Deep Learning-Driven One-Shot Dual-View 3-D Reconstruction for Dual-Projector System

Yiming Li, Zhuang Li, Chaobo Zhang, *Member, IEEE*, Min Han, Fengxiao Lei, Xiaojun Liang, *Member, IEEE*, Xiaohao Wang, Weihua Gui, *Member, IEEE*, and Xinghui Li, *Member, IEEE*

Abstract—Fringe projection profilometry (FPP) is an extensively used active 3-D measurement technique. However, it faces challenges in achieving synchronous improvement of measurement range, speed, accuracy, and shadow issues. To meet the demand for rapid 3-D reconstruction of two projection views only using a single-shot phase-shifting grating, we initially propose a fast and large-scale 3-D reconstruction system based on deep learning with a dual-projector and single-camera configuration, named DL_DPSL. The key feature is that the entire measurement process only requires two projectors to project one image simultaneously. To validate the effect, we constructed a simulated and a real system and corresponding simulation (∼3300 sets) and real dataset (∼1000 sets), respectively. The experimental results show that based ResUNet the DL_DPSL system can recover the phases of two projection views from one-shot superimposed phase-shifting grating within 0.018 s inclusive projection time when using two NVIDIA GeForce RTX 3090 graphics cards. Additionally, despite reducing the projection time by half, in the simulation dataset compared to the sequential projection, the mean absolute error was reduced by a maximum of 20%. And similar performance improvements in the continuous area of the measured objects in the real dataset. The DL_DPSL system has better accuracy and stability compared to the current fastest deep learning system. DL_DPSL breaks the existing dual-projector structured light 3-D measurement system paradigm, providing a new solution to promote the development and industrial application of FPP technology, with broad application prospects. (Simulation and real datasets

Yiming Li and Xiaohao Wang are with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: liyiming21@mails.tsinghua.edu.cn; wang.xiaohao@sz.tsinghua.edu.cn).

Zhuang Li, Min Han, and Fengxiao Lei are with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China (e-mail: liz22@mails.tsinghua.edu.cn; hanm21@mails.tsinghua.edu.cn; leifx21@mails.tsinghua.edu.cn).

Chaobo Zhang and Xiaojun Liang are with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: zhangchb@pcl.ac.cn; liangxj@pcl.ac.cn).

Weihua Gui is with the School of Automation, Central South University, Changsha 410083, China, and also with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: gwh@csu.edu.cn).

Xinghui Li is with the Tsinghua Shenzhen International Graduate School and the Tsinghua Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: li.xinghui@sz.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3343782

are publicly available on the following GitHub repository: https://github.com/LiYiMingM/DPSL3D-measurement.)

Index Terms—Deep learning, fringe projection profilometry (FPP), real dataset, single-shot superimposed phase-shifting grating, two projection views 3-D reconstruction.

I. INTRODUCTION

FRINGE projection profilometry (FPP) is a popular active 3-D measurement technique [1], [2]. A typical FPP setup consists of a camera, a projector and a computer. The measurement process is that, first the projector projects coded gratings onto the measured object, meanwhile the camera captures the distorted fringe image modulated by its surface height. Then the computing unit decodes absolute phase information. Finally, the phase information can be converted to 3-D world coordinates by combining it with the calibration of the system. Temporal phase unwrapping (TPU) is crucial in FPP to restore a clear absolute phase [3], owning high performance in the presence of significant discontinuities or spatially isolated surfaces compared with Fourier transform profilometry. We first extract the phase information wrapped in $(-\pi, \pi)$ using the arctan function [4], and then we need to retrieve the 2π discontinuity characteristic to unwrap the wrapped phase. Due to its outstanding qualities in high precision, non-contact, quick speed, and surface measurement, FPP technology has been extensively utilized to measure a variety of physical variables such as displacement, strain, surface profile, refractive index, etc. Phase unwrapping is one of the most significant steps in FPP. Apart from the FPP field, phase unwrapping technology is also used in the fields of optical interferometry, synthetic aperture radar [5], digital holographic interferometry [6], and magnetic resonance imaging [7] to calculate the contours and physical data.

FPP technology currently suffers from two fundamental conflicts. First, the disparity between measurement range and measurement accuracy. One of the most notable shortcomings of the FPP is the lack of point clouds due to shadowing and occlusion issues [8]. Besides, owning to the restricted fixed resolution of the projector and the limitations of the projection method, a larger measurement range will inevitably sacrifice accuracy. The second conflict is the issue between measurement precision and speed. In TPU, increasing the number of fringe frames eventually slows down measurement speed however improves accuracy [9].

Many attempts have been made to address the above issues. To resolve the first conflict between measurement range and accuracy, dual-projector or multi-projector systems are the preferable alternatives in terms of enhancing the measurement range and resolving the shadow issue, while simultaneously guaranteeing measurement accuracy [10], [11]. Apart from broadening the measuring area, the dual projectors system may also enhance the signal-to-noise ratio of light intensity and remove the bimodal multi-path issue. When dual projectors work simultaneously in the dual-projector system, it unquestionably can cut down the measurement time. Nevertheless, if more than one projector is running at once, phase-shifting gratings will superimpose together.

Aiming at the second conflict between measurement accuracy and speed, deep learning provides a new solution. In recent years, deep learning has outperformed conventional techniques in areas such as face recognition, defect detection, automatic driving, etc. Due to the rising need for high-speed and precise 3-D measurement, deep learning technology provides faster and more robust predictions similarly in FPP, computational imaging, fluorescence lifetime imaging area [12], etc. Especially in FFP, deep learning scheme covers many aspects including fringe denoising [13], image information enhancement [14], color channel separation [15], phase modulation [16], phase unwrapping [17], error compensation [18], depth reconstruction [19], and others.

To solve the above two challenges, we designed and implemented the simulated and real deep learning-driven dual-projector and single-camera 3-D measurement system with corresponding datasets. Based on ResUNet [20], the experiment performed that it could predict the absolute phase of two projections using only one superimposed grating, and proofed that deep learning can separate superimposed phase-shifting grating and reconstruct a larger range object.

Our contributions mainly include the following four points.

1) *First single-frame dual-view 3-D reconstruction:* We first propose a deep learning-driven fast dual-projector and single-camera 3-D reconstruction system (DL_DPSL) to address the fast 3-D reconstruction of two projection views and resolving shadow and occlusion issues based on the ResUNet, only using an one-shot superimposed phase-shifting grating. The projection time for dual-projection views is reduced by 50%, compared with the current fastest end-to-end 3-D reconstruction based on deep learning.

2) *First dual-projector simulation system:* We present the first simulation system for dual-projector and single-camera (Simu_DL_DPSL), which simulates gaussian surfaces of different shapes, high-order harmonic noise, dual-projector fringe modulation, the process of unwrapping phase, etc. The simulation system generates objects of various shapes, which can benefit correlated research.

3) *First real and simulated datasets:* We establish the first real and simulated datasets consisting of the single-frame superimposed phase-shifting grating and absolute phases of dual-projection viewpoints. They verify the excellent performance in measurement speed and accuracy of the DL_DPSL system.

4) *Fast, wide range and high precision:* This technology can be extended to the application scenario of multi-projectors with a large field of view, when the light in a superimposed area only comes from two projectors. Due to the abundant information in the superimposed phase-shifting grating, the measurement accuracy and stability improve simultaneously.

II. RELATED WORK

To address shadow and occlusion difficulties, dual projectors/multi-projectors systems are frequently employed. However, because of grating superposition generated by simultaneous projection, the two projectors are restricted to project serially which increase the projection time. How to achieve faster and extremely accurate structured light 3-D measurement has long been a research focus. With deep learning's broad applicability in the field of computational optics, numerous work have demonstrated that deep learning scheme outperforms traditional FPP. The objective of our work in this research is to get effective phase information for two views from one superimposed phase-shifting grating without compromising accuracy based on deep learning, so that two projectors can project simultaneously and reduce the projection time.

A. Dual-Projector Structured Light 3-D Measurement System

The dual-projector device is undoubtedly an excellent solution to balance the measurement range and accuracy. In addition, dual-projector measurement can also solve the inherent shadow and occlusion problems of the structured light system. The author once proposed a six-figures separation method for separating superimposed phase-shifting gratings [21]. By designing the projection process of grating encoding and decoding, the dual-view absolute phases can be separated effectively from superimposed gratings. However, this method is limited to the four-step phase-shifting and dual-frequency phase unwrapping method and the speed boost is simply going from serial eight frames to parallel six frames in terms of projection frames. Zhang et al. [22] proposed a separation method based on a color camera and projector to separate the projections of different colors. Despite this, the measurement accuracy is greatly affected by color crosstalk among the three channels. Yu et al. [23] used temporal-frequency multiplexing method to computationally decouple the projected patterns, and find the lookup table method to solve the phase. An innovative plane-sweeping-based technique is presented by Wang et al. [24] addressed this interference from common depth cameras. The suggested technique recovers the depth information for both overlapped and non-overlapped regions by considering the correlation between the infrared pictures and the various projectors as well as the correlation between the infrared images. To separate the superimposed patterns using explicit derivative computations, Je et al. [25] disambiguated numerous projectors by multiplexing the projector patterns' orientations.

The above methods either have a poor separation effect or have quite complex and stringent experimental conditions. Faster 3-D reconstruction remains a challenge and an ongoing goal in many practical problems [26]. Currently, there is

no efficient method for extracting phase information from superimposed fringes in dual-projector systems.

