need to perform phase unwrapping to remove the ambiguities and correctly extract the absolute phase contribution. In this work, through the obtained multi-frequency fringe wrapped phase maps, we use the temporal phase unwrapping method to eliminate the phase ambiguity in the time domain pixel by pixel. Projection distance minimization (PDM) is an optimal method for solving multi-frequency temporal phase unwrapping. Assuming that three groups of fringe patterns with fringe wavelength $\lambda = [\lambda_1, \lambda_2, \lambda_3]^T$ are obtained by the phase-shifting method, the corresponding relative phase is $\varphi = [\phi_1, \phi_2, \phi_3]^T$, and the absolute phase $\Phi = [\Phi_1, \Phi_2, \Phi_3]^T$ and the wrapped phase satisfy the

following relationship:

$$\Phi = \varphi + 2\mathbf{k}\pi, \quad (10)$$

where $\mathbf{k} = [k_1, k_2, k_3]^T$ is the integer fringe order vector, $k_{1,2,3} \in [0, K - 1]$, and K denotes the number of used fringes. The task of phase unwrapping is to determine the unique fringe orders \mathbf{k} of the wrapped phase, and then obtain the absolute phase maps Φ from Eq. (10). To achieve the goal that the relative phase φ can be successfully unwrapped without ambiguities within the desired measurement range, the fringe wavelength combination should be selected appropriately. On the one hand, given that the projection pattern has W pixels along the horizontal axis wherein the sinusoidal fringe intensity change, on the other hand, considering the fact that the least common multiple of the wavelength combination determines the maximum range of unambiguous phase wrapping along the absolute phase axial direction [17,57], the selected three different wavelengths $\lambda_1, \lambda_2, \lambda_3$ should satisfy the following inequality to exclude phase ambiguous:

$$LCM(\lambda_1, \lambda_2, \lambda_3) \geq W, \quad (11)$$

where $LCM()$ represents the least common multiple function. Refer to the optimal wavelength selection strategy [17,58], the wavelengths λ_n should be sufficiently small to allow more higher accuracy measurement. In particular, we also select the same wavelength combination as the three high-frequency modulation wavelength of the generated CFP (refer to Section 2.1).

$$\lambda_n = \lambda_{\phi_n}, \quad (12)$$

After examining that the pairs of wrapped phase values are unique, the fringe orders k_1, k_2 , and k_3 of the three phase maps can be determined, then we can acquire the high-accuracy absolute phase as part of the high-quality network training datasets.

2.3. One-to-three single-shot phase retrieval network

The crucial step of FPP is to retrieve high-precision and unambiguous phase distribution. Ideally, a monocular FPP system (as shown in Fig. 4) can use only one dense fringe image to robustly achieve high-quality phase unwrapping and absolute 3D reconstruction for complex scenes. However, limited by the number of fringe projection patterns required for 3D imaging, the current traditional FPP methods are still unable to robustly complete high-quality phase recovery under the premise of one projection. To this end, inspired by the recent successful applications of deep learning techniques on FPP, we combine a deep convolutional neural network with a single composite fringe image to develop a one-to-three single-shot phase retrieval network. The flow chart of the proposed method is shown in Fig. 5, which mainly includes: data preprocessing and network model construction, phase analysis and phase recovery based on deep convolutional neural network, phase-to-height mapping.

After training and testing different networks, such as ResNet [59], U-Net [60], and U-Net derivative networks (such as MultiResUNet [61], etc.), we finally choose the U-Net network that takes into account versatility and practicability for model training, and make the following fine-tuning of the network based on the prototype structure of the U-Net network: (1) In order to prevent overfitting caused by the bigger network, the designed U-Net network is reduced by one layer, changed from 5 layers to 4 layers, and we use Dropout which is one of the most effective and most commonly used regularization techniques for neural networks to further fight overfitting. (2) We set the network as a one-to-three convolutional neural network with single input and three outputs, the input channel is a composite fringe image designed in Section 2.1, and the three output channels are the numerator term and the denominator term of the wrapped phase arctangent function, and the coarse absolute phase term.

The improved U-Net network architecture is illustrated in Fig. 5(b). It consists of a contracting path (left side) and an expansive path (right side) [60]. The contracting path (also called Encoder)

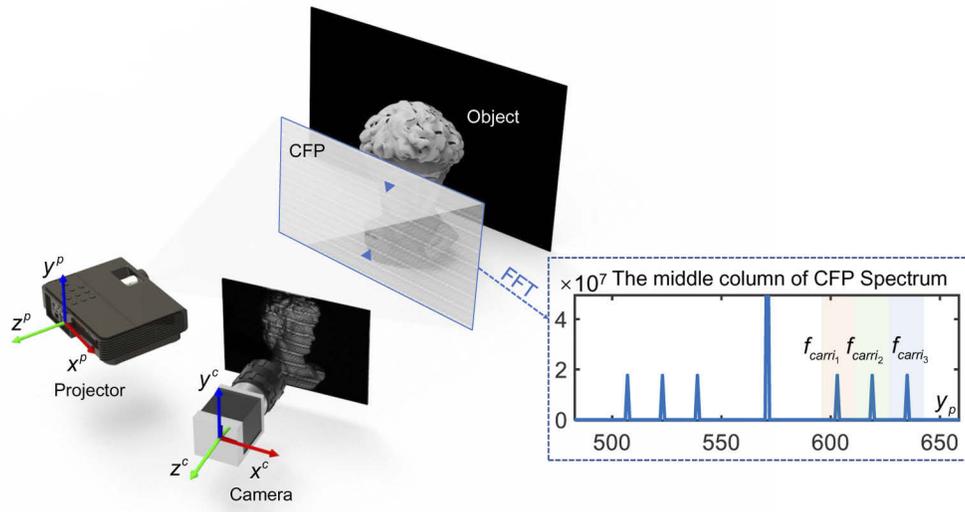


Fig. 4. Hardware system and the middle column distribution of the frequency spectrum generated by the Fourier transform of the designed composite fringe pattern (CFP).

