Phase-shifting guided deep learning for single-frame fringe analysis

AUTHOR ONE,[1] AUTHOR TWO,[2,*] AND AUTHOR THREE[2,3]

[1]*Peer Review, Publications Department, Optica Publishing Group, 2010 Massachusetts Avenue NW, Washington, DC 20036, USA*
[2]*Publications Department, Optica Publishing Group, 2010 Massachusetts Avenue NW, Washington, DC 20036, USA*
[3]*Currently with the Department of Electronic Journals, Optica Publishing Group, 2010 Massachusetts Avenue NW, Washington, DC 20036, USA*
[*]*opex@optica.org*

Abstract: Single-shot fringe projection profilometry (FPP) has been an ultimate goal to researchers. With the implementation of deep neural networks (DNNs), single-shot 3D reconstruction has been achieved through different approaches. For most DNN fringe analysis solutions, all captured fringes are involved in ground-truth making process but just few fringes serve as the input during network training. Namely, the information provided for DNN is compressed into single-shot input and corresponding ground-truth feature map. Therefore, these models may not have strong generalization and physically consistent predictions. To overcome this problem, we propose an universal framework which inputs all fringes in the phase-shifting sequence to 'teach' DNN models about the governing physical rules during training(using loss functions to give constraint about the phase and modulation), and only need single-frame input during testing. This is an overall optimized framework where DNN architecture is designed to maximize the utilization of the fringe sequence and the fringes are specially encoded to provide effective information for DNN. Experiments show that the proposed approach can achieve more robust and more precise fringe analysis than previous method, and can obtain high-quality 3D reconstructions from a single fringe image.

1. Introduction

Three-dimensional (3D) measurement technologies, as valuable tools for perceiving the real 3D world, provide data bases for the reconstruction of the geometric shape of objects and subsequent 3D modeling, detection, and recognition. Among them, optical non-contact methods have been widely applied to reverse engineering, industrial inspection, virtual reality, digital relics, medical examination, etc [1]. Fringe projection profilometry (FPP), as one of the most promising optical 3D measurement methods, attracts increasing attention due to its simple hardware configuration, high measurement accuracy and efficiency, and dense reconstruction point cloud. FPP projects structured light patterns (usually sinusoidal fringes, or fringe patterns which have spatially continuous phase distribution) on to the object using a digital projector, and captures the fringes deformed by object surface using a synchronously triggered camera. By applying proper fringe analysis algorithms to obtain the fringe's phase information, the 3D shape of the object can be reconstructed using the pre-calibrated geometric parameters of the FPP system. Usually, fringe analysis contains two stages, phase retrieval and phase unwrapping. During phase retrieval, the wrapped phase of each pixel is recovered from the captured intensity map. Because of mathematical feature of arctangent function, the phase value is wrapped between $-\pi$ and π . Phase unwrapping is to eliminate the periodic 2π jumps, ensuring the unambiguity when mapping the absolute phase to 3D coordinates. The common phase retrieval methods include transform-based methods and multi-frame fitting methods. The former extracts phase information in frequency domain using a meticulously designed band-pass filter, with regular domain transformation methods including Fourier Transform (FT) [2], Windowed Fourier Transform (WFT) [3, 4], and

Wavelet Transform (WT) [5]. Although transform-based methods need only a single frame to perform phase retrieval, the phase quality is relatively low especially when meeting with nonuniform ambient light, discontinuities surface and isolated objects [6–8]. Phase-shifting (PS) often use 3 or more fringes with different intensity map to carry out Least-square (LS) fitting algorithm, thus realizing pixel-wise phase calculation. The principle of PS ensures its high accuracy, and makes it quite robust to various measurement environment [9]. Phase unwrapping methods mainly include spatial phase unwrapping, temporal phase unwrapping (TPU) [10,11] and stereo phase unwrapping (SPU) [12,13]. Considering the hardware configuration, unwrapping algorithm stability and speed, TPU is still the common choice. However, both PS and TPU sacrifice measurement efficiency for more reliable 3D result, which severely restrict its application in dynamic scene. To apply FPP to moving objects measurement and simultaneously maintain the high-resolution result and measurement robustness, researchers have made great efforts to reduce the fringes needed for one single 3D reconstruction [14–17] and proposed some compensation algorithms as remedies [18–20]. However, these methods still fail to meet the ideal circumstance that the absolute phase could be recovered from a single-shot fringe pattern.

In recent years, deep learning has become a powerful tool for digging implicit pattern from massive data, which is promising in solving ill-posed questions. Feng et al. [21] first introduce deep neural network (DNN) into structured light fringe analysis. By learning the intermediate result in PS algorithm, they realized single frame phase retrieval. Another PS inspired method is to generate several phase-shifting fringe patterns from a single fringe using deep learning, and apply traditional PS algorithm to carry out phase retrieval [22,23]. However, these methods can only acquire wrapped phase and need extra fringes to perform phase unwrapping. Qian et al. [24] combined deep learning with SPU, successfully making DNN to recover fringe order from two cameras' view. To improve the phase unwrapping robustness, they also proposed a color fringe encoding strategy combined with TPU. Li et al. [25,26] proposed strategies to composite two or three fringes with different spatial frequency into a single pattern, thus theoretically containing all the information needed for single frame fringe analysis. Though the mixed pattern cannot be decoupled through transform-based method because of spectrum overlap, it can be automatically separated by data-driven DNN. However, the accuracy is not as excellent compared with single-frequency fringe analysis using DNN. The methods mentioned above as well as some other DNN-based methods [27–30], have a common feature that DL is only taken as a data-driven 'black box', lacking convincing interpretation and sufficient integration with prior knowledge. There have been some attempts on this problem. Fan et al. [31] first introduced unsupervised learning into fringe analysis. By virtually reprojecting fringe patterns onto the output depth-map, they construct a self-supervised constraint for label-free learning. However, the computer simulated dataset makes it not suitable for practical applications with complex surface reflectivity and ambient light. Yu et al. [32,33] constructed three physical constraints in DNNs' loss functions based on WFT and SPU, realizing robust phase unwrapping even in a small training dataset. However, the accuracy is lower than previous label-based DNN when the dataset contains sufficient scenes. Yin et al. [34] combined transform-based method with deep learning. This method shifts the learning domain from spatial to frequency, which achieves single frame fringe analysis with a light network. However, the learning process is completely guided by ground-truth, which still lacks guidance of physical laws or prior knowledge.