B. Deep Learning-Driven Absolute Phase Recovery Technology

The prediction of phase images is similar with a regression task in computer vision tasks. Deep learning networks have achieved good grades in image prediction [27], high-resolution imaging [28], image segmentation [29], image generation [29], etc. The UNet network [30] and some of its variants are one of the most important networks for computer vision regression tasks [31]. Deep learning has been recently introduced into structured light 3-D measurement area, which can directly transform a single fringe into the numerator image, denominator image, wrapped phase, order map, unwrapping phase or depth image, and simultaneously preserve the object's edges [16] and achieve higher accuracy. Nguyen et al. [19] demonstrated that UNet outperforms fully convolutional networks (FCN) and autoencoder networks (AEN) in end-to-end 3-D reconstruction. Wan et al. [32] proposed a phase retrieval network-FrANet which use two subnetworks to compensate for the missing information for a single UNet. Zuo's team published several work based on the single step reconstruction of numerator and denominator graphs, which could avoid the challenges of bypassing abrupt 2π in the wrapped phase and produce a good quality phase estimate [33]. Qian et al. [4] presented a deep learning-enabled geometric constraints and phase unwrapping approach for single-shot measurement. Huang et al. [34] proved that in predicting order k, the HRnet is superior to other representation methods such as Phasenet 2.0, DeeplabV3, Goldstein's Branch Cut, least-squares phase unwrapping (LSPU) and calibrated phase unwrapping based on least-squares and iterations (CPULSI) method in terms of noise resistance and larger range's phase change. Wang et al. [35] proposed one-step phase unwrapping network from one wrapped phase to unwrapping phase based on UNet and residual network. Yu et al. [36] designed a fringe pattern transformation (FPTNet) for the fringe-to-fringe prediction, transforming a single or two fringes into the phase retrieval required fringes. Song et al. [37] proposed a super-resolution phase retrieval network (SRPRnet). Specifically, a phase-shifting module is proposed to extract multi-scale features with different phase shifts. Wang et al. [20] reconstructed wavefront aberration phase from the distorted image using a convolution neural network (CNN)-ResUNet inspired by UNet, residual block [38] and inception module [39].

The performance of the deep learning technology in structured light 3-D measurement has already surpassed the traditional method. However so far there is no work to verify whether the deep learning network has the ability to extract the correct phase in the one-shot superimposed phase shifts. Besides some work is currently only verified in the simulation dataset, which is not representative that they can also get stable and better results in the real scene.

III. METHODS

We proposed a dual-projector structured light 3-D measurement system based on deep learning (DL-DPSL) and the key feature is the simultaneous projection of the two projectors. The system is the fastest end-to-end structured light 3-D reconstruction technology that only needs a one-shot image as the input to obtain dual-view point clouds as the output. The entire measurement process is shown in Fig. 1, mainly including three parts. First, we established measurement system composed of dual projectors and one camera. What's more, based on the simultaneous projection, we built the dataset of real objects ($\sim$1000 pairs). Input is a one-shot superimposed phase-shifting image, and ground truth (gt) is dual-side absolute phases. Second, we trained and verified the dataset using ResUNet to obtain the output of dual-side unwrapping phases. Finally, we reconstructed the 3-D point cloud for two perspectives based on precalibrated system parameters. Furthermore, in Fig. 1(a), the light intensity projected by the projector 1 and 2 are recorded as L_1 and L_2 separately. Then a superimposed stripe I is captured by the camera

$$L_1(x, y) = \cos(2\pi f_1 x - \phi_1)$$
$$L_2(x, y) = \cos(2\pi f_2 x - \phi_2) \quad (1)$$
$$I = a(x, y) + b_1(x, y)L_1(x, y) + b_2(x, y)L_2(x, y). \quad (2)$$

In (1) and (2), where L_1, L_2, and I respectively represent the sinusoidal grating projected by projector 1, 2 as well as the one-shot superimposed phase-shifting grating image captured by the camera. f_1, f_2, ϕ_1, and ϕ_2 means the grating frequency and the initial phase of the light projected by projectors 1 and 2 separately. And (x, y) represents the image pixel. $a(x, y)$ and $b(x, y)$ are the light intensity relating to ambient light and modulation intensity relating to the object's surface reflectivity, respectively. The label 1 and 2 in the lower right of all elements represent projectors 1 and 2.

A. Algorithms for Obtaining the Ground Truth

The standard N-step phase-shifting algorithm is one of the most widely used methods in the FPP area [40], because of its simple operation, high accuracy, and strong robustness. Figs. 3(c) and 4(b) show the process of absolute phase recovery (gt) using N-steps and multifrequency phase-shifting algorithms. The sinusoidal phase-shifting grating is expressed as

$$I_n(x, y) = a(x, y) + b(x, y)\cos(2\pi f_0 x - 2\pi n/N) \quad (3)$$

where I_n represents the nth ($n = 0, 1, 2, \ldots, N-1$) sinusoidal grating image captured by the camera. (x, y) represents the image pixel. a, b, f_0, and N are ambient light intensity, surface reflectivity intensity modulation, frequency of projecting grating, and the number of phase shifts. The wrapped phase shown in the process B of Fig. 3(c) can be calculated by the following arctan function [41]:

$$\phi(x, y) = \tan^{-1} \frac{\sum_{n=0}^{N-1} I_n(x, y) \sin(2\pi n/N)}{\sum_{n=0}^{N-1} I_n(x, y) \cos(2\pi n/N)}. \quad (4)$$

The wrapped phase calculated in (4) belongs to $[-\pi, \pi]$, which means the wrapped phase has 2π phase discontinuities. Multifrequency TPU method provides the best unwrapping
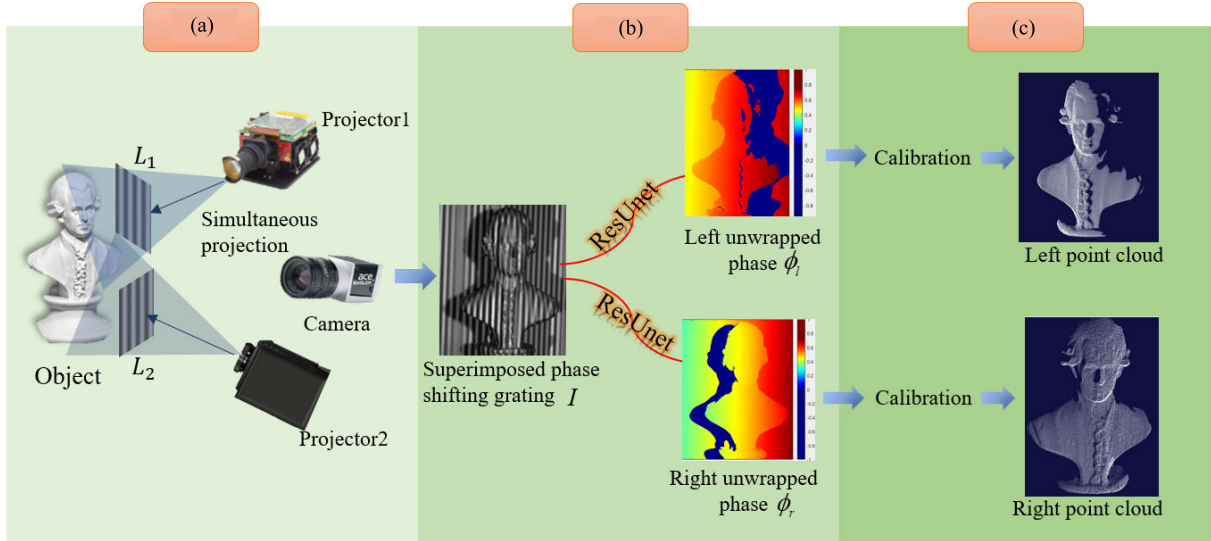
Fig. 1. Whole measurement process for the proposed end-to-end (one-shot image as input, dual-side large-scale point clouds as output) structured light 3-D reconstruction method for dual-side simultaneous projection. (a) Diagram of hardware setup and projection mode. (b) Prediction process diagram for two perspectives phase predicted by ResUNet using a single grating image. We obtain the output of dual-side absolute phases ϕ_l and ϕ_r from one-shot superimposed phase-shifting grating I. (c) Reconstruction process diagram for two-perspective 3-D point cloud from corresponding absolute phase by system calibration.
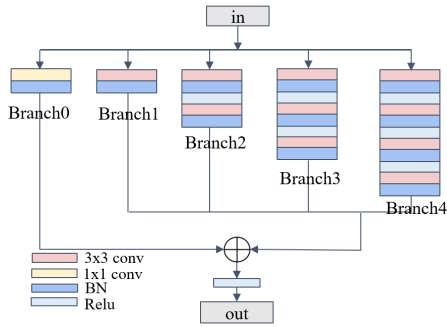

Fig. 2. Residual module.

reliability in the statistical sense and the largest unambiguous measurement range [3]

$$\begin{cases} \Phi_h(x, y) = \phi_h(x, y) + 2\pi k_h(x, y) \\ \Phi_l(x, y) = \phi_l(x, y) + 2\pi k_l(x, y) \end{cases} \quad (5)$$

$$k_h(x, y) = \mathrm{Round}\left[\frac{(\lambda_l/\lambda_h)\phi_l(x, y) - \phi_h(x, y)}{2\pi}\right] \quad (6)$$

where k_h, ϕ_h, and Φ_h represent the fringe orders, wrapped phase, and absolute phase of high-frequency phase-shifting grating. k_l, ϕ_l, and Φ_l correspond to the low-frequency items. In summary, we used four-frequency and six-step phase-shifting method in simulation datasets and four-frequency and 12-step phase-shifting method in real datasets to obtain the complete phase information (gt).

B. ResUNet

ResUNet [20] is the end-to-end deep learning network we used for this separation task shown in Fig. 3(b). Network input is one high-frequency superimposed grating, and network output is a single-view absolute phase image. ResUNet

mainly includes the UNet structure, the residual connection [38], and the inception module [39]. The UNet architecture is widely popular in the field of regression tasks. The inception model was introduced in the GoogLeNet and has subsequently been widely adopted after various improvements. It utilizes 1×1 convolutions for dimensionality reduction and augmentation. By stacking more convolutions in modules of the same size, it enables the extraction of richer features. The residual connection can be implemented as skip connections, where the input of a unit is directly added to its output and then activated. This design allows for more free-flowing information and preserves both fine-grained details and semantic information from the original input data.