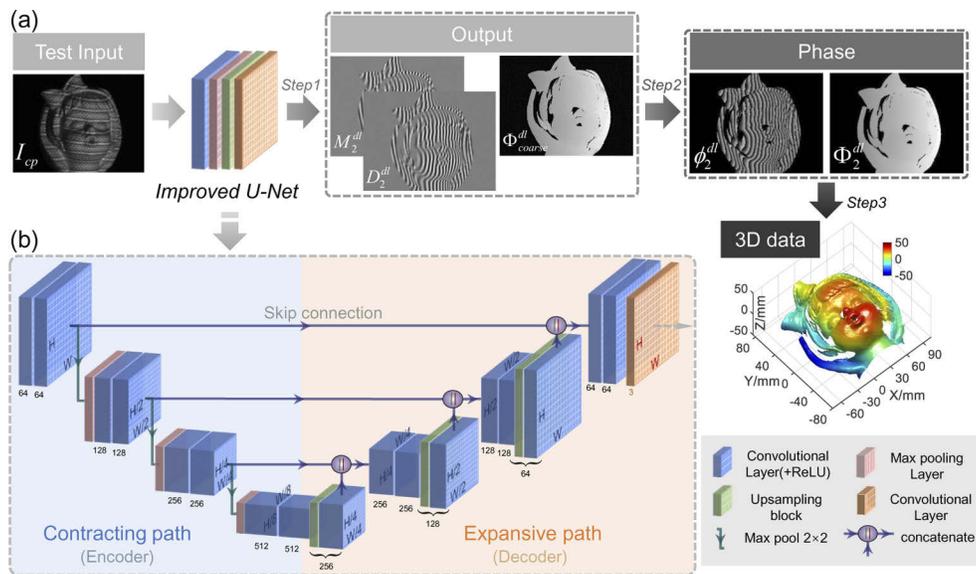


Fig. 5. Flowchart of our proposed approach. (a) Input the test data, output the numerator M_{dl} , denominator D_{dl} and the low-accuracy absolute phase Φ_{coarse} through the trained network model, then obtain the high-accuracy absolute phase by post-processing and reconstruct the 3D information by the calibration parameters. (b) The improved U-Net network architecture.

follows a typical convolutional network architecture, including two convolutional layers (“SAME” padding) that are repeatedly applied, and each convolution is followed by a rectified linear unit (ReLU) and a 2×2 max pooling layer. In each convolution layer of the network, the size of the convolution kernel is 3×3 , which is used for feature extraction, and the padding is set to “same” to ensure that the size of the feature map (H, W) remains the same after each convolution. After the max pooling layer, the size of the feature map will be downsampled with a stride of 2, the size of the feature map will become $(H/2, W/2)$, and the number of feature channels will be doubled. It should be noted that the linear rectification unit (ReLU) in each convolutional layer is one of the important factors to ensure that the deep learning network can be trained, and its operation is as follows:

$$r(\xi) = \max(0, \xi) = \begin{cases} 0, & \text{if } \xi \leq 0 \\ \xi, & \text{otherwise} \end{cases}, \quad (13)$$

where ξ represents an independent variable. The above process needs to be performed 4 times. Especially in the last execution, max pooling is no longer performed, but the feature map is directly sent to the expansion path.

Each step in the expansion path (also called Decoder) includes an upsampling layer followed by a 2×2 convolutional layer, a concatenate, and two 3×3 convolutional layers followed by a ReLU. The upsampling layer used for feature mapping halves the number of feature channels and doubles the size of the feature map. A concatenate merges the upper layer and the corresponding feature map from the contracting path through skip connection to retain more dimensional/location information. This critical step will facilitate the subsequent layers to freely choose between shallow and deep features, which is more advantageous for the semantic segmentation task of deep neural networks. At the final layer, a 1×1 convolutional layer (“SAME” padding) is used to map the required three 3D tensors. The improved U-Net network has a total of 18 convolutional layers.

Next, we will discuss the specific procedures of our algorithm.

Step 1: In order to retrieve high-quality wrapped phase and absolute phase information, we input the three phase-wavelength and three carrier-frequency composite fringe images captured by the camera different test scenarios into the trained improved U-Net network, where the three short wavelengths of the composite fringe pattern are $\lambda_{\phi_1} = 9$, $\lambda_{\phi_2} = 11$, $\lambda_{\phi_3} = 13$ (satisfying Eq. (11)), and the three carrier frequencies are set to $f_{carr1} = 32$, $f_{carr2} = 48$, $f_{carr3} = 64$, respectively. From the perspective of the robustness and accuracy of phase recovery in our algorithm, we finally chose the wrapped phase numerator term M_2 , denominator term D_2 , and absolute phase Φ_2 corresponding to the second wavelength λ_{ϕ_2} as the network label to train our network. Considering the physical model of the traditional phase-shifting algorithm, we choose to predict the numerator and denominator terms instead of directly predicting the wrapped phases. Compared with the network structure that directly connects the fringe pattern to the phase, this strategy bypasses the difficulty of the wrapped phase with the 2π phase truncation and effectively removes the influence of the surface reflectivity variations, so as to achieve higher quality phase analysis to predict the high-quality wrapped phase. Inspired by the traditional composite fringe projection profilometry described in Section 2.1, three coprime short-wavelength fringes that can achieve unambiguous phase retrieval in the time domain are combined into one pattern through three carrier frequency. Compared with a single-frequency fringe pattern as the network input, the three-phase-wavelength and three-carrier-frequency composite fringe pattern ensures the unambiguity of the network input, and at the same time ensures that the phase can be unambiguously unwrapped during the absolute phase retrieval process.

Step 2: After predicting the numerator M_2^{dl} , denominator D_2^{dl} , and coarse absolute phase Φ_{coarse}^{dl} of the composite fringe image through the trained improved U-Net network, the high-quality

wrapped phase map of the second wavelength ϕ_2^{dl} can be calculated:

$$\phi_2^{dl} = \arctan \frac{M_2^{dl}}{D_2^{dl}}. \quad (14)$$

Then, high-quality absolute phase Φ_2^{dl} can be obtained:

$$\Phi_2^{dl} = \phi_2^{dl} + 2\pi \cdot \text{round}[(\Phi_{coarse}^{dl} - \phi_2^{dl}) / (2\pi)], \quad (15)$$

where *round* represents the rounding function. Although the existence of the objects' surface reflectivity α makes the deep learning training model can only predict "coarse" absolute phase with low-precision, its accuracy is sufficient to provide the correct fringe order of the high-quality wrapped phase. The final high-precision absolute phase can be obtained through the high-quality wrapped phase and the correct fringe order.