In this work, we present a phase-shifting guided deep learning for single-frame fringe analysis which is a universal solution for integrating prior knowledge into DL phase retrieval. Through weight-sharing strategy, we construct N same subnetwork (same structure and same parameter space) to analyze N step PS fringes separately, set constraints on N-group net predictions in loss functions according to the phase shifts, and keep only one single subnetwork in the architecture during testing. Eventually, all PS fringes in projection sequence are utilized in the training stage and single-frame fringe analysis in the testing stage is simultaneously ensured. To better guide

the neural network's convergence, 'frequency-proportioned phase-shifting encoding strategy for dual-frequency composite fringes' (fpPSC) and 'optimal phase shifting rules' (OPSR) are proposed to generate an informative fringe projection sequence. Compared with label-based DL method, our method achieves impressive accuracy improvement both in phase retrieval and phase unwrapping, bringing high quality 3D measurement for dynamic complex surface.

The remainder of this paper is organized as follows. In Section **Principle**, fpPSC, OPSR and the DNN workflow are introduced successively. In Section **Experiments**, comparison results and analysis are presented. In Section **Discussion and Conclusion**, the possible explanation for DL phase retrieval is discussed to help readers better understand the mutual influences between fringe encoding and network architecture, and the extensive application prospect is given. The theory demonstration, formula deduction and other detailed information involved in the comprehensive optimized framework are provided in **Supplemental document**.

2. Principle

2.1. Dual-frequency phase-shifting fringe patterns encoding strategy

Since we use all the phase-shifting fringes instead of the first one in each projection sequence to optimize the neural network, the fringe patterns should be carefully designed to provide optimal guidance for DNN. In brief, the PS fringes should share the same wave mode and obey OPSR. Concrete demonstrations are as follows.

In previous work, dual-frequency PS fringes are defined similar to single frequency standard N-step phase-shifting fringes, which can be written as:

$$I_n^p = a^p + b_H^p \cos(2\pi f_H x^P + 2\pi n/N\ ) + b_L^p \cos(2\pi f_L x^p + 2m\pi n/N\ ) \tag{1}$$

where I_n^p is the intensity of a pixel in the projector, a^p is the background intensity, b_H^p and b_L^p mean the modulation of the high- and low-frequency fringe respectively, f_H and f_L are the corresponding fringe frequencies, n represents the phase shift steps ($n = 0, 1, 2, ..., N-1$), N is the total number of phase shift ($N \geq 5$), and m is an integer to ensure the unambiguity (see **Supplement 1**(Section S1)).

The corresponding captured image can be expressed as:

$$I_n^c = A^c + B_H^c \cos(\Phi_H + 2\pi n/N\ ) + B_L^c \cos(\Phi_L + 2m\pi n/N\ ) \tag{2}$$

where I_n^c is the intensity of a pixel in the camera, Φ_H and Φ_L represent the phase information of the high- and low-frequency fringe respectively. In this way, the phase-pair can be obtained by Eq. (3), which is the same as single frequency circumstances:

$$\varphi_H = \tan^{-1} \frac{\sum_{n=0}^{N-1} I_n^c \sin(2\pi n/N\ )}{\sum_{n=0}^{N-1} I_n^c \cos(2\pi n/N\ )}\ , \varphi_L = \tan^{-1} \frac{\sum_{n=0}^{N-1} I_n^c \sin(2m\pi n/N\ )}{\sum_{n=0}^{N-1} I_n^c \cos(2m\pi n/N\ )} \tag{3}$$

However, this 'standard phase-shifting encoding strategy for dual-frequency composite fringes' (stPSC) is not suitable in this work because each fringe pattern in the phase-shifting projection sequence has a different wave mode (Fig.1(b)). Intuitively, keeping the fringe wave mode unchanged during phase-shifting helps DNN extracts the underlying feature map more easily (see Section **Discussion and conclusion** for detailed analysis). Hence, we proposed a 'frequency-proportioned phase-shifting encoding strategy for dual-frequency composite fringes' (fpPSC) (Fig.1(a)) which can be written as:

$$I_n^p = a^p + b_H^p \cos(2\pi f_H x^p + k f_H \pi n/N\ ) + b_L^p \cos(2\pi f_L x^p + k f_L \pi n/N\ ) \tag{4}$$

where k is the 'PS-coefficient' which controls the total phase-shifting range.