ResUNet retains UNet's U-shaped structure of encoder, decoder, and skip-connection operation, allowing for the extraction of features from input images and restoration to their original sizes. Additionally, it incorporates skip connections to preserve texture information of input images, which ensures that each feature map before upsampling in the decoding stage incorporates more low-level features. Residual module shown in Fig. 2 combines inception structure and residual connection. The key point of inception module is that the aggregation of different size convolutions in parallel allows for the extraction of features at different scales within the feature maps. The four right paths (Branch1–Branch4) employ singular, dual, triple, and quadruple 3×3 convolutions, respectively, to enhance the efficacious extraction of features across varying scales. Furthermore, the concatenations of subbranches' output facilitate the amalgamation of features from disparate scales. Branch0 undergoes dimensionality reduction and nonlinear feature extraction through a 1×1 convolution, and the output is residual connected with the aggregated output of the other four branches. Residual connection facilitates easier propagation of information to subsequent layers, thereby mitigating information loss and addressing the degradation
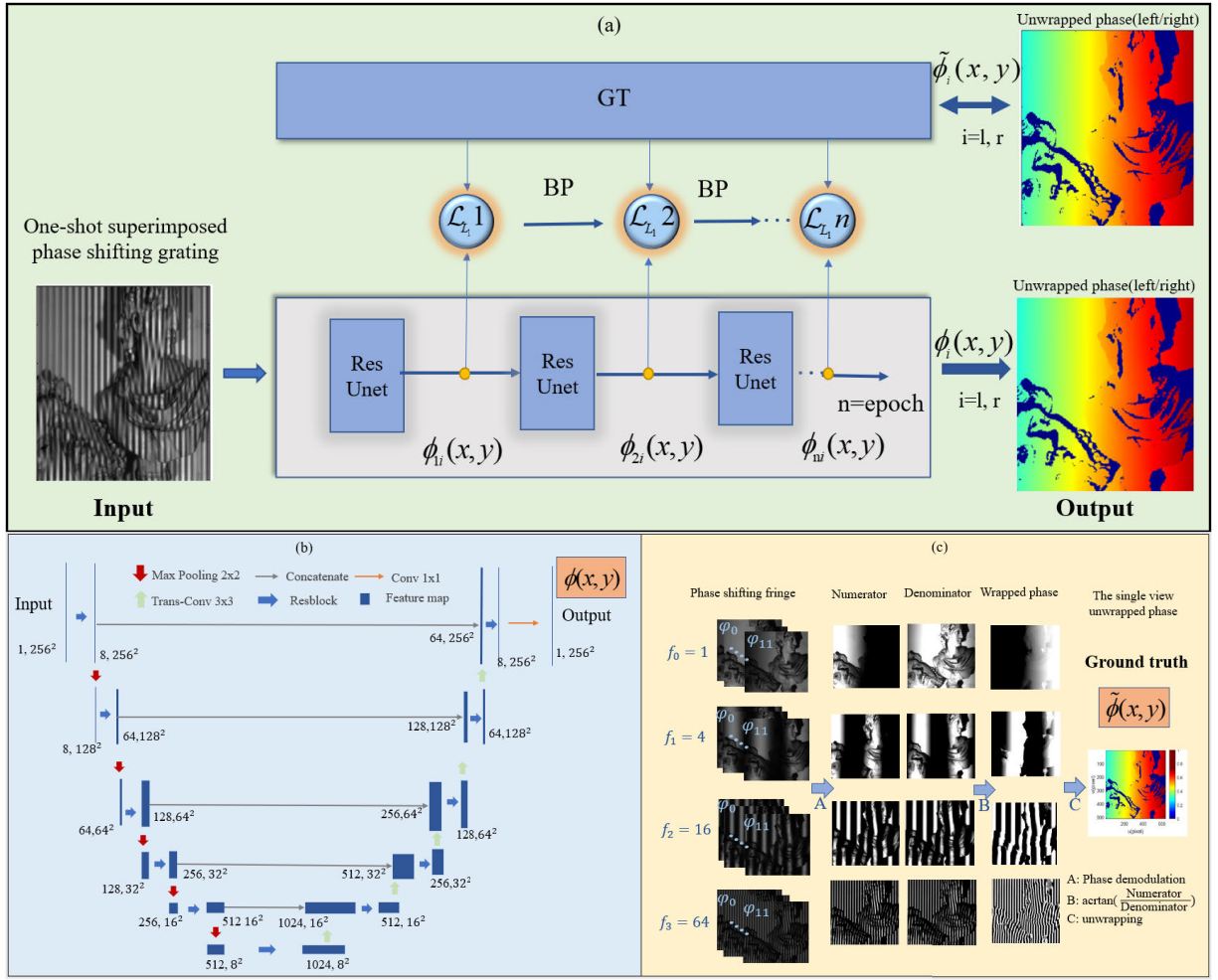
Fig. 3. Training process diagram of DL_DPSL system. (a) Training process of ResUNet. The input is an one-shot superimposed phase-shifting grating. The network is ResUNet. The output and gt is $\phi_i(x, y)$ and $\tilde{\phi}_i(x, y)$, respectively. $\phi_{ni}(x, y)$ represents the ResUNet's nth-epoch output, where $i = (l, r)$, the l, r representing the left view and right view separately. $L_{L_1}n$ represents the nth-epoch $L1$ loss between gt and output. (b) Structure of ResUNet. ResUNet mainly includes five downsampling modules (max pooling), five upsampling modules (transposed convolution), and five concatenate modules. Feature extraction adopts the Resblock module. The (m, n^2) surrounds each feature map, where m represents the number of channels and n represents the side length of the image. (c) Process of obtaining the gt. It shows the flowchart of the four-frequency and 12-step phase-shifting unwrapping method. After generating 48 phase-shifting gratings, we obtain the corresponding-frequency numerator and denominator by process A (phase demodulation). Then by process B (arctan function), we get the wrapped phase of each frequency. Finally, by the process C (phase unwrapping), we calculate the absolute phase (gt) $\tilde{\phi}(x, y)$.

problem. Additionally, it greatly improves training speed and enables increased depth of the network. By extracting feature with the residual module, the prediction accuracy is raised dramatically and the overfitting problem is effectively addressed.

On the whole, in the encoding part, the input image is passed through five residual blocks for feature extraction, and each residual block has a max pooling after it for downsampling. In Fig. 3(b), in (m, n^2), m represents the number of channels in each layer, and n represents the length of the image. Each time the feature image passes through a layer of residual module, the number of channels increases to m shown in Fig. 3(b), and after one Max pooling the image size is reduced by half. Besides, in the decoding part, the deepest feature images are passed upward to five residual blocks for feature recovery, and each residual block has a transposed convolution after it for upsampling. Meanwhile, the gray arrow connects the encoding part and the decoding parts through the concatenate structure, fusing the feature map of the two parts and then passing upward together. In contrast to the dimensional changes during

encoding, each time the feature image passes through a layer of residual module, the number of channels decreases to the number shown in Fig. 3(b). Then through a transposed convolution the image size is doubled. The end of the network module is a 1×1 convolution, which restores the feature map to a single channel for output.

C. Training Process of ResUNet

The DL_DPSL system relies heavily on real datasets to train its networks. The training process of ResUNet is shown in Fig. 3(a), where $\phi_{ni}(x, y)$ represents the ResUNet's nth-epoch output, and $L_{L_1}n$ represents the nth-epoch $L1$ loss between gt ($\tilde{\phi}(x, y)$) and output ($\phi(x, y)$) calculated by (7). To reduce the loss and decrease the gap between the output and gt, calculation process is shown in Fig. 3(c), we estimated the gradient of each weight parameter for the loss function by backpropagation (BP) and subsequently updated each weight parameter using gradient descent. After n iterations of the
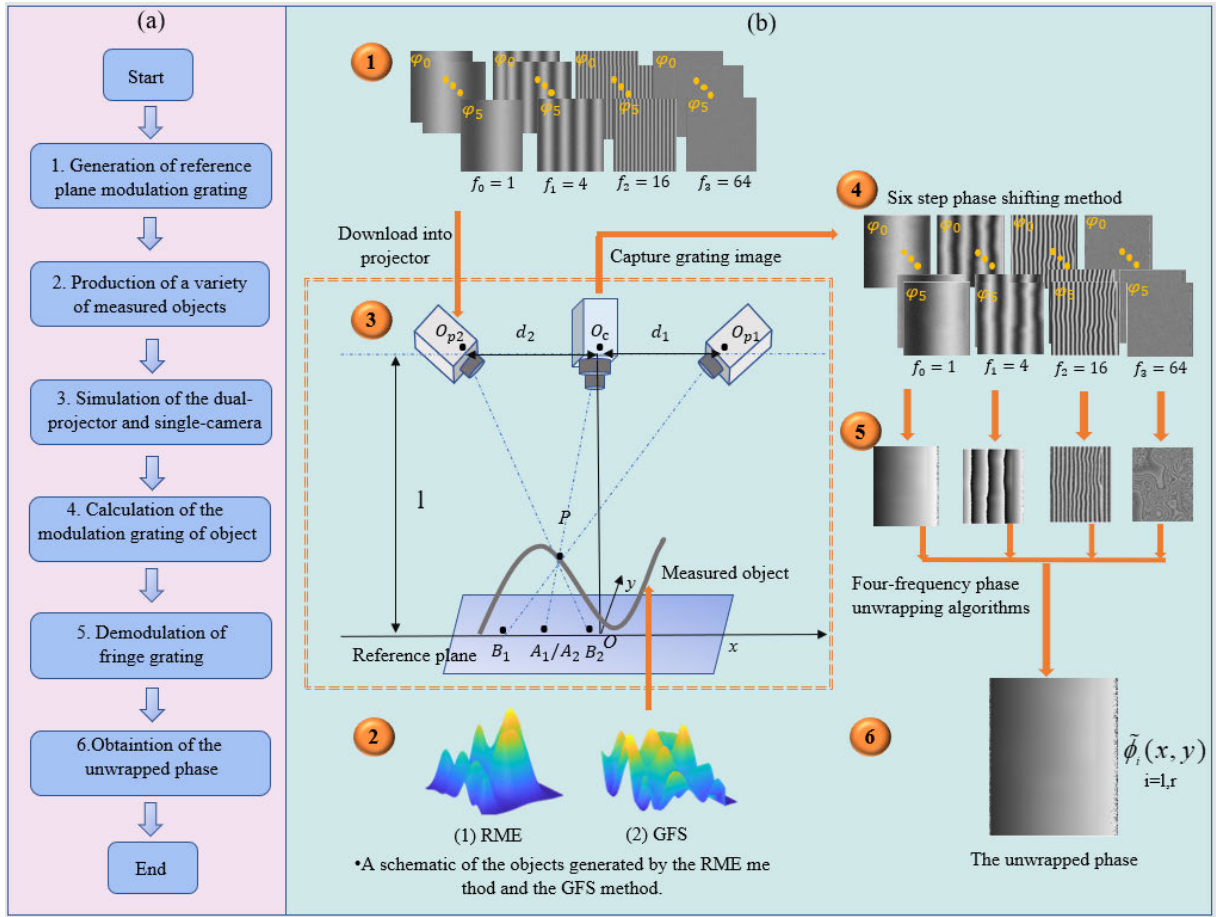
Fig. 4. Simu-DL_DPSL system's creation process for generating simulation datasets. (a) Process description of each step in (b). (b1) Reference plane modulation gratings generated by four-frequency ($f_0 = 1$, $f_1 = 4$, $f_2 = 16$, $f_3 = 64$) and six-step ($\varphi_n = 2\pi(n-1)/6$, $n = 1, 2, \ldots, 6$) phase-shifting algorithm. (b2) Method for object generation by RME and GFS methods. (b3) Measurement system that simulates the measurement situation of dual projectors and one camera. (b4) 24-shot phase-shifting images modulated by object height captured by the camera. (b5) Wrapped phase of four frequency gratings. (b6) Calculation of object's gt $\tilde{\phi}(x, y)$ (absolute phase).