Step 3: After acquiring the high-accuracy absolute phase, the 3D information of the objects can be reconstructed by utilizing the phase-to-height mapping relationship and the calibration parameters of the FPP system [62]. The relation between the phase and the height coordinates can be written as

$$\begin{cases} x_p = \Phi_2^{dl} W / (2\pi N_{\lambda_2}) \\ Z_w = M_z + N_z / (C x_p + 1) \end{cases}, \quad (16)$$

where x_p is the projector x coordinate, W is the horizontal resolution of the projection pattern, N_{λ_2} is the fringe density, M_z and N_z , and C are the constants derived from calibration parameters.

3. Experiments

To verify the performance of our method, we construct a monocular fringe projection system (Fig. 4), which consists of a digital light processing (DLP) projector (Texas Instruments DLP LightCrafter 4500) with an WXGA resolution (912×1140) DMD and an industrial camera (Basler ace acA640-750 μm) with 640×480 resolution. The camera with the ON Semiconductor PYTHON 300 CMOS sensor delivers 751 fps Frame Rate at VGA resolution. Under the condition of satisfying the above Eq. (11) and Eq. (12), we select $\{9, 11, 13\}$ wavelength combinations to provide unambiguous phase unwrapping for the whole projection range ($LCM(9, 11, 13) = 1287 > 912$). The field of view (FOV) of the measurement system is about $210 \text{ mm} \times 160 \text{ mm}$, and the distance between the camera and the region of interest is 400 mm approximately.

In the supervised training mode, the unambiguous inputs and the corresponding accurately known outputs are required. Figure 6 shows some typical shooting scenes of the training datasets. As mentioned above, a set of input and output network training data includes composite fringe images I_{cp}^c , as well as the numerators M_2 , denominators D_2 and the absolute phases Φ_2 , where M_2 and D_2 are calculated by the 12-step PS method, and Φ_2 is obtained by the three-frequency TPU with PDM method. We collect 1000 sets of data for different scenarios including simple, complex, and isolated objects, and divide them into training sets, validation sets and test sets at a ratio of 8:1:1. The training data sets are used to determine the network weight (Fig. 6); the validation data sets are used to determine when to stop training; after training, the performance is evaluated by test data set that have never been trained.

The constructed neural network is computed on a desktop with Intel Core i7-7800X CPU and a GeForce GTX 1080 Ti GPU (NVIDIA) under the Python deep learning framework Keras with the TensorFlow platform (Google). The optimizer chooses the Adam optimization scheme, which is used to update the network weights with the loss value and achieve better gradient propagation, and its default initial learning rate lr is set to 0.0001. Batch Normalization is adopted mini-batch

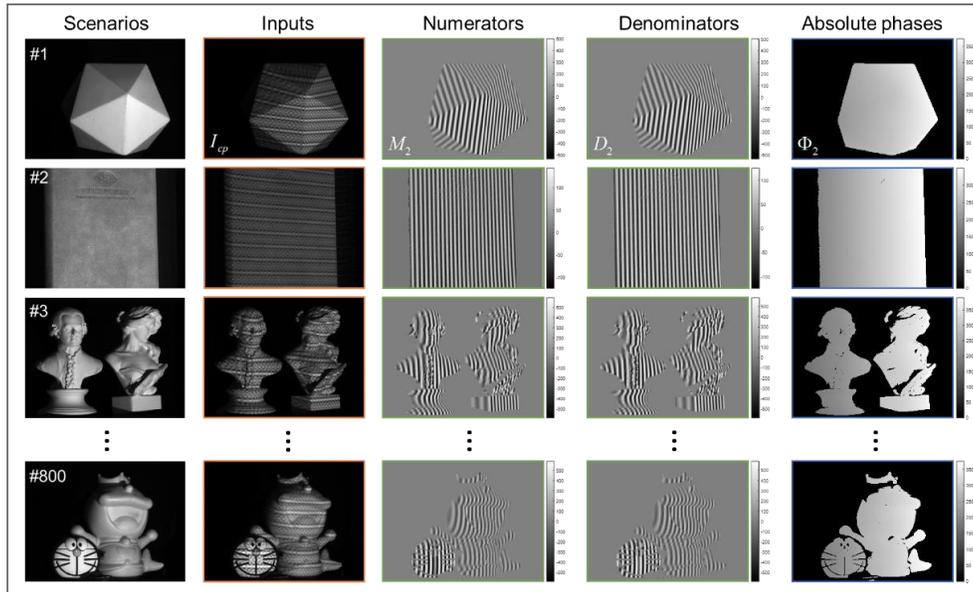


Fig. 6. Part of training datasets. Each row shows the network training scenario and labels. The training dataset includes the input image I_{cp} and three output terms of the second (modulated) frequency fringe: numerator M_2 , denominator D_2 , and absolute phase Φ_2 . The composite fringe images as input are captured in different scenes, including objects with simple or complex surfaces, continuous surfaces or isolated objects, and objects with different materials. These scenes shown in turn are an icosahedral triangle, a customised notebook with “SCILab” logo, Beethoven and a harp girl, as well as a plastic toy and earphone case. The ground-truth data is calculated by 12-step PS and three-frequency TPU with PDM method.

gradient descent scheme (mini-batch size = 2). The loss function we select in this neural network is mean squared error (MSE), which compares the predicted value with the target value after each batch in each epoch and generates a loss value. At the same time, the root mean squared error (RMSE) is calculated after each epoch to help visually monitor the training process. After 200 epochs were trained on the NVIDIA graphics card, the training loss and validation loss of the network converged. Moreover, due to data enhancement (background removal and normalization of input images) and the improvement of the network structure, the entire training time of our network only takes about 3 hours. We can directly put the captured and processed composite fringe image into the trained network model to retrieve the absolute phase map of target object and complete the offline 3D measurement. The network model prediction speed of our approach is about 15 fps.