Figure 1 shows a group of fpPSC and stPSC patterns with $N = 6$, $f_H = 5$(/1000pixel), $f_L = 2$(/1000pixel), $a^p = 0.509$, $b_H^p = 0.255$, $b_L^p = 0.235$ in order to cover the entire dynamic range of the projector (without loss of generality, here we assume the projector has a dynamic range of [0, 1]). For clear and concise illustration, we choose $k = 500$ for fpPSC, $m = 2$ for stPSC, and only present the first three fringe patterns (I_0^p,I_1^p,I_2^p) in the projection sequence. It can be seen clearly from Fig.1 that fpPSC patterns move as one while stPSC patterns move in a mess.

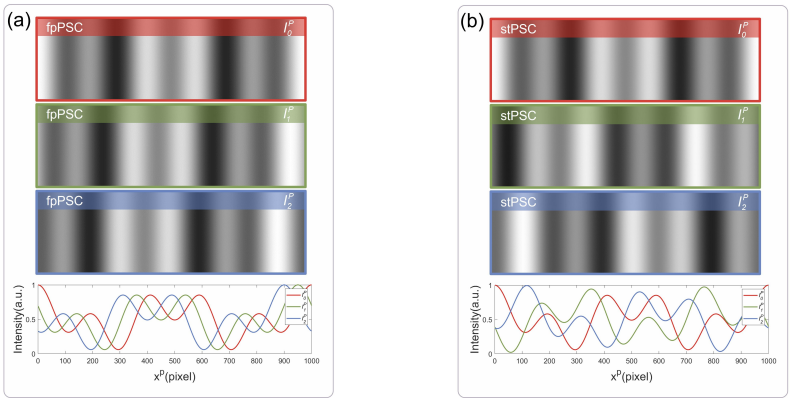

Fig. 1. Phase-shifting patterns of fpPSC and stPSC ($N = 6$).

Therefore, fpPSC expressed as Eq. (4) is chosen to encode the phase-shifting fringes for projection. The corresponding captured images can be expressed as:

$$I_n^c = A + B_H^c \cos(\Phi_H + \Delta_n) + B_L^c \cos(\Phi_L + \delta_n) \tag{5}$$

where $\Delta_n = k f_H \pi n/N$, $\delta_n = k f_L \pi n/N$, and the phase-pair can be obtained by LS fitting algorithm (see **Supplement 1**(Section S1) for details):

$$\varphi_H = \tan^{-1} \frac{p_{3,1}}{p_{2,1}} \text{ , } \varphi_L = \tan^{-1} \frac{p_{5,1}}{p_{4,1}} \tag{6}$$

where $p_{i,j}$ is the element of 5×1 column matrix P decided by I_n^c , Δ_n, and δ_n.

To obtain high-quality LS fitting result and offer DNN informative guidance, the phase shifts of both high- and low-frequency components should follow the optimal phase-shifting rules (OPSR): Regulate each phase shift $\Delta_n(\delta_n)$ into 2π period by mod operation and view these phase shifts as sampling points within a 2π range:

1) Total number of samples N should be as large as possible.

2) The overall sampling range should be as wide as possible.

3) The distribution of sampling points should be as even as possible.

Therefore, when generating specific fpPSC fringe projection sequence, f_H, f_L, N, and k should be determined properly to make the phase shift $\Delta_n(\delta_n)$ satisfy the OPSR. Detailed demonstration will be introduced in **Supplement 1**(Section S2).

2.2. Phase-shifting guided deep learning for single frame fringe analysis

In previous DL fringe analysis methods, single frequency standard N-step phase-shifting is often adopted together with TPU to make ground-truth (GT) for DNN. Targeting at single frame fringe

analysis, researchers tend to choose the first fringe pattern in each projection sequence as the DNN input, and use corresponding label as outputs' GT to guide the DNN convergence. We call that 'label-based method'. In this way, a large number of fringe patterns are wasted during network training process. To fully exploit the fringes captured during GT making process, we designed a weight-sharing network architecture to perform phase-shifting guided deep learning, which uses all the PS fringes as the network input during training and is capable of single-frame-input prediction during testing. The network architecture and overall flowchart (after DNN being properly trained) are shown in Fig.2 and Fig.3.

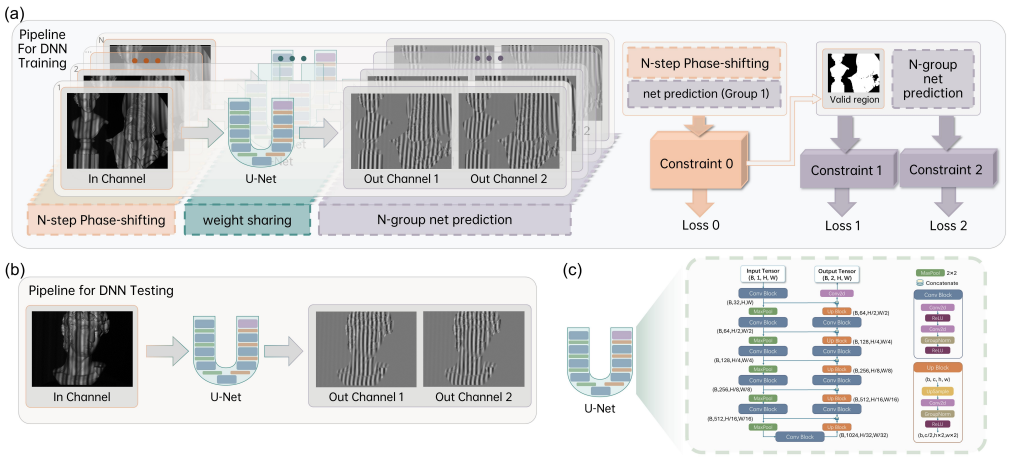

Fig. 2. DNN framework. (a) Pipeline for training stage; (b) Pipeline for testing stage; (c) Unet architecture.