epoch, we completed the network's training until the loss function effectively and stably converged.

The ResUNet was trained by the Pytorch framework (version 2.0.1+cu117+Python 3.9.13) using an NVIDIA GeForce RTX 3090 GPU. The $L1$ norm is the absolute average error between predicted output and gt in the regression model. We use the $L1$ loss as the loss function for graph-to-graph regression

$$\mathcal{L}_{L_1} = \frac{1}{N} \sum_{x,y} \left| \phi(x, y) - \tilde{\phi}(x, y) \right| \tag{7}$$

where L_{L_1} is the value of mean absolute error, and $\phi(x, y)$ and $\tilde{\phi}(x, y)$ (shown in Fig. 3) are the predicted output and gt values, respectively. N represents the amount of input image pairs.

IV. ESTABLISHMENT AND VERIFICATION OF THE SIMULATION SYSTEM

To determine the effectiveness of this technique as soon as possible, simulation and experiment verification are conducted. To provide a more accurate description of ResUNet's capability for absolute phase recovery, we conduct a performance comparison based on our simulated and real datasets among

UNet [30], HRnet [34] and ResNet18. UNet and HRnet were introduced in "related work." ResNet18, a traditional network for regression tasks [38], [42] with comparable parameters, also took part.

A. Setup of the Simulation System

As shown in Fig. 4, we propose a simulation system of dual-projector (two projectors project simultaneously) structured light 3-D measurement, named Simu-DL_DPSL and generate corresponding simulated dataset. There are ~3000 image pairs (with 256×256 pixels) in the training dataset and ~300 image pairs (with 256×256 pixels) in test datasets. A single-shot high-frequency superimposed phase-shifting grating captured by the camera in the case of simultaneous projection of dual projectors is used as the input of the simulation dataset. We choose the grating with a fringe frequency of 64 and an initial phase of $2\pi/3$ as the input in the simulation dataset. And the gt is the absolute phase when the projectors project in turn. Simulation objects are unique Gaussian surfaces with heights ranging from 0 to 60 mm. Fig. 4 illustrates the creation process of the Simu-DL_DPSL system and the generation of simulation datasets.
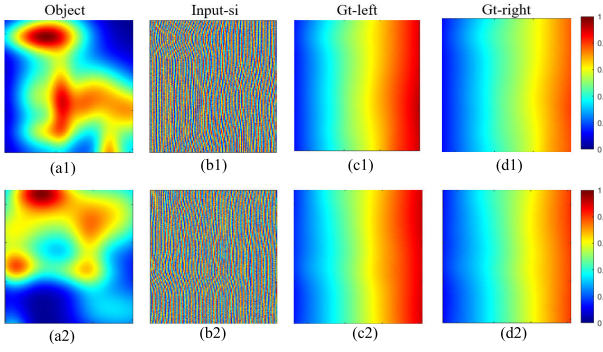
Fig. 5. Two representative objects in simulation dataset. (a1) Depth map of objects 1. (b1) Input (one-shot superimposed phase-shifting grating) of the measured object 1 when the two projectors projected at the same time. (c1) gt (absolute phase) of object 1 in the left view. (d1) gt (absolute phase) of object 1 in the right view. As for object 2, the naming of (a2)–(d2) is consistent with object 1.

To illustrate the phase solution process more clearly, Fig. 4 only shows the situation in which two projectors project sequentially, using projector 2 as an example. The superimposed grating is formed by (2). First, we generated reference plane modulation gratings in four-frequency ($f_0 = 1$, $f_1 = 4$, $f_2 = 16$, $f_3 = 64$) and six-step ($\varphi_n = 2\pi(n-1)/6$, $n = 1, 2, \ldots, 6$) phase-shifting algorithm. After that, 24 grating images were downloaded into the projector 2. Second, we established a variety of measured objects different from each other primarily employing random matrix enlargement (RME) [43] and Gaussian function superposition (GFS) [44], [45] methods. Third, we established a measurement system that simulates the measurement situation of dual projectors and one single camera. Subsequently, dual projectors projected simultaneously, and the camera captured the one-shot superimposed phase-shifting image. Finally, we got the object's phase information using a four-frequency and six-step phase-shifting unwrapping algorithm. Besides, high-order harmonics noise is the most dominant type of noise in phase-shifting profilometry (PSP). Nonsinusoidal waveforms can be approximated as an ideal sinusoidal function distorted by high-order harmonics. To imitate the real-world measurement environment, high-order harmonic noise is added to the camera's raster image. As a result, we adopted the sixth harmonic noise whose amplitudes are 0.2, 0.001, 0.004, 0.002, 0.001, and 0.001.

B. Object Generation Method for Simulation

To generate the diverse measured objects in the simulation dataset, we adopted RME, and GFS, as illustrated in Fig. 4(b2). The mixed-use of the two methods makes the network more robust and stronger. To introduce randomness into the RME method, a small square matrix is generated with an initial matrix size of 2-8, which is then interpolated and linearly mapped into a large matrix to ensure continuity. GFS generates some (ours: 15) Gaussian functions with varying center positions (ours: random values between 20 and 236 mm in x- and y-directions), standard deviations (SDs, ours: random values between 10 and 40 mm in x- and y-axes), and amplitudes (ours: random values between 10 and 20 mm). Finally, add these Gaussian functions together. Data continuity can be ensured by a Gaussian function, while randomness can

be introduced by arbitrarily Gaussian parameters and random superposition styles.

C. Phase-Height Model of Dual Projectors Measurement System

We built the phase-height model for the dual-projector structured light 3-D measurement system, as shown in Fig. 4(b3). Dual projectors in the imaging system allow it to measure phase information twice at the same location. It raises the system's visual presentation, measurement precision, and interference-prevention capabilities. Equation (8) displays the height of point P in the scenario of simultaneous projection by two projection devices. The relationship between the height and the phase of the double projection is as follows:

$$h = \frac{l}{1 + \frac{2\pi}{\lambda} \frac{d_1+d_2}{(\theta_{B_1}-\theta_{A_1})+(\theta_{B_2}-\theta_{A_2})}}.$$

In (8) and Fig. 4(b3), where P is a random point on the measured surface. h is the height of point P relative to the reference plane. O_C, O_{P1}, and O_{P2} are the optical centers for the camera, projector 1, and projector 2 separately. The distance between O_{P1} and O_C, and the distance between O_{P2} and O_C, are d_1 and d_2, respectively. $d_1 = d_2 = 150$ mm. l is the vertical distance between the reference plane and the plane where O_C, O_{P1}, and O_{P2} are located and $l = 600$ mm. θ_{B_1} and θ_{A_1}, respectively, represent the phases of B_1 and A_1 under the projection of projector 1. θ_{B_2} and θ_{A_2}, respectively, represent the phases of B_1 and A_1 projected by projector 2.