3.1. Qualitative evaluation

Through “learning” from a large number of data sets, the properly trained neural network can “de-multiplex” high-resolution, spectrum-crosstalk-free phases from the multiplexing composite fringe and directly reconstruct a high-accuracy absolute phase map for single-shot, unambiguous 3D surface imaging. We conducted static and dynamic experiments in several different scenarios to test the trained deep convolutional neural network and verify the superiority of the proposed method over traditional methods.

After analyzing different coding schemes of single-shot structured light illumination (see Sec.1 for detailed analysis), we designed a three-carrier three-frequency composite fringe pattern. Here, we respectively projected the designed composite fringe pattern and the direct composite three-frequency fringe pattern to the scenes. Figure 7 shows the comparison between the designed composite fringe image (Fig. 7(a)) and the direct composite three-frequency fringe image (Fig. 7(b)), and the spectrum distribution of these two kinds of frequency-multiplexing-coded images are shown in Fig. 7(c) and (d), from which we can see that: (1) the directly composite image has serious spectrum aliasing, while our designed composite pattern can separate three close high-frequency information through three carrier frequencies; (2) although the designed fringe pattern avoids spectrum aliasing to some extent, its spectrum is easily affected by the system parameters between the projector and the camera, which results in a slight unknown variation of the three carrier frequencies $f_{carrier}$ in the orthogonal direction. Therefore, it is difficult to demodulate the high-precision phase information through the traditional FT method that uniformly filters three high-frequency channels of the captured composite image by the band-pass filters at the center of $f_{carrier}$. Our deep learning-based method will solve these obstacles at once through a trained convolutional neural network.

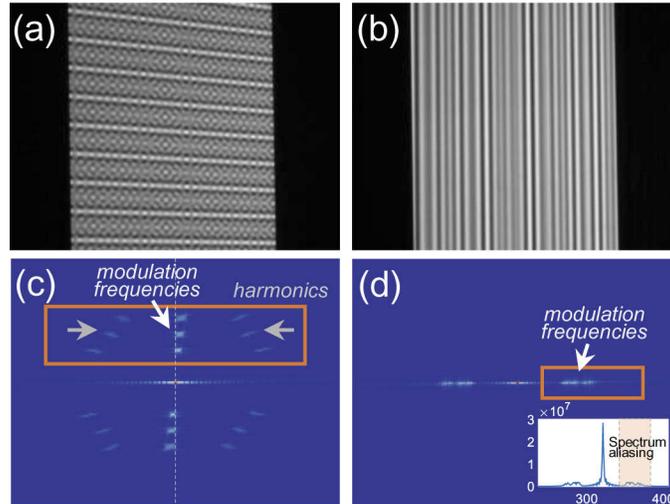


Fig. 7. Comparison of two kinds of frequency-multiplexing-coded schemes. (a) Image of a flat plate obtained by projecting the designed three short-wavelength superimposed three carrier-frequency composite fringe image. (b) Image of a flat plate obtained by projecting the direct composite three-frequency fringe image. (c) Spectrum distribution of (a). (d) Spectrum distribution of (b).

To test the performance of the trained neural network, we measured two static scenarios that include single and multiple isolated objects with different surface roughness. The captured row input composite fringe images $I_{cp}^p(x, y)$ are shown in the first columns of Fig. 8. Note that our neural network has never seen these scenarios during the training phase. After preprocessing these captured composite fringe images, we directly input them into the trained neural network to predict the numerators M_2^{dl} , denominators D_2^{dl} and coarse absolute phase Φ_{coarse}^{dl} of the input fringe images. The results are shown in the second to fourth columns of Fig. 8, where the estimated numerator and denominator are fed into Eq. (14) to obtain the wrapped phase map, and then the unwrapped phase Φ_2^{dl} distribution shown in the fifth column can be retrieved from the calculated wrapped phase and the estimated coarse absolute phase according to Eq. (15). As we can see, the phase ambiguity has been completely eliminated. Furthermore, through the pre-calibrated

parameters of the camera-projector FPP system and the phase-height mapping (Eq. (16)), we converted the unwrapped phase maps into 3D rendered geometries. In Fig. 9, we compared 3D reconstruction results for the traditional FT method using Guan's coding scheme [36], and our learning-based frequency multiplexing-coded method to ground truth. And to quantitatively analyze the phase quality, Figs. 9(c), (e), (h), and (j) show the corresponding unwrapped phase error maps of the entire measurement area. In the investigation, the phases calculated by the 12-step phase-shifting and three-frequency temporal phase unwrapping with projection distance minimization are serve as ground truth phase maps. Due to the influence of severe spectrum aliasing and frequency shift transform, the phase error of the traditional phase retrieval method using Guan's coding scheme is more obvious than our proposed CDLP method. Specifically, to further quantify this trend, we report the mean absolute error (MAE) of unwrapped phase in Fig. 9. Compared with the tradition FT method, our proposed method reduces the projection mode from three to one without losing the accuracy of phase recovery, improving the time resolution without changing the spatial resolution. Compared with the traditional method with Guan's coding scheme, the proposed method improves the phase recovery accuracy by nearly an order of magnitude.

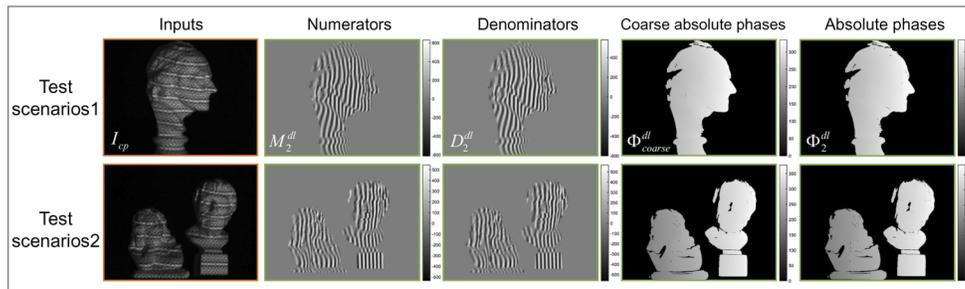


Fig. 8. The prediction results of the two static test scenarios. Each row shows the input composite image $I_{cp}^p(x, y)$, the estimated results of numerator M_2^{dl} , denominator D_2^{dl} , coarse absolute phase Φ_{coarse}^{dl} , and the final absolute/unwrapped phase Φ_2^{dl} .