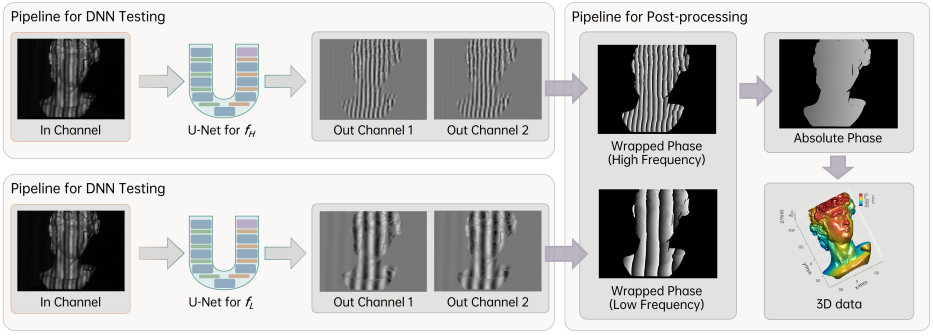

Fig. 3. Overall flowchart of the proposed approach to single-frame 3D shape measurement.

2.2.1. Training stage

To train a network for retrieving a certain component's phase (high- or low-frequency component of the composite fringe), the detailed procedure is as follows.

Step1: Generate fpPSC fringes based on Eq.(4). Referring to previous research [25], we choose $f_H = \frac{1}{19}(\text{pixel}^{-1})$ and $f_L = \frac{1}{51}(\text{pixel}^{-1})$, which has been proved effective for DL single-frame phase retrieval and unwrapping. Then we choose $N = 12$ and $k = 124$ to make the phase shifts satisfying OPSR (see **Supplement 1**(Section S2) for details). In order to cover the entire dynamic range of the projector [0,255], we set $a_p = 130$, $b_H^P = 65$, $b_L^P = 60$.

Step2: Project, capture and data preprocess. For each scene, we project 12 fpPSC fringes and capture the corresponding deformed fringe intensity maps. Then we carry out data normalization and file format conversion (see **Supplement 1**(Section S3.B) for details) before feeding them into the neural network.

Step3: Feed preprocessed data into a DNN architecture with 12 weight-sharing subnetworks and construct corresponding loss functions. Here we choose Unet as subnetwork backbone for its great balance between FLOPs and performance (see **Supplement 1**(Section S3.B) for Unet architecture). For each scene, 12 captured fringe patterns are fed into 12 weight-sharing Unet respectively, namely each Unet receives a single frame input. Then the 12 Unet with totally same parameter space analyze the 12 fringes in parallel and give 12-group predictions. We set the n-th Unet prediction as a two-channel tensor $y^n_{DL} = \left(M^n_{DL}, D^n_{DL}\right)$. The first channel gives the numerator term M^n_{DL} of arctangent function in LS algorithm (like Eq. (6)) and the second channel gives the corresponding denominator term D^n_{DL}. The reasons for this setting are as follows. For one thing, it bypasses the difficulty of predicting wrapped phase with periodic 2π jump and enables higher quality phase extraction compared with end-to-end method [30, 35]. For another, the definition of numerator and denominator term naturally ensures its capability for calculating wrapped phase φ and modulation B (see **Supplement 1**(Section S1) for details), which is convenient for the construction of the following constraints.

Taking high-frequency component as an example, noting group n-th prediction as $y^n_{DL\_H} = \left(M^n_{DL\_H}, D^n_{DL\_H}\right)$, there must exists three constraints.

Constraint1: Use group n-th prediction to calculate the phase value $\varphi^n_{DL\_H}$ of the first group (first fringe in each phase-shifting projection sequence) according to the phase shifts.

$$\varphi^n_{DL\_H} = tan^{-1}\frac{M^n_{DL\_H}}{D^n_{DL\_H}} - \frac{k f_H n\pi}{N} \tag{7}$$

No matter which index n is chosen, the obtained $\varphi^n_{DL\_H}$ should keep the same (relative position in a 2π period), namely:

$$\varphi^n_{DL\_H} \equiv \varphi^0_{DL\_H} \ (\mathrm{mod}\ 2\pi) \tag{8}$$

Constraint2: Use group n-th prediction to calculate the modulation $B^n_{DL\_H}$.

$$B^n_{DL\_H} = \left((M^n_{DL\_H})^2 + (D^n_{DL\_H})^2\right)^{0.5} \tag{9}$$

No matter which index n is chosen, the obtained $B^n_{DL\_H}$ should keep the same, namely:

$$B^n_{DL\_H} \equiv B^0_{DL\_H} \tag{10}$$

Constraint1&2 give out the relative relationship between different groups' output prediction, but is unable to construct the absolute relationship between input and output. This may lead to a result that the 12-group predictions follow the relative relationship well but cannot give the correct exact value. To eliminate the difficulty for DNN to find the absolute mapping rules itself, we need another constraint.

Contraint0: Calculate the first group's ($n = 0$) $y_{LS\_H} = (M_{LS\_H}, D_{LS\_H})$ using LS algorithm (see **Supplement 1**(Section S1) for details) to constraint the value of $\varphi^0_{DL\_H}$ and $B^0_{DL\_H}$, thus setting an absolute benchmark for the relative relationship in Constraint1&2, namely:

$$\varphi_{LS\_H} = tan^{-1}\frac{M_{LS\_H}}{D_{LS\_H}} = \varphi^0_{DL\_H} \tag{11}$$

$$B_{LS\_H} = \left((M_{LS\_H})^2 + (D_{LS\_H})^2\right)^{0.5} = B^0_{DL\_H} \tag{12}$$

According to the above Constraint0&1&2, we generate 3 modules (Fig.2(a)) to construct corresponding 3 loss functions (Loss0&1&2, see **Supplement 1**(Section S3.A) for details).