D. Verification and Evaluation of the Simulation System

We obtain the simulation dataset containing one-shot dual-projection superimposed phase-shifting grating and dual-view absolute phases based on the simu-DL_DPSL system. Take two simulation objects as examples, as shown in Fig. 5, where (a1) and (a2) denote the depth map of objects 1 and 2, respectively. Fig. 5(b1) and (b2) represent the superimposed phase-shifting grating of the measured objects 1 and 2, respectively, when the two projectors are projected at the same time. Fig. 5(c1) and (c2) represent the left view gt (absolute phase) of objects 1 and 2, respectively. Fig. 5(d1) and (d2) represent the right view gt (absolute phase) of objects 1 and 2, respectively. We evaluate the 300 pairs of testing results in simultaneous projection by a trained ResUNet, as shown in Fig. 3(b), compared with the end-to-end deep learning system when projectors project sequentially (DL_SPSL). The method to get left and right absolute phase is called DL_SPSL (left) and DL_SPSL (right). Methods we use to get the absolute phase of left and right views when the dual projectors projected simultaneously are named DL_DPSL (left) and DL_DPSL (right), respectively. Table I represents DL_DPSL and DL_SPSL's absolute phase error of 300 sets of data in the simulation testing dataset predicted by ResUNet, UNet, HRnet and ResNet18. MAE refers to the average error. RMSE is the root mean square error, and SD is the standard deviation. Note that the data in Table I show the normalized phase, and it needs to be multiplied by 130π to restore the real phase value. In the

TABLE I
GENERAL EVALUATION OF PHASE ERROR ON SIMULATION TESTING DATASET

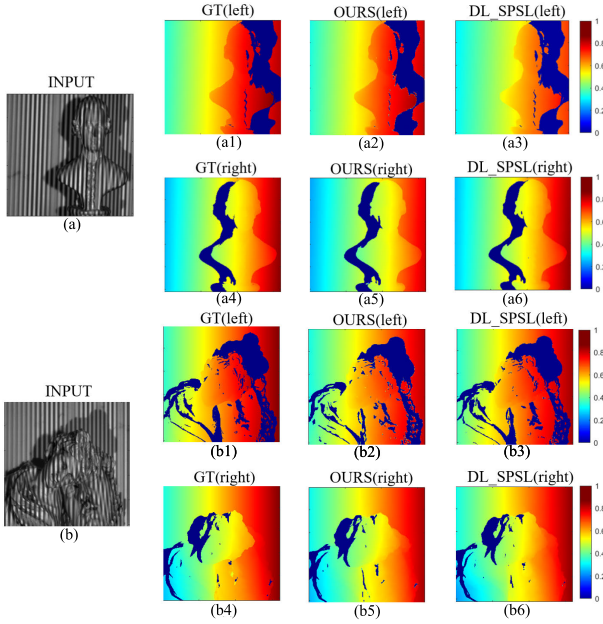| Method	Projection view	MAE(10e-4) ↓		RMSE(10e-4) ↓		SD(10e-4) ↓	
		DL_SPSL	DL_DPSL(ours)	DL_SPSL	DL_DPSL(ours)	DL_SPSL	DL_DPSL(ours)
HRnet	left	16.5	26.5	82.3	108.7	1.7	7.8
	right	71.4	72.7	665.0	657.3	17.4	6.8
ResNet18	left	3.5	3.5	6.6	6.7	7.3	7.3
	right	2.9	2.5	5.9	5.1	6.8	5.4
UNet	left	18.6	9.0	63.1	14.4	13.2	12.3
	right	8.0	6.9	13.4	11.1	11.5	10.0
ResUNet	left	3.4	**2.8**	5.0	**4.0**	1.8	**2.4**
	right	2.0	**1.9**	3.2	**3.3**	2.1	**2.0**

Fig. 6. Input and gt of real dataset, as well as the output of ResUNet using DL_DPSL and DL_SPSL system-examples of representative real objects A and B. (a) Input of object A (one-shot superimposed grating). (b) Input of object B (one-shot superimposed grating). (a1) and (a4) gt (absolute phase) from the left and right viewpoints, respectively. (a2) and (a5) Network output (absolute phase) of DL_DPSL from left and right projector viewpoint separately. (a3) and (a6) Network output (absolute phase) of DL_SPSL from left and right projector viewpoint separately. The naming of (b1)–(b6) is consistent with (a1)–(a6).
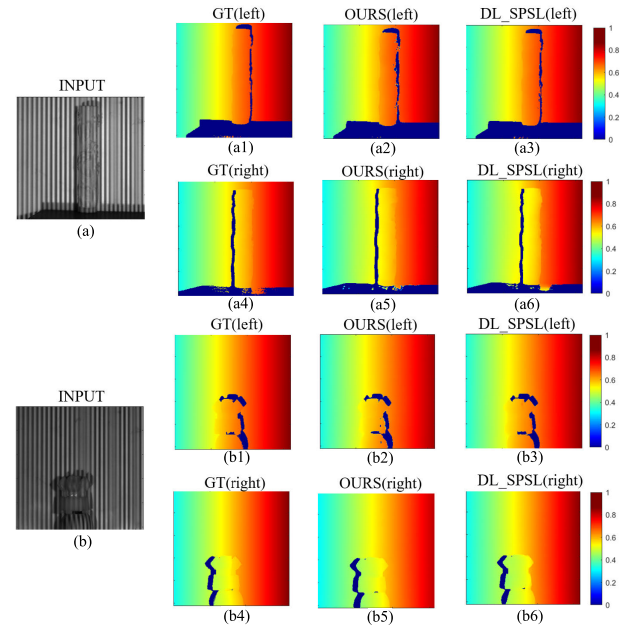

Fig. 7. Generalization ability description of DL_DPSL (our) and DL_SPSL system's measurement performance. Taking two industrial parts that have never been seen by the network as the example. (a) and (b) Dataset's input (one-shot superimposed shifting grating) of object C and D, respectively. (a1) and (a4) gt (absolute phase) in two views (left and right), respectively. (a2) and (a5) DL_DPSL's network output (absolute phase) of left and right projector's view separately. (a3) and (a6) DL_SPSL's network output (absolute phase) of left and right projector's view separately. The naming of (b1)–(b6) is consistent with (a1)–(a6).

training settings, the initial learning rate and the epoch of four conditions are 0.01 and 300 separately.

From Table I, we can see that the absolute phase prediction effect of our scheme on the simulation dataset is better than the result of individual projection on the whole. It is verified that the scheme based on deep learning network to effectively extract the projection phases on both sides from the single-frame superimposed phase shift grating is feasible. Second, ResUNet's prediction effect and prosperity capabilities are better than ResNet18, UNet, and HRnet. All in all, DL_DPSL can reduce measurement time by 50% while improving overall accuracy compared to DL_SPSL based on the ResUNet network for dual-view absolute phase measurements. Simulation system's impressive performance motivated us to create a real system to further verification.

V. ESTABLISHMENT AND VERIFICATION OF THE REAL SYSTEM

The schematic of experiment device of real DL_DPSL system is shown in Fig. 1(a). The projector 1 is the Texas Instrument pro6500-rgb-235, with a resolution of 1920×1080 and a refresh rate of 100 fps at 8 bit. The projector 2 adopts TengJu Technology version III, with a resolution of 1280×720 and a refresh rate of 100 fps at 8 bit. The camera is the Balser aca 2040-120 um grayscale camera, with a resolution of 2048×1536 and a frame rate of 100 fps. In the following reconstruction task, we adopted unidirectional structured light calibration method for system calibration [46]. The triggering capabilities of the projector and camera guaranteed synchronous projection of the projector and sampling of the camera. Similarly with simulated system,

TABLE II
OVERALL PHASE ERROR EVALUATION OF THE REAL TESTING DATASET

Method	Projection view	MAE(10e-3) ↓		RMSE(10e-2) ↓		SD(10e-3) ↓	
		DL_SPSL	DL_DPSL(ours)	DL_SPSL	DL_DPSL(ours)	DL_SPSL	DL_DPSL(ours)
HRnet	left	11.7	13.4	5.3	6.2	8.2	10.7
	right	53.9	7.9	25.0	3.8	411	5.0
ResNet18	left	14.4	18.7	6.0	7.4	10.6	12.7
	right	14.2	15.8	4.0	4.9	10.6	12.7
UNet	left	19.0	23.0	6.4	7.8	14.0	16.1
	right	18.4	19.5	4.6	5.2	12.8	13.8
ResUNet	left	8.2	**12.2**	5.3	**6.7**	6.4	**10.14**
	right	7.3	**7.7**	3.1	**3.9**	6.93	**6.54**

we also compared the prediction performance of ResUNet with UNet, ResNet18, and HRnet on real datasets.

A. Introduction of the Real Dataset

Based on the setup in Fig. 1 and the method in Section II, we construct the real dataset by measuring abundant real objects. In the real dataset, there are ∼900 image pairs in the training dataset and ∼80 image pairs in the testing dataset with the size of 640 × 512 pixels. It contains more than 200 plaster statues, wooden objects, and polystyrene artifacts, including basic geometric shapes (spheres, triangle pyramids, prisms), abundant human statues, and animal statues in combination. The volume of the measured objects is within 10 × 10 × 10 cm. In the DL_DPSL system, the input is the one-shot superimposed grating image obtained from simultaneous projection by two projectors. We choose the grating with a fringe frequency of 64 and an initial phase of $4\pi/3$ as the input of the real dataset. In the case of the left view, DL_SPSL's input is the left projection grating when two projectors are projected sequentially. The gt of DL_SPSL is the same as that of the DL_DPSL system. Taking two representative real objects A and B as the example, Fig. 6 shows the input and gt of the real dataset, along with the output of ResUNet under DL_DPSL and DL_SPSL system. Fig. 6(a) and (b) show the input (one-shot superimposed grating) of objects A and B. Fig. 6(a1), (a4), (b1), and (b4) correspond to the left and right view gt (absolute phase) for objects A and B, respectively. Fig. 6(a2) and (b2) illustrate the network output (absolute phase) of left projector's view for objects A and B in DL_DPSL, respectively. Fig. 6(a5) and (b5) illustrate the network output (absolute phase) of right projector's view respectively for objects A and B in DL_DPSL. Fig. 6(a3) and (b3) illustrate the network output (absolute phase) of left projector's view for objects A and B in DL_SPSL, respectively. Fig. 6(a6) and (b6) illustrate the network output (absolute phase) of right projector's view for objects A and B in DL_SPSL, respectively. Fig. 7 shows the generalization ability of the system. Objects C and D are 3-D printed industrial parts made of polystyrene, where C is an embossed cylinder carved with dragon patterns and D is a special-shaped structural workpiece. More importantly, the network has never encountered objects C and D, which are used to assess and validate the system's generalization capability. Fig. 7(a) and (b) show the input of the dataset, specifically the one-shot superimposed grating, for objects C and D, respectively. Fig. 7(a1), (a4), (b1), and (b4) correspond to the gt (absolute phase) observed from left and right viewpoints, for objects C and D, respectively. Fig. 7(a2) and (b2) showcase the network output (absolute phase) from the viewpoint of the left projector for objects C and D in DL_DPSL, respectively. Fig. 7(a5) and (b5) illustrate the network output (absolute phase) from the viewpoint of the right projector for objects C and D in DL_DPSL, respectively. Fig. 7(a3) and (b3) illustrate the network output (absolute phase) from the viewpoint of the left projector for objects C and D in DL_SPSL, respectively. Fig. 7(a6) and (b6) illustrate the network output (absolute phase) from the viewpoint of the right projector for objects C and D in DL_SPSL, respectively.