Although Guan's method avoided a significant degree of spectrum loss, but the slight frequency shifts error brought considerable loss to the 3D reconstruction. From the results of our method, it can be seen that the deep learning-based frequency-multiplexing-coded method obtains a higher quality 3D reconstruction, which is almost comparable to the reference 3D model reconstructed by 12-step PS method (the Ground truth). Moreover, our method only needs one composite fringe image to reconstruct absolute 3D information, and the reconstruction efficiency is 36 times higher than that of the reference method. This experiment verified that the deep learning-based frequency-multiplexing-coded method can not only effectively overcome the adverse effects, such as spectrum aliasing, spectrum leakage, and channel crosstalk, but also achieve high-precision absolute phase retrieval and high-quality absolute 3D surface reconstruction from a single-frame fringe image.

In the second experiment, we measured two sets of moving objects. a rotating bow girl model and a moving Voltaire plaster model, to verify the ability of our method in dynamic scenes. Figures 10(a) and (d) respectively show the raw image of a certain frame in the two captured videos, Figs. 10(b) and (e) are the corresponding 3D reconstruction results using our method in the selected moments, and Figs. 10(c) and (f) further show the 360-degree point cloud registration results. During the measurement, a single-frame composite fringe pattern was continuously projected on the surface of the object, meanwhile, a monochrome camera simultaneously captured the gray fringe image of each frame. We can see that our method is fundamentally immune to phase-shifting errors induced by object motion thanks to its single-shot nature. Consequently, it

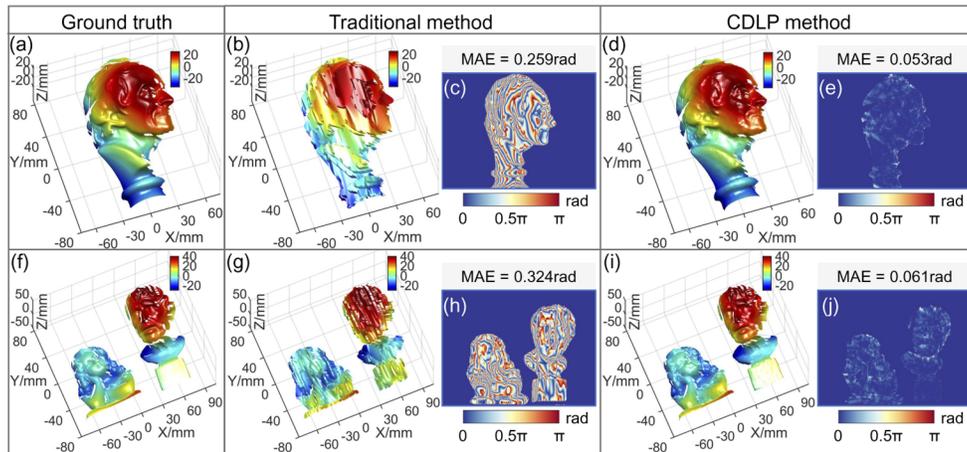


Fig. 9. 3D reconstruction results of Ground truth, traditional method, and our proposed CDLP method in two measurement scenes. (a), (f) 3D reconstruction result of the Ground truth (12-step PS with number-theoretical method). (b) (c), (g) (h) 3D reconstruction result and its corresponding absolute phase error map of traditional method (FT method with Guan's coding scheme). (d) (e), (i) (j) 3D reconstruction result and its corresponding absolute phase error map of CDLP method (composite fringe projection deep learning profilometry).

is suitable for dynamic 3D imaging of rapidly moving objects. The whole measurement process of the rotating statues are shown in Fig. 10 (Multimedia views).

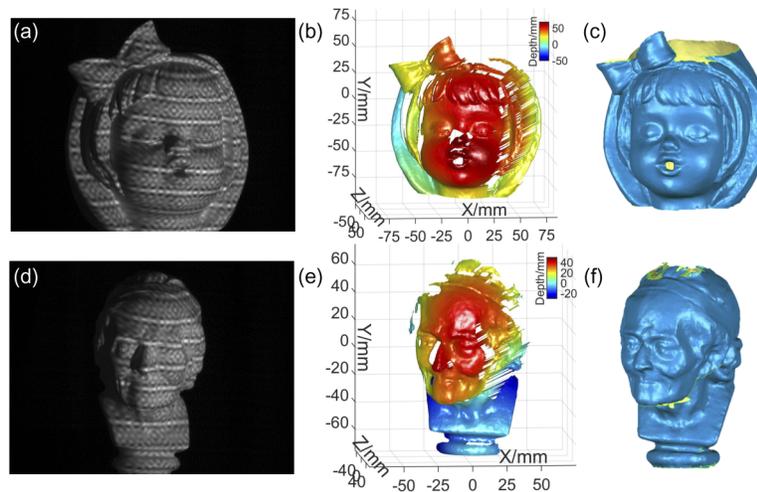


Fig. 10. The dynamic 3D measurement results of a rotating bow girl model and a moving Voltaire plaster model. (a), (d) The captured composite fringe images at two different moments. (b), (e) The corresponding 3D results reconstructed by our method. (c), (f) Registration results. (Multimedia views: see [Visualization 1](#) and [Visualization 2](#) for the whole measurement process of these two scenes)