After proper training, we got a network that can extract phase of the high-frequency component in a composite fringe. Similarly, we change the f_H to f_L in Eq. (8), change $y_{LS\_H} = (M_{LS\_H}, D_{LS\_H})$ and all the corresponding value from high- to low-component, then we got another network responsible for low-frequency phase retrieval. In testing stage, two trained network responsible respectively for high- and low-frequency phase extraction are used.

2.2.2. Testing stage

During testing stage (Fig.3), we need just the first one of 12-step PS fringes as the network input. For retrieving the wrapped phase (both high- and low-frequency component of the composite fringe), absolute phase and 3D results, detailed procedure is as follows.

Step1: Generate 1 composite fringe, which is the first of the 12 fpPSC fringes in training stage.

Step2: Project, capture and data preprocess.

Step3: Feed preprocessed data into a DNN (trained for high-frequency component) with only 1 subnetwork, which is the first (or any one) of 12 weight-sharing subnetworks in training stage. Similarly, feed the data into another DNN (trained for low-frequency component). Then we got the numerator and denominator terms of the arctangent function for retrieving both high- and low-frequency wrapped phase φ_{DL_H} and φ_{DL_L}.

Step4: Use the DNN output to perform phase retrieval and temporal phase unwrapping. The wrapped phase-pair φ_{DL_H} and φ_{DL_L} can be expressed as:

$$\varphi_{DL\_H} = tan^{-1}\frac{M^0_{DL\_H}}{D^0_{DL\_H}}\ ,\ \varphi_{DL\_L} = tan^{-1}\frac{M^0_{DL\_L}}{D^0_{DL\_L}} \tag{13}$$

where $M^0_{DL\_H}$, $D^0_{DL\_H}$ is the output of DNN for high-frequency, $M^0_{DL\_L}$, $D^0_{DL\_L}$ is the output of DNN for low-frequency.

The absolute phase corresponding to the wrapped phase of two frequencies can be expressed as:

$$\begin{cases} \Phi_{DL\_H} = \varphi_{DL\_H} + 2\pi k_{DL\_H} \\ \Phi_{DL\_L} = \varphi_{DL\_L} + 2\pi k_{DL\_L} \end{cases} \tag{14}$$

where k_{DL_H} and k_{DL_L} are the integer fringe order, which can be obtained through number-theoretical approach (one of the TPU algorithm). With least common multiple $LCM(f_H^{-1}, f_L^{-1})$ exceeding the lateral resolution W of the projected pattern, every phase-pair $(\varphi_{DL\_H}, \varphi_{DL\_L})$ is unique within the lateral range. Therefore, we can establish lookup table (LUT) to determine the fringe order [10]:

$$(k_{DL\_H}, k_{DL\_L}) = LUT\ [Stair(x, y)] \tag{15}$$

where $Stair$ function can be expressed as:

$$Stair(x, y) = round(\frac{\varphi_{DL\_H}}{2\pi f_L} - \frac{\varphi_{DL\_L}}{2\pi f_H}) \tag{16}$$

After getting the fringe order, we can get the absolute phase by Eq. (20).

Step5: Map absolute phase to 3D coordinates. For better anti-noise performance, we choose the absolute phase of high-frequency component to perform 3D mapping [9]. By utilizing the pre-calibrated parameters of the FPP system, 3D information of the objects can be reconstructed.

In this Section, we present a training strategy that fully utilizes all PS fringes in training stage and keeps single frame fringe analysis ability in testing stage. By directly construct loss functions according to prior knowledge like Eq.(8) and Eq.(10), we forced neural networks to 'learn' the mathematical form of the output, instead of just give a label and leave it confused about what the output is used for.

3. Experiment

To test the performance of the proposed phase-shifting guided deep learning approach, we construct a monocular FPP system, which includes a LightCrafter 4500Pro (912 × 1140 resolution) and a Basler acA640-750 μm cameras (640 × 480 resolution). The size of the measuring field is about 240 mm × 200 mm. We collect 1000 scenes including continuous and isolated surfaces with different reflectivity distribution, and divide them into train set, validation set and test set (800:100:100). Fig.4 shows some typical scenes in the training datasets.

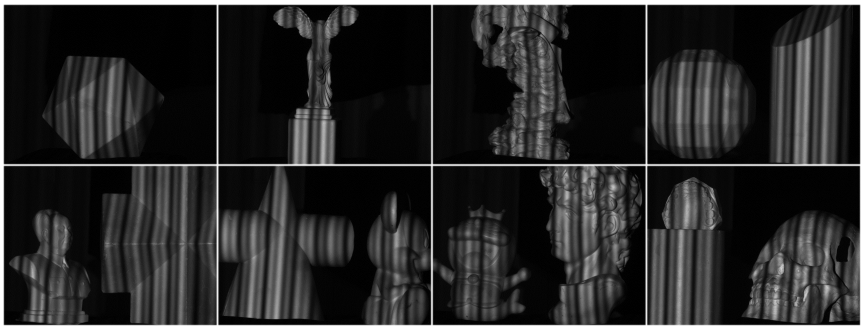

Fig. 4. Typical scenes in training datasets. E.g. single smooth surfaces, single complex surfaces and the different combinations of them forming isolate surfaces.