We assess the measurement accuracy of our DL_DPSL system in terms of the absolute phase and the correctness of the reconstructed 3-D point cloud. Additionally, we compare the phase measurement errors between our DL_DPSL system and the existing DL_SPSL system using the real object's testing dataset.

B. Evaluation of Absolute Phase Prediction Accuracy

In our comparative analysis, we employ an advanced absolute phase recovery solution called deep-learning-driven single perspective structured light 3-D measurement (DL_SPSL). It currently holds the distinction of being the fastest unwrapping solution available. The DL_SPSL is trained to solve the one-side absolute phase when two projectors are projected in succession. In this section, we compare the absolute phase prediction capabilities of ResUNet with UNet, ResNet18 and HRnet based on real datasets. As shown in Table II, the prediction effect of ResUNet is much better than the other three networks. Besides, in the absolute phase analysis of the left and right viewing angles, the DL_DPSL (ours) and DL_SPSL schemes have generally approximative values in terms of MAEloss, RMSEloss, and SD indicators. Note that the data in Table II is normalized phase and it needs to be multiplied by 400 to restore the real phase value. In the training settings, when DL_SPSL (left), DL_DPSL (left), DL_SPSL (right), and DL_DPSL (right) achieved the best convergence effect, the batch size, the initial learning rate and the epoch of four situations is 8, 0.01, and 300 separately. So next
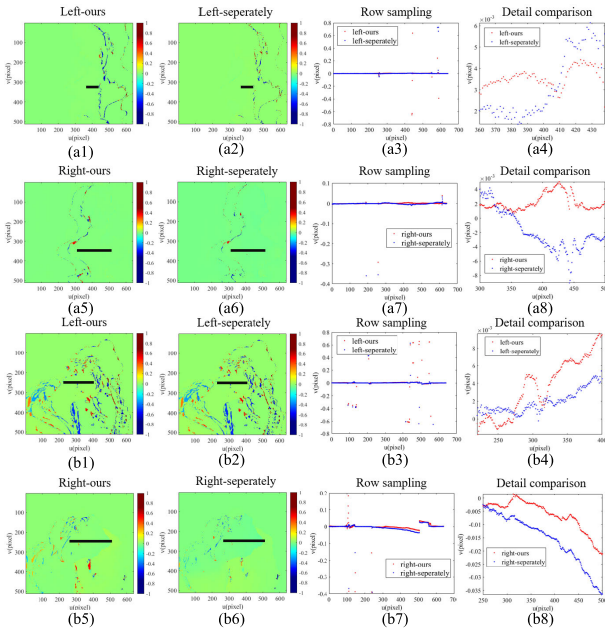
Fig. 8. Error analysis of dual-side absolute phases of two representative objects in the testing dataset of DL_DPSL and DL_SPSL system. (a1) Left-view phase error diagram of DL_DPSL system. (a2) Left-view phase error diagram of DL_SPSL system. (a3) Row sampling result where the black line is located in (a1) and (a2) of DL_DPSL and DL_SPSL system. The blue and red dots represent DL_SPSL and DL_DPSL separately. (a4) Detail extraction of (a3). (a5) Right-view phase error diagram of DL_DPSL system. (a6) Right-view phase error diagram of DL_SPSL system. (a7) Row sampling result where the black line is located in (a5) and (a6) of DL_DPSL and DL_SPSL system. The blue and red dots represent DL_SPSL and DL_DPSL separately. (a8) Detail extraction of (a7). The naming of the (b1)–(b8) is consistent with (a1)–(a8).

we further compare the ResUNet prediction results in our scheme and the separate projection scheme. Fig. 8 illustrates the overall and local sampling prediction phase error for both our DL_DPSL system and DL_SPSL system. The red and blue point depicts the error results of DL_DPSL (our) and DL_SPSL system separately. As for object A, Fig. 8(a1) and (a5) show the phase error of DL_DPSL (our) system in left and right projector's view, respectively. Fig. 8(a2) and (a6) show the phase error of DL_SPSL system in left and right projector's view, respectively. Fig. 8(a3) shows the result of row sampling with the row aligning with the black line observed in Fig. 8(a1) and (a2). And Fig. 8(a4) is the detail comparison of Fig. 8(a3). For object B, the error description and image naming convention follow the same pattern as that of object A. For instance, Fig. 8(b1)–(b8) correspond to the respective sub-figures. Fig. 9 shows the phase error for two representative objects that the network has never seen. Objects C and D are both 3-D printed structural components with unique shapes made of polystyrene. The error description and image naming of objects C and D are consistent with that of object A, such as Fig. 9(a1)–(a8). Figs. 8 and 9 describe the prediction results of phase accuracy error intuitively in color depth between DL_DPSL and DL_SPSL. From Figs. 8(a1), (a2), (a5), (a6), (b1), (b2), (b5), and (b6) and 9(a1), (a2), (a5), (a6), (b1), (b2), (b5), and (b6), it can be seen that at the edge and place where the height value has a huge change, the color of the measurement error of the DL_DPSL system is deeper than the result of the DL_SPSL system, which means that
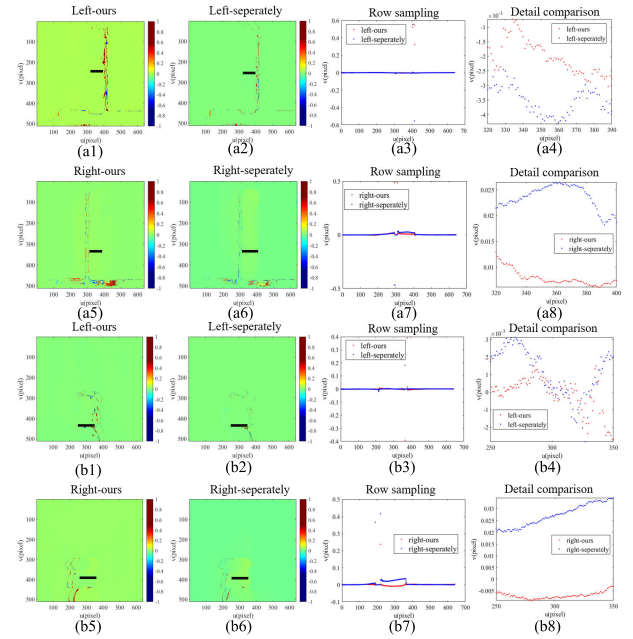
Fig. 9. Generalization ability test of 3-D point cloud generation between DL_DPSL and DL_SPSL system. Analysis the dual-side absolute phase error of two objects which have never been seen by the network. Objects C and D are distinct 3-D-printed structural components composed of polystyrene. (a1) Left-view phase error diagram of DL_DPSL system. (a2) Left-view phase error diagram of DL_SPSL system. (a3) Row sampling result where the black line is located in (a1) and (a2) of DL_DPSL and DL_SPSL system. The blue and red dots represent DL_SPSL and DL_DPSL separately. (a4) Detail extraction of (a3). (a5) Right-view phase error diagram of DL_DPSL system. (a6) Right-view phase error diagram of DL_SPSL system. (a7) Row sampling result where the black line is located in (a5) and (a6) of DL_DPSL and DL_SPSL system. The blue and red dots represent DL_SPSL and DL_DPSL separately. (a8) Detail extraction of (a7). The naming of the (b1)–(b8) is consistent with (a1)–(a8).
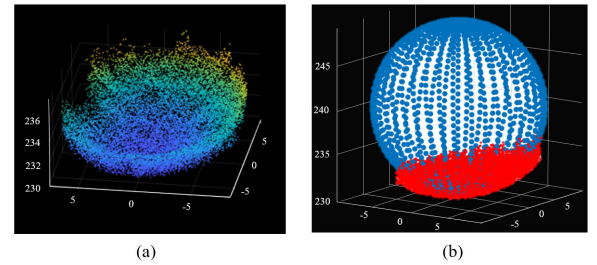

Fig. 10. Point cloud and radius fitting effect of the standard sphere. (a) Standard sphere's 3-D point cloud predicted by DL_DPSL system. (b) Radius fitting result optimized by the Levenberg–Marquardt algorithm.

the prediction accuracy of the DL_DPSL system is weaker than that of the DL_SPSL system. For the error comparison of detail sampling, Figs. 8(a3), (a4), (a7), (a8), (b3), (b4), (b7), and (b8) and 9(a3), (a4), (a7), (a8), (b3), (b4), (b7), and (b8) show clear that the DL_DPSL system performs better in measurement accuracy at the location where the measured object's height is essentially continuous.