3.2. Quantitative evaluation

Last but not least, we respectively measured a standard ceramic plate (Fig. 11) and a pair of standard ceramic spheres (Fig. 12) to quantitatively evaluate the 3D reconstruction precision of the proposed method. It is noted that the structural parameters of the standard ceramic balls have been calibrated by the coordinate measuring machine, and their radii are $RA = 25.3999$ mm and $RB = 25.3983$ mm, respectively. The center-to-center distance of the standard ceramic balls is $D = 100.1563$ mm with an uncertainty of $1.0 \mu\text{m}$. We produced the measurement results of the plate and two spheres and performed plane and spherical fitting on the measurement results. Their errors are shown in Figs. 11(b), (c), Figs. 12(c1), (c2), and Figs. 12(d1), (d2). The radii of reconstructed spheres are $RA_{dl} = 25.5246$ mm and $RB_{dl} = 25.2901$ mm, with the mean absolute error (MAE) of 0.0531 mm and 0.0506 mm. The measured center distance is $D_{dl} = 100.2027$ mm with the deviation of $\Delta d = 0.0464$ mm. Additionally, the root mean square (RMS) error of sphere A and sphere B are respectively 0.066 mm and 0.062 mm, as shown in Figs. 12(c2) and (d2). This experiment proves that our method can provide high-quality 3D measurements using only a single fringe image.

4. Conclusions

In this work, we have proposed a deep learning-based single-shot composite fringe projection profilometry (CFPP), which combines a deep learning technology with a specially designed spatial frequency multiplexing coding strategy to achieve single-frame, high-precision, unambiguous 3D shape reconstruction. According to experimental results, this deep learning-based method can perform high-quality 3D shape measurements on discontinuous and/or mutually isolated objects in fast motion. Compared with the existing high-speed 3D imaging method based on multi-frame images, our method is fundamentally immune to motion errors. Compared with the traditional FT and frequency multiplexing FT methods, our approach can effectively overcome the adverse effects, such as spectrum aliasing, spectrum leakage, and channel crosstalk. Using only a single composite fringe pattern, the 3D imaging quality of the proposed method is comparable to the performance of the traditional 12-step PS method. Besides, the trained network model can be fully automatic to achieve high-quality 3D measurement without tuning parameters.

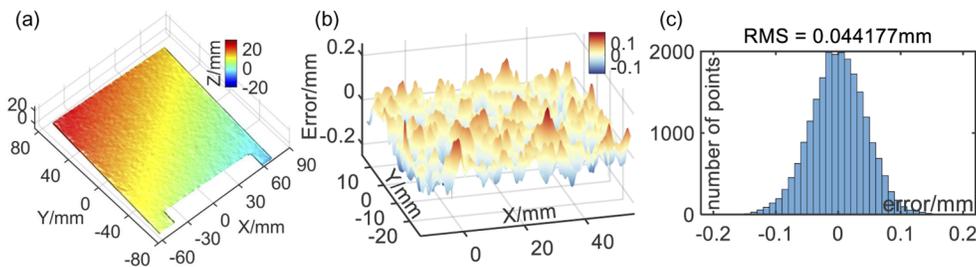


Fig. 11. Precision analysis of standard ceramic plate. (a) 3D reconstruction results by our method. (b) Error distribution. (c) RMS error.

Deep learning technology has thoroughly “permeated” into almost all tasks of optical metrology and has delivered some pretty impressive results. This paper intends to point out that with its powerful learning capabilities, deep learning technology can break the limitations of various influencing factors in traditional single-frame 3D imaging algorithms and achieve impressive results for single-shot, instantaneous absolute 3D shape measurement of discontinuous and/or isolated objects. However, the underlying reasons behind these successes of deep learning prediction remain unclear at this stage. Many researchers are still skeptical and maintain a wait-and-see attitude towards its applications in high-risk scenarios, such as industrial inspection and

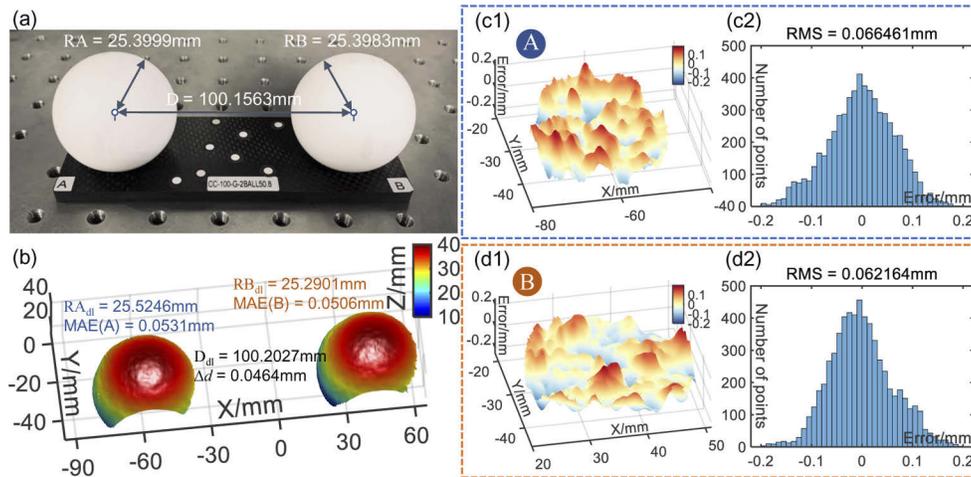


Fig. 12. Precision analysis of a pair of standard ceramic spheres. (a), (b) 3D reconstruction results by our method. (c1), (c2) The error distribution and corresponding RMS error of sphere A. (d1), (d2) The error distribution and corresponding RMS error of sphere B.

medical care. But it can be envisaged that with the further development of artificial intelligence technology, the continuous improvement of computer hardware performance, and the further development of optical information processing techniques, these challenges will gradually be solved in the near future. Deep learning will thus play a more significant role and make a more far-reaching impact in optics and photonics.

Funding. National Natural Science Foundation of China (62075096); Leading Technology of Jiangsu Basic Research Plan (BK20192003); Jiangsu Provincial “One belt and one road” innovation cooperation project (BZ2020007); Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0273); Fundamental Research Funds for the Central Universities (30919011222, 30920032101, 30921011208).