Each scene we captured 36 fringes including 24 single frequency fringes (12 high-frequency and 12 low-frequency fringes with standard PS, '12+12 PS' for short) and 12 proposed frequency-proportioned phase-shifting composite fringes ('12 fpPSC' for short). The former 24 fringes are used to obtain high quality phase map based on LS algorithm, which serve as the standard to perform error analysis in this section and the label of control group DL method. The latter 12 fringes are used as the training input of proposed method.

To be concrete, the control group method chooses the first frame of fpPSC fringe sequence as the input and outputs the corresponding numerator and denominator, using MSE loss function to guide the convergence. We call that label-based method. The proposed method uses 3 loss functions to guide the convergence, where y_{LS} (label) in Loss0 obtained by the latter '12 fpPSC' fringes has a bit lower quality. When calculating Mean Absolute Error (MAE) in the later error analysis, for convenience and uniform standard, we choose the high-quality label as GT (obtained by '12+12 PS') for evaluating both label-based method (control group method) and PS-guided method (proposed method). The constructed network is computed on Cloud Sever AutoDL using Intel(R) Xeon(R) Platinum 8358P CPU and an NVIDIA A40(48GB) GPU, with deep learning framework Pytorch (Facebook).

To verify the superiority of our approach over label-based method, we measured a hundred different scenes that were not seen by neural networks during training stage. We choose the most representative three scenes to show their visualized error map and present the error distribution of whole 100 test scenes (Fig.5).

For the first scene (#1), a David plaster, we can see that complex regions having disordered reflectivity distribution near ear, nose, hair, neck are measured with large errors, both in high- and low-frequency predictions. When using our method, the MAE of absolute phase is greatly reduced from 0.087 rad to 0.045 rad.

For the second scene (#2), a cylinder and a skull, we can see that the phase of smooth areas with gentle depth changes is retrieved accurately, while that of regions having steep depth changes are measured with large errors. When using our method, the MAE of absolute phase is reduced from 0.074 rad to 0.042 rad.

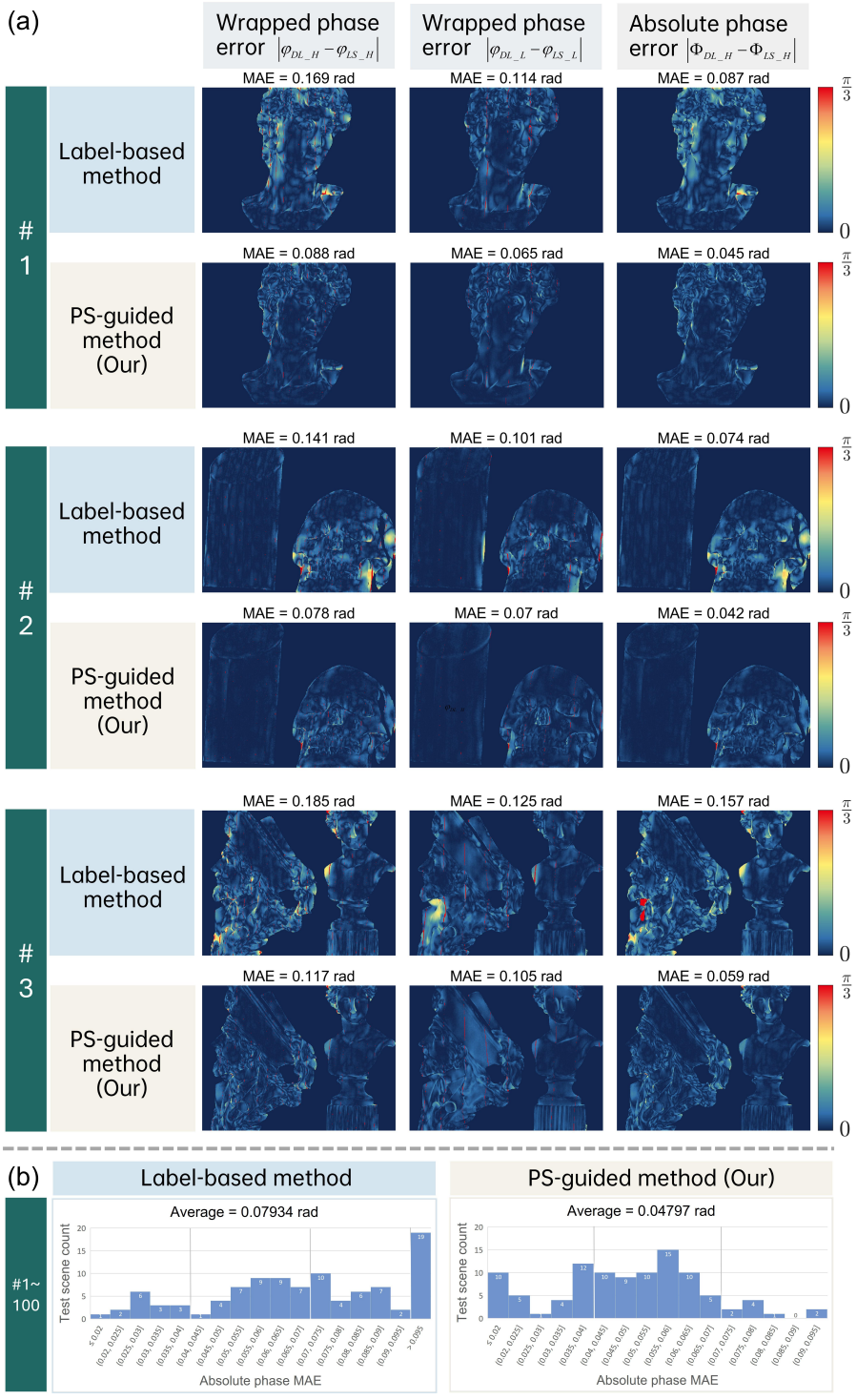

Fig. 5. Comparison between label-based method and PS-guided method. (a) Error maps. (b) Error distribution.