Taken together, first, the performance of ResUNet is better than that of UNet, HRnet and ResNet18 in terms of absolute phase prediction analysis, based on the real dataset. Second, in terms of superimposed phase extraction capability, the prediction error of the simultaneous projection strategy is close to that of single projection. Besides, at the continuous positions of the measured object, the superimposed phase shift grating has a higher precision effect because it contains
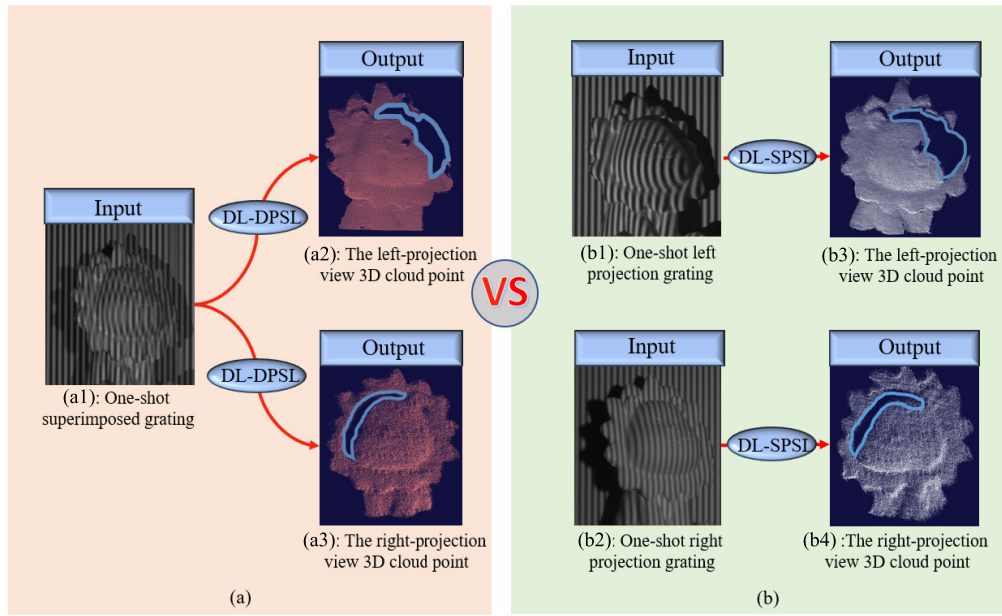
Fig. 11. Comparison of 3-D point cloud reconstruction effects of complex surface objects between DL_DPSL system and DL_SPSL system. In (a), (a1) represents the one-shot input image, (a2) and (a3), respectively, represent the left-view and right-view point cloud output reconstructed by the DL_DPSL system. In (b), (b1) and (b2) show the left-view and right-view input separately, (b3) and (b4) respectively represent the left-view and right-view point cloud reconstructed by the DL_SPSL system.

TABLE III
DURATION OF THREE TECHNIQUE ABSOLUTE PHASE ACQUISITION

Method	Projection time (ms) ↓	Unwrapping algorithm ↓ time (ms)	Totol duration ↓ (ms)
Six-figures separation method	12 fps×10 ms	1900	2020
DL_SPSL	2 fps×10 ms	8	28
DL_DPSL-ours	**1 fps×10 ms**	**8**	**18**

more phase information. Furthermore, next we perform a 3-D reconstruction of the bilateral absolute phase predicted by the ResUNet network.

C. Evaluation of 3-D Reconstruction Accuracy

To gauge the measurement precision of the 3-D point cloud generated by our proposed DL_DPSL system, we conducted eight reconstructions of a standard sphere utilizing the testing dataset. The true radius of the standard sphere is measured to be 10.009 mm. From these reconstructions, we calculated the average errors and SD in the reconstructed radius of the standard sphere. The average errors of the reconstructed standard sphere radius are 0.325 mm (right view) and 0.319 mm (left view), respectively. The SDs are 0.234 mm (right view) and 0.156 mm (left view), respectively. Fig. 10 shows the 3-D point cloud and radius fitting effect of the standard sphere. The 3-D reconstruction effect of standard sphere shown in Fig. 10(a). The fitting result optimized by Levenberg–Marquardt algorithm of the sphere's radius is depicted in Fig. 10(b).

We qualitatively compare the 3-D measurement effect of the DL_DPSL with the (DL_SPSL) system on objects having complicated contours in addition to the standard sphere,

as shown in Fig. 11. In Fig. 11(a1)–(a3) are the one-shot input (superimposed grating), left and right view 3-D point clouds reconstructed by the DL_DPSL system, respectively. Fig. 11(b1) and (b2) show the left-view and right-view input separately, Fig. 11(b3) and (b4) are the left and right view-points clouds reconstructed by the DL_SPSL system. The point clouds in Fig. 11(a2) and (a3) along with their respective point clouds in Fig. 11(b3) and (b4) perform nearly identically in terms of the total accuracy performance. The parts circled by the blue curve in Fig. 11(a2), (a3), (b3), and (b4) depict the missing point cloud due to the convexity of the object surface and the viewing angle. The 3-D point clouds from two different views can balance one another out, and enhance the point cloud of the darkened area, and broaden the measurement range.

D. Evaluation of Absolute Phase Acquisition Speed

In the simulated dataset, the test time of the one image (with an image size of 256 × 256 pixels) is ∼1.7 ms. In the real dataset, the test time of the one image (with an image size of 512 × 640 pixels) is ∼8 ms. Table III shows the comparison result among DL_DPSL system (ours) with DL_SPSL system (existing deep-learning scheme) and six-figures separation method [21] (traditional separation construction scheme). In view of the time consumption is consistent in the calibration and reconstruction stages by the above three methods during the 3-D measurement process. We just assess the time variation during the absolute phase acquisition. Taking Texas Instrument pro_6500 with 100 fps at 8-bit as an example, as for the projection time, DL_SPSL and our DL_DPSL method needs 120, 20, and 10 ms, respectively. The projection time are improved by 50% and 92% compared with DL_SPSL system and six-figures separation method, respectively. For the time consumption of unwrapping algorithm, it is 1900, 8, 8 ms,

respectively. Another argument is that for the six-figures separation method's unwrapping step, we use the 8-core CPU model 11th Gen Intel[1] Core[2] i7-1165G7 at 2.80 and 1.69 GHz.

In conclusion, because there is no need to wait for two projectors to project consecutively, the DL_DPSL system cuts projection time in half compared to the currently quickest absolute phase reconstruction method (DL_SPSL). The two perspectives' 3-D point clouds of the measured items can be produced using an one-shot superimposed phase shift grating image. Moreover, from the verification results of the simulation dataset and the details of the continuous height of the measured objects in the real dataset, the DL_DPSL system has better accuracy and stability. The precision and stability of 3-D reconstruction for the phase recovery of continuous surfaces are all improved since the superimposed grating image provides more effective phase information. The DL_DPSL system can reconstruct 3-D point clouds from two perspectives, effectively compensating for the lack of point clouds caused by occlusion.

VI. DISCUSSION AND CONCLUSION

Deep learning-driven structured light 3-D measurement technology stands as the fastest 3-D measurement method in FPP area currently. Its predictive capabilities in accuracy rival those of conventional methods without compromise. Nevertheless, cumbersome aspects persist in the implementation of large-scale end-to-end systems that two projectors associated with one gray camera are still required to project sequentially. To overcome the aforementioned challenges, we first propose a ResUNet-based dual-projector single-camera structured light 3-D measurement system named DL_DPSL. One key feature of this novel system is that two projectors can project simultaneously, and we can obtain dual-side 3-D point from just one superimposed phase-shifting fringe image. We also constructed the simulation and real system and corresponding simulated (∼3300 sets) and real dataset (∼1000 sets), respectively, for experimental verification. Experiments based on simulated datasets and real datasets have shown that the DL_DPSL method (our) has higher accuracy and stability in network prediction where the height of the object is relatively continuous. In comparison to the DL_SPSL method, which is currently the fastest end-to-end phase calculation method, and shortens the measurement time by 50%. Additionally, the simulation system can directly extract dataset from the grating image to the depth image, which may significantly cut labor costs and budget expenses and support the development of deep learning technology in FPP.

Importantly, it must be reiterated that the accuracy of DL_DPSL has not yet surpassed the level achieved by traditional multifrequency PSP. Various avenues exist for enhancing system accuracy. Taking three methods as examples, first, DL_DPSL only uses end-to-end phase prediction, and there are many more accurate methodologies, such as molecule prediction or phase unwrapping order prediction; Second, although the dataset constructed in this article is richer than the existing

public datasets, the number is still limited. Besides, large differences among objects in real dataset limit the predictive power of networks; Third, at present, our system only uses one loss function, and augmenting the system with multiple loss functions will elevate precision.

In summary, the developed DL_DPSL simulation and actual system, as well as the associated datasets, break the existing solution paradigm of dual-projector structured light 3-D measurement technology (SLM) system. The projecting time is halved while maintaining nearly unaltered measurement precision. This advancement offers a new solution for the industrial application of fast and large-scale SLM. This technology presents intriguing prospects for future extensions, such as predicting dual-side depth maps from one single superimposed phase-shifting grating and expanding simultaneous projection to multiple projectors, etc.

ACKNOWLEDGMENT

REFERENCES

[1] S. S. Gorthi and P. Rastogi, "Fringe projection techniques: Whither we are?" *Opt. Lasers Eng.*, vol. 48, no. 2, pp. 133–140, Feb. 2010.

[2] J. Geng, "Structured-light 3D surface imaging: A tutorial," *Adv. Opt. Photon.*, vol. 3, no. 2, pp. 128–160, 2011.

[3] 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.*, vol. 85, pp. 84–103, Oct. 2016.

[4] J. Qian et al., "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement," *APL Photon.*, vol. 5, no. 4, Apr. 2020.

[5] H. Yu, Y. Lan, J. Xu, D. An, and H. Lee, "Large-scale L^0-norm and L^1-norm 2-D phase unwrapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4712–4728, Aug. 2017.

[6] R. G. Waghmare, P. R. Sukumar, G. S. Subrahmanyam, R. K. Singh, and D. Mishra, "Particle-filter-based phase estimation in digital holographic interferometry," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 33, no. 3, pp. 326–332, 2016.

[7] J. Dong et al., "Simultaneous phase unwrapping and removal of chemical shift (SPURS) using graph cuts: Application in quantitative susceptibility mapping," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 531–540, Feb. 2015.

[8] B. Shi, Z. Ma, J. Liu, X. Ni, W. Xiao, and H. Liu, "Shadow extraction method based on multi-information fusion and discrete wavelet transform," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[9] W. Yin et al., "High-speed three-dimensional shape measurement using geometry-constraint-based number-theoretical phase unwrapping," *Opt. Lasers Eng.*, vol. 115, pp. 21–31, Apr. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143816618311266

[10] Y. Jin, Y. Chang, J. Wang, M. Li, L. Ren, and Y. Chen, "The measurement method for the size of the hole on the part surface based on grating image processing," *IEEE Access*, vol. 8, pp. 29159–29168, 2020.