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

- J. Geng, “Structured-light 3d surface imaging: a tutorial,” *Adv. Opt. Photonics* **3**(2), 128–160 (2011).
- S. Feng, C. Zuo, T. Tao, Y. Hu, M. Zhang, Q. Chen, and G. Gu, “Robust dynamic 3-d measurements with motion-compensated phase-shifting profilometry,” *Opt. Lasers Eng.* **103**, 127–138 (2018).
- B. Pan, H. Xie, Z. Wang, K. Qian, and Z. Wang, “Study on subset size selection in digital image correlation for speckle patterns,” *Opt. Express* **16**(10), 7037–7048 (2008).
- Y. Hu, Q. Chen, S. Feng, and C. Zuo, “Microscopic fringe projection profilometry: A review,” *Opt. Lasers Eng.* **135**, 106192 (2020).
- S. S. Gorthi and P. Rastogi, “Fringe projection techniques: whither we are?” *Opt. Lasers Eng.* **48**(2), 133–140 (2010).
- X. Su and Q. Zhang, “Dynamic 3-d shape measurement method: a review,” *Opt. Lasers Eng.* **48**(2), 191–204 (2010).
- S. Feng, L. Zhang, C. Zuo, T. Tao, Q. Chen, and G. Gu, “High dynamic range 3d measurements with fringe projection profilometry: a review,” *Meas. Sci. Technol.* **29**(12), 122001 (2018).
- J. Qian, S. Feng, T. Tao, Y. Hu, K. Liu, S. Wu, Q. Chen, and C. Zuo, “High-resolution real-time 360 3d model reconstruction of a handheld object with fringe projection profilometry,” *Opt. Lett.* **44**(23), 5751–5754 (2019).
- J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, “A state of the art in structured light patterns for surface profilometry,” *Pattern Recognit.* **43**(8), 2666–2680 (2010).
- L. Gao, J. Liang, C. Li, and L. V. Wang, “Single-shot compressed ultrafast photography at one hundred billion frames per second,” *Nature* **516**(7529), 74–77 (2014).
- S. Heist, C. Zhang, K. Reichwald, P. Kühmstedt, G. Notni, and A. Tünnermann, “5d hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light,” *Opt. Express* **26**(18), 23366–23379 (2018).

12. D. Qi, S. Zhang, C. Yang, Y. He, F. Cao, J. Yao, P. Ding, L. Gao, T. Jia, J. Liang, Z. Sun, and L. V. Wang, "Single-shot compressed ultrafast photography: a review," *Adv. Photonics* **2**(1), 014003 (2020).
13. J. Qian, S. Feng, M. Xu, T. Tao, Y. Shang, Q. Chen, and C. Zuo, "High-resolution real-time 360° 3d surface defect inspection with fringe projection profilometry," *Opt. Lasers Eng.* **137**, 106382 (2021).
14. S. Lei and S. Zhang, "Flexible 3-d shape measurement using projector defocusing," *Opt. Lett.* **34**(20), 3080–3082 (2009).
15. S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," *Opt. Eng.* **45**(12), 123601 (2006).
16. S. Heist, P. Lutzke, I. Schmidt, P. Dietrich, P. Kühmstedt, A. Tünnermann, and G. Notni, "High-speed three-dimensional shape measurement using gobo projection," *Opt. Lasers Eng.* **87**, 90–96 (2016).
17. C. Zuo, T. Tao, S. Feng, L. Huang, A. Asundi, and Q. Chen, "Micro fourier transform profilometry (μ ftp): 3d shape measurement at 10,000 frames per second," *Opt. Lasers Eng.* **102**, 70–91 (2018).
18. K. Liu, Y. Wang, D. L. Lau, Q. Hao, and L. G. Hassebrook, "Dual-frequency pattern scheme for high-speed 3-d shape measurement," *Opt. Express* **18**(5), 5229–5244 (2010).
19. C. Zuo, Q. Chen, G. Gu, S. Feng, F. Feng, R. Li, and G. Shen, "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," *Opt. Lasers Eng.* **51**(8), 953–960 (2013).
20. C. Zuo, Q. Chen, G. Gu, S. Feng, and F. Feng, "High-speed three-dimensional profilometry for multiple objects with complex shapes," *Opt. Express* **20**(17), 19493–19510 (2012).
21. Y. Zhang, Z. Xiong, and F. Wu, "Unambiguous 3d measurement from speckle-embedded fringe," *Applied optics* **52**(32), 7797–7805 (2013).
22. S. Feng, Q. Chen, and C. Zuo, "Graphics processing unit-assisted real-time three-dimensional measurement using speckle-embedded fringe," *Appl. Opt.* **54**(22), 6865–6873 (2015).
23. T. Tao, Q. Chen, J. Da, S. Feng, Y. Hu, and C. Zuo, "Real-time 3-d shape measurement with composite phase-shifting fringes and multi-view system," *Opt. Express* **24**(18), 20253–20269 (2016).
24. S. Heist, P. Kuehmstedt, A. Tuennermann, and G. Notni, "Theoretical considerations on aperiodic sinusoidal fringes in comparison to phase-shifted sinusoidal fringes for high-speed three-dimensional shape measurement," *Appl. Opt.* **54**(35), 10541–10551 (2015).
25. M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," *J. Opt. Soc. Am.* **72**(1), 156–160 (1982).
26. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-d object shapes," *Appl. Opt.* **22**(24), 3977–3982 (1983).
27. Q. Kemaoy, "Two-dimensional windowed fourier transform for fringe pattern analysis: principles, applications and implementations," *Opt. Lasers Eng.* **45**(2), 304–317 (2007).
28. Q. Kemaoy, "Windowed fourier transform for fringe pattern analysis," *Appl. Opt.* **43**(13), 2695–2702 (2004).
29. L. Huang, Q. Kemaoy, B. Pan, and A. K. Asundi, "Comparison of fourier transform, windowed fourier transform, and wavelet transform methods for phase extraction from a single fringe pattern in fringe projection profilometry," *Opt. Lasers Eng.* **48**(2), 141–148 (2010).
30. Z. Zhang, Z. Jing, Z. Wang, and D. Kuang, "Comparison of fourier transform, windowed fourier transform, and wavelet transform methods for phase calculation at discontinuities in fringe projection profilometry," *Opt. Lasers Eng.* **50**(8), 1152–1160 (2012).
31. J. Zhong and J. Weng, "Spatial carrier-fringe pattern analysis by means of wavelet transform: wavelet transform profilometry," *Appl. Opt.* **43**(26), 4993–4998 (2004).
32. X. Su, W. Chen, Q. Zhang, and Y. Chao, "Dynamic 3-d shape measurement method based on ftp," *Opt. Lasers Eng.* **36**(1), 49–64 (2001).
33. X. Su and W. Chen, "Fourier transform profilometry: a review," *Opt. Lasers Eng.* **35**(5), 263–284 (2001).
34. C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," *Opt. Lasers Eng.* **85**, 84–103 (2016).
35. M. Takeda, Q. Gu, M. Kinoshita, H. Takai, and Y. Takahashi, "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**(22), 5347–5354 (1997).
36. C. Guan, L. Hassebrook, and D. Lau, "Composite structured light pattern for three-dimensional video," *Opt. Express* **11**(5), 406–417 (2003).
37. C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, and Q. Chen, "Phase shifting algorithms for fringe projection profilometry: A review," *Opt. Lasers Eng.* **109**, 23–59 (2018).
38. H.-M. Yue, X.-Y. Su, and Y.-Z. Liu, "Fourier transform profilometry based on composite structured light pattern," *Opt. Laser Technol.* **39**(6), 1170–1175 (2007).
39. M. Lu, X. Su, Y. Cao, Z. You, and M. Zhong, "Modulation measuring profilometry with cross grating projection and single shot for dynamic 3d shape measurement," *Opt. Lasers Eng.* **87**, 103–110 (2016).
40. J. Pages, J. Salvi, C. Collewet, and J. Forest, "Optimised de bruijn patterns for one-shot shape acquisition," *Image Vis. Comput.* **23**(8), 707–720 (2005).
41. Z. Zhang, D. P. Towers, and C. E. Towers, "Snapshot color fringe projection for absolute three-dimensional metrology of video sequences," *Appl. Opt.* **49**(31), 5947–5953 (2010).