For the third scene (#3), a pirate and a goddess, we can see that the low-frequency wrapped phase error may extremely influence the phase unwrapping accuracy. In label-based method, the red region indicating large error in the absolute phase map is caused by the error in low-frequency wrapped phase. When using our method, the error is eliminated well, thus the MAE of absolute phase is sharply reduced from 0.157 rad to 0.059 rad.

It needs to be mentioned that, the reason why the absolute phase MAE is considerably smaller than the wrapped phase MAE is that jump area of the predicted wrapped phase and that of the GT wrapped phase do not completely coincide. The errors caused by these very slight misalignments (often only one pixel apart) will be automatically canceled after phase unwrapping.

Fig.5(b) shows the error distribution of the whole 100 test scenes. The horizontal axis divides absolute phase MAE into every 0.005 intervals, while the vertical axis shows the number of scenes within the corresponding intervals. In label-based method, there are 19 scenes MAE over 0.095 rad, while in our method, only 9 scenes MAE between 0.07rad and 0.095 rad with not a single scene over 0.095 rad. It can be seen clearly that, the error distribution has been effectively transferred from high-MAE part to low-MAE part. This strongly testifies the excellence of our PS-guided method.

Based on calibration parameters of the projector-camera FPP system, the 3D reconstruction results can be obtained through phase-height mapping. Fig.6 shows the comparison results of three methods: the 12+12 PS with TPU (the ground-truth generation method) (Fig.6(a,d,g)), label-based method (Fig.6 (b,e,h)), PS-guided method (Fig.6(c,f,i)). Our PS-guided method successfully achieves more precise and robust 3D results.

It can be seen clearly that, the error reduction in high-frequency wrapped phase leads to less absolute phase error around complex regions, bringing better surface detail in 3D reconstructions. Meanwhile, the error reduction in low-frequency wrapped phase leads to more robust phase unwrapping, bringing less huge mistakes in 3D results. Therefore, PS-guided method deserves to be applied in both high- and low-frequency phase retrieval.

To further verify the effectiveness of the proposed loss functions, we carry out the ablation study for high-frequency phase retrieval of fpPSC fringes. Fig.7 shows the loss curve in training stage (Fig.7(a)), the wrapped phase error map of one scene in testing stage (Fig.7(b)) and the wrapped phase error distribution of all scenes in testing stage (Fig.7(c)). Four colors are used to represent different loss combinations: orange (complete loss functions), blue (Loss0+Loss2), green (Loss0+Loss1), gray(Loss0). It needs to be mentioned that the loss data used to draw loss curves is not the backward loss, but the spatial part of Loss0. Therefore, the curves of different losses are able to be compared. In Fig.7(a), we can infer that both Loss1(the constraint for phase) and Loss2 (the constraint for modulation) have the positive effect, and Loss1 plays a relatively more important role. Corresponding results can be seen in Fig.7(b) (gradually descending MAE value) and Fig.7(c) (gradually left-shift MAE distribution).

4. Discussion and conclusion

In our opinion, phase-shifting based on least-square algorithm and deep learning are both fitting method for phase retrieval. The common target of them is to get rid of the ambient illumination and nonuniform surface reflectivity that hinder the extraction of phase distribution. The pixel-wise influence of the two types of interference factors appears in average intensity $A(x, y)$ and intensity modulation $B(x, y)$ of the captured intensity map $I(x, y)$ [9]. For LS algorithm, each scene is measured independently by projecting a series of fringes with different $I(x, y)$ through adjusting the phase. Supposing the ambient light and surface reflectivity stay unchanged during the projection sequence, the interference as well as the needed phase can be mathematically fitted through LS algorithm. However, when meeting with time-varying ambient light, or even dynamic/moving objects (which means changing surface reflectivity) within the projection sequence, phase-shifting algorithm is by no means available. Different