[11] M. Zhong, X. Hu, F. Chen, C. Xiao, D. Peng, and S. Zhang, "Autofocusing method for a digital fringe projection system with dual projectors," *Opt. Exp.*, vol. 28, no. 9, pp. 12609–12620, 2020.

[12] D. Xiao, N. Sapermsap, Y. Chen, and D. D. U. Li, "Deep learning enhanced fast fluorescence lifetime imaging with a few photons," *Optica*, vol. 10, no. 7, pp. 944–951, Jul. 2023. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-10-7-944

[13] W. Jeon, W. Jeong, K. Son, and H. Yang, "Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks," *Opt. Lett.*, vol. 43, no. 17, pp. 4240–4243, 2018.

[14] H. Yu, D. Zheng, J. Fu, Y. Zhang, C. Zuo, and J. Han, "Deep learning-based fringe modulation-enhancing method for accurate fringe projection profilometry," *Opt. Exp.*, vol. 28, no. 15, pp. 21692–21703, Jul. 2020. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-28-15-21692

---

[1] Registered trademark.
[2] Trademarked.

[15] J. Qian et al., "Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry," *Opt. Lett.*, vol. 45, no. 7, pp. 1842–1845, 2020.

[16] S. Feng, C. Zuo, W. Yin, G. Gu, and Q. Chen, "Micro deep learning profilometry for high-speed 3D surface imaging," *Opt. Lasers Eng.*, vol. 121, pp. 416–427, Oct. 2019.

[17] G. E. Spoorthi, S. Gorthi, and R. K. S. S. Gorthi, "PhaseNet: A deep convolutional neural network for two-dimensional phase unwrapping," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 54–58, Jan. 2019.

[18] E. Aguénounon, J. T. Smith, M. Al-Taher, M. Diana, X. Intes, and S. Gioux, "Real-time, wide-field and high-quality single snapshot imaging of optical properties with profile correction using deep learning," *Biomed. Opt. Exp.*, vol. 11, no. 10, pp. 5701–5716, 2020.

[19] H. Nguyen, Y. Wang, and Z. Wang, "Single-shot 3D shape reconstruction using structured light and deep convolutional neural networks," *Sensors*, vol. 20, no. 13, p. 3718, Jul. 2020.

[20] K. Wang et al., "Deep learning wavefront sensing and aberration correction in atmospheric turbulence," *PhotoniX*, vol. 2, no. 1, pp. 1–11, Dec. 2021.

[21] Y. Li, X. Qu, F. Zhang, and Y. Zhang, "Separation method of superimposed gratings in double-projector structured-light vision 3D measurement system," *Opt. Commun.*, vol. 456, Feb. 2020, Art. no. 124676.

[22] Y. Zhang, X. Qu, Y. Li, and F. Zhang, "A separation method of superimposed gratings in double-projector fringe projection profilometry using a color camera," *Appl. Sci.*, vol. 11, no. 3, p. 890, Jan. 2021.

[23] Y. Yu, D. L. Lau, M. P. Ruffner, and K. Liu, "Dual-projector structured light 3D shape measurement," *Appl. Opt.*, vol. 59, no. 4, pp. 964–974, 2020.

[24] J. Wang, C. Zhang, W. Zhu, Z. Zhang, Z. Xiong, and P. A. Chou, "3D scene reconstruction by multiple structured-light based commodity depth cameras," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 5429–5432.

[25] C. Je, K. H. Lee, and S. W. Lee, "Multi-projector color structured-light vision," *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1046–1058, Oct. 2013.

[26] S. Zhang, "Absolute phase retrieval methods for digital fringe projection profilometry: A review," *Opt. Lasers Eng.*, vol. 107, pp. 28–37, Aug. 2018.

[27] H. Zhang et al., "Image prediction for limited-angle tomography via deep learning with convolutional neural network," 2016, *arXiv:1607.08707*.

[28] X. Hu, M. A. Naiel, A. Wong, M. Lamm, and P. Fieguth, "RUNet: A robust UNet architecture for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019.

[29] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Munich, Germany: Springer, 2015, pp. 234–241.

[31] T. Falk et al., "U-net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, Jan. 2019.

[32] M. Wan, L. Kong, and X. Peng, "Single-shot three-dimensional measurement by fringe analysis network," *Photonics*, vol. 10, no. 4, p. 417, 2023.

[33] S. Feng et al., "Fringe pattern analysis using deep learning," *Adv. Photon.*, vol. 1, no. 2, p. 1, Feb. 2019, Art. no. 025001, doi: 10.1117/1.ap.1.2.025001.

[34] W. Huang, X. Mei, Y. Wang, Z. Fan, C. Chen, and G. Jiang, "Two-dimensional phase unwrapping by a high-resolution deep learning network," *Measurement*, vol. 200, Aug. 2022, Art. no. 111566. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224122007813

[35] K. Wang, Y. Li, Q. Kemao, J. Di, and J. Zhao, "One-step robust deep learning phase unwrapping," *Opt. Exp.*, vol. 27, no. 10, pp. 15100–15115, 2019.

[36] H. Yu et al., "Dynamic 3-D measurement based on fringe-to-fringe transformation using deep learning," *Opt. Exp.*, vol. 28, no. 7, pp. 9405–9418, 2020.

[37] J. Song, K. Liu, A. Sowmya, and C. Sun, "Super-resolution phase retrieval network for single-pattern structured light 3D imaging," *IEEE Trans. Image Process.*, vol. 32, pp. 537–549, 2023.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[40] J. Wang and Y. Yang, "Phase extraction accuracy comparison based on multi-frequency phase-shifting method in fringe projection profilometry," *Measurement*, vol. 199, Aug. 2022, Art. no. 111525. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224122007461

[41] J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surfaces and lenses," *Appl. Opt.*, vol. 13, no. 11, pp. 2693–2703, 1974.

[42] V. Kumar, V. Tripathi, and B. Pant, "Content based fine-grained image retrieval using convolutional neural network," in *Proc. 7th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2020, pp. 1120–1125.

[43] M. Xu, C. Tang, Y. Shen, N. Hong, and Z. Lei, "PU-M-Net for phase unwrapping with speckle reduction and structure protection in ESPI," *Opt. Lasers Eng.*, vol. 151, Apr. 2022, Art. no. 106824.

[44] G. E. Spoorthi, R. K. S. S. Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach," *IEEE Trans. Image Process.*, vol. 29, pp. 4862–4872, 2020.

[45] J. Zhang and Q. Li, "EESANet: Edge-enhanced self-attention network for two-dimensional phase unwrapping," *Opt. Exp.*, vol. 30, no. 7, pp. 10470–10490, 2022.

[46] S. Zhang, "Flexible and high-accuracy method for uni-directional structured light system calibration," *Opt. Lasers Eng.*, vol. 143, Aug. 2021, Art. no. 106637.

Yiming Li is currently pursuing the Ph.D. degree in instrument science and technology, jointly cultivated, with Tsinghua University, Shenzhen, China, and the Peng Cheng Laboratory, Shenzhen.

Her research interests include intelligent precision 3-D measurement, structured light 3-D systems, and computer vision.

Zhuang Li was born in Henan, China. He is currently pursuing the M.S. degree with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

His research interests include 3-D measurement with structured light and computer vision.

Chaobo Zhang (Member, IEEE) received the master's and Ph.D. degrees in civil engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2017 and 2020, respectively.

He is currently a Post-Doctoral Fellow with the Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen, China. His research interests include intelligent sensing, big data analysis, deep learning, and computer vision for industrial applications.

Min Han is currently pursuing the Ph.D. degree with the Division of Advanced Manufacturing, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

His research interests include structured light 3-D reconstruction, micro-electro-mechanical systems (MEMS)-based fringe projection profilometry (FPP) systems, and computational imaging.

Xiaohao Wang received the B.S. and Ph.D. degrees in precision instrumentation and mechanics from Tsinghua University, Beijing, China, in 1994 and 1999, respectively.

He is currently a Professor with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His current research interests include instrumentation, measurement and control technology, equipment manufacturing, marine equipment, medical devices, micro-electro-mechanical systems (MEMS), sensors, optical engineering, etc.

Fengxiao Lei was born in Yongzhou, Hunan, China, in 2000. He received the B.S. degree in material processing engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2021. He is currently pursuing the M.S. degree with the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

He is involved in research on structured light 3-D reconstruction at the Tsinghua Shenzhen International Graduate School, Tsinghua University.

Weihua Gui (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in automatic control engineering from Central South University, Changsha, China, in 1976 and 1981, respectively.

From 1986 to 1988, he was a Visiting Scholar at the University of Duisburg-Essen, Duisburg, Germany. From 1991 to 2018, he was a Full Professor with the School of Information Science and Engineering, Central South University, where he has been a Full Professor with the School of Automation since 2019. Since 2013, he has been an Academician with the Chinese Academy of Engineering, Beijing, China. His main research interests include the modeling and optimal control of complex industrial processes, fault diagnoses, and distributed robust control.

Xiaojun Liang (Member, IEEE) received the B.S. degree in engineering mechanics and aerospace from Tsinghua University, Beijing, China, in 2012, and the Ph.D. degree in mechanical engineering and applied mechanics from the University of Pennsylvania, Philadelphia, PA, USA, in 2017.

From 2017 to 2020, he was a Senior Research and Development Engineer with the Intelligent Robot Division, JD Group, Beijing. Since 2021, he has been an Associate Research Fellow at the Peng Cheng Laboratory, Shenzhen, China. His research interests include mechanical modeling and analysis, industrial intelligent systems, and industrial artificial intelligence (AI) technology.

Xinghui Li (Member, IEEE) received the B.S. degree in mechanical engineering from Wuhan University, Wuhan, China, in 2008, the M.S. degree in mechatronic engineering from Xi'an Jiaotong University, Xi'an, China, in 2011, and the Ph.D. degree in nanomechanics from Tohoku University, Sendai, Japan, in 2014.

He is currently an Associate Professor with the Tsinghua Shenzhen International Graduate School, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, China. His current research interests include intelligent instruments, precision measurement, computer vision, and machine vision.