42. G. Sansoni and E. Redaelli, "A 3d vision system based on one-shot projection and phase demodulation for fast profilometry," *Meas. Sci. Technol.* **16**(5), 1109 (2005).
43. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**(8), 921–943 (2019).
44. S. Feng, C. Zuo, L. Zhang, W. Yin, and Q. Chen, "Generalized framework for non-sinusoidal fringe analysis using deep learning," *Photonics Res.* **9**(6), 1084–1098 (2021).
45. S. Feng, Q. Chen, G. Gu, T. Tao, L. Zhang, Y. Hu, W. Yin, and C. Zuo, "Fringe pattern analysis using deep learning," *Adv. Photonics* **1**(2), 025001 (2019).
46. S. Feng, C. Zuo, Y. Hu, Y. Li, and Q. Chen, "Deep-learning-based fringe-pattern analysis with uncertainty estimation," *Optica* **8**(12), 1507–1510 (2021).
47. J. Shi, X. Zhu, H. Wang, L. Song, and Q. Guo, "Label enhanced and patch based deep learning for phase retrieval from single frame fringe pattern in fringe projection 3d measurement," *Opt. Express* **27**(20), 28929–28943 (2019).
48. S. Feng, C. Zuo, W. Yin, G. Gu, and Q. Chen, "Micro deep learning profilometry for high-speed 3d surface imaging," *Opt. Lasers Eng.* **121**, 416–427 (2019).
49. S. Van der Jeught and J. J. Dirckx, "Deep neural networks for single shot structured light profilometry," *Opt. Express* **27**(12), 17091–17101 (2019).
50. W. Yin, Q. Chen, S. Feng, T. Tao, L. Huang, M. Trusiak, A. Asundi, and C. Zuo, "Temporal phase unwrapping using deep learning," *Sci. Rep.* **9**(1), 20175 (2019).
51. J. Qian, S. Feng, T. Tao, Y. Hu, Y. Li, Q. Chen, and C. Zuo, "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3d shape measurement," *APL Photonics* **5**(4), 046105 (2020).
52. J. Qian, S. Feng, Y. Li, T. Tao, J. Han, Q. Chen, and C. Zuo, "Single-shot absolute 3d shape measurement with deep-learning-based color fringe projection profilometry," *Opt. Lett.* **45**(7), 1842–1845 (2020).
53. Z.-W. Li, Y.-S. Shi, C.-J. Wang, D.-H. Qin, and K. Huang, "Complex object 3d measurement based on phase-shifting and a neural network," *Opt. Commun.* **282**(14), 2699–2706 (2009).
54. H. Nguyen, Y. Wang, and Z. Wang, "Single-shot 3d shape reconstruction using structured light and deep convolutional neural networks," *Sensors* **20**(13), 3718 (2020).
55. Y. Zheng, S. Wang, Q. Li, and B. Li, "Fringe projection profilometry by conducting deep learning from its digital twin," *Opt. Express* **28**(24), 36568–36583 (2020).
56. H. Nguyen, N. Dunne, H. Li, Y. Wang, and Z. Wang, "Real-time 3d shape measurement using 3lcd projection and deep machine learning," *Appl. Opt.* **58**(26), 7100–7109 (2019).
57. T. Pribanić, S. Mrvoš, and J. Salvi, "Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning," *Image Vis. Comput.* **28**(8), 1255–1266 (2010).
58. H. Li, Y. Hu, T. Tao, S. Feng, M. Zhang, Y. Zhang, and C. Zuo, "Optimal wavelength selection strategy in temporal phase unwrapping with projection distance minimization," *Appl. Opt.* **57**(10), 2352–2360 (2018).
59. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 770–778.
60. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
61. N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Netw.* **121**, 74–87 (2020).
62. S. Feng, C. Zuo, L. Zhang, T. Tao, Y. Hu, W. Yin, J. Qian, and Q. Chen, "Calibration of fringe projection profilometry: A comparative review," *Opt. Lasers Eng.* **143**, 106622 (2021).