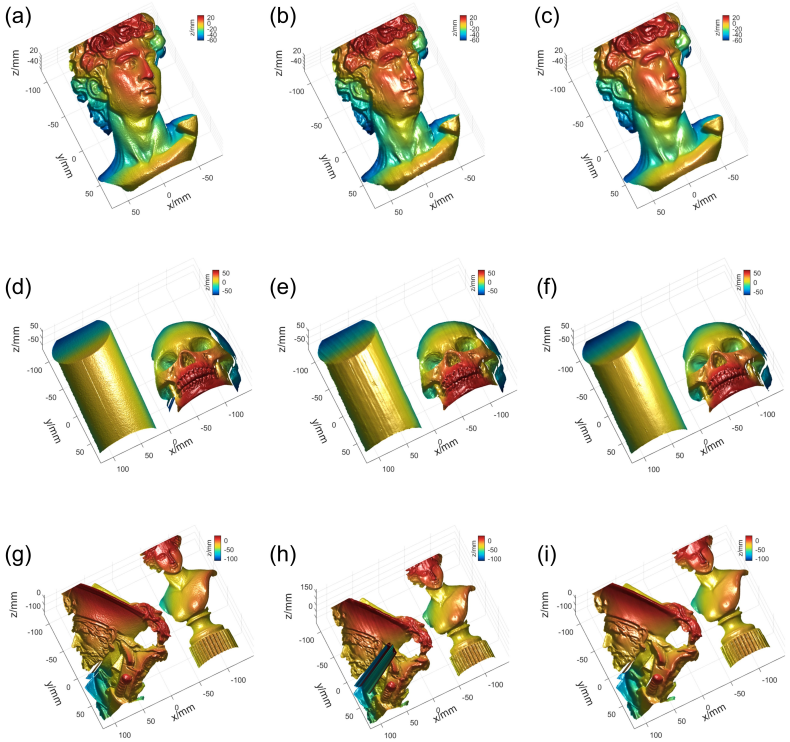

Fig. 6. 3D reconstruction results of three methods. (a,d,g) 3D results of GT (12 high-frequency and 12 low-frequency fringes with standard phase-shifting and number-theoretical phase unwrapping); (b,e,h) 3D results of label-based method; (c,f,i) 3D results of our method

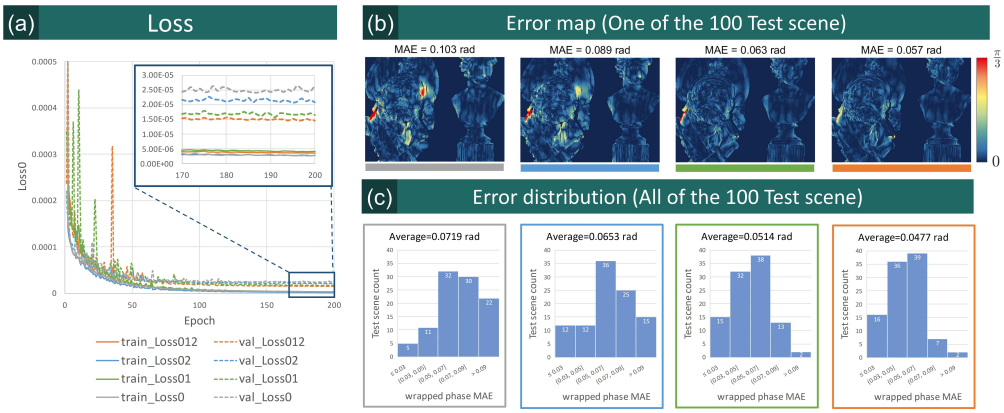

Fig. 7. Ablation study of 3 loss functions. (a) Training loss curves. (b) Error map of one scene. (c) Error distribution of all test scenes.

from PS repeatedly carrying out scene-independent calculation, deep learning method utilizes a large number of scenes to 'learn' the optimal mapping rules from the input intensity map to target feature map. Since some convolutional neural networks (like Unet) tend to concentrate on local information, the input intensity distribution within local receptive field can be classified into approximately limited types. Therefore, limited convolutional kernels can find an optimal non-linear combination that satisfies the vast majority of scenes in the train set. And once the convolutional kernels are capable of handling situations in train set, they can tackle with other situations in test set (same measurement environment, same local reflectivity distribution with those in train set). Concrete analysis is as follows.

Considering the components of captured intensity maps of different scenes, there are projection light with fixed wave mode, ambient light with stable distribution and surface reflectivity with various distribution. Dividing the whole camera view into several receptive fields, we suppose γ as the number of types corresponding to local ambient light distribution. Under static measurement environment and fixed camera view, γ is completely a constant value. Similarly, we suppose η as the number of types corresponding to local surface reflectivity distribution. Though there must be infinite types of local reflectivity distribution, large training data can include most cases that ensures any so-called 'never-before-seen' scene in test set can find a similar distribution in train set in a local view. If we classify similar local reflectivity distribution into one type, η can be seen as a finite value. Besides, we suppose ρ as the number of types corresponding to local projection wave mode. Since the projected fringes are fixed, ρ is ideally a constant value. However, the fringe distortion due to object surface may slightly influence the local projection distribution (wave mode) in captured fringe pattern. Thus, ρ should be seen as a finite value. In actual experiment, the camera sensor just captures the intensity map, which is the combination of ambient light, projected fringe, and surface reflectivity. Therefore, type-number of local intensity distribution ω is co-decided by γ, ρ, η. Supposing DNN have a certain capacity for ω, the less γ and ρ exist, the more η can be recognized.

Using Ω to represent the ability of chosen DNN (concerning structure, loss functions, optimizer, etc.), supposing the training strategy enables DNN to achieve best capacity, we define the phase retrieval ability of a trained DNN as Γ:

$$\Gamma \propto N_s \cdot \frac{\Omega \cdot \bar{\omega}}{\gamma \cdot \rho} \tag{17}$$

where N_s is the number of scenes in train set, $\bar{\omega}$ mean the average type-number of local intensity distribution in one scene (one projection sequence). Therefore, to achieve better DNN phase retrieval, concrete measures can be conducted. From the positive correlation factors, we should create a large amount of training data containing various scenes, construct strong neural networks with leading loss functions, and try to enlarge the type-number of local intensity distribution. From the negative correlation factors, we should experiment in uniform ambient illuminations (dark room the best), and avoid using complex fringe patterns.

Having the analysis above, all the techniques used in this work can be understood more clearly. In previous work, DNN composite fringe analysis is adopted to achieve single frame 3D measurement. However, DNN single-frequency fringe phase retrieval does have a better accuracy over composite fringes. The reason just lies in the type-number of local projection wave mode ρ. Single-frequency fringes have far less ρ for its periodic wave mode, while composite fringes designed for number-theoretic TPU have no periodicity. In Section 2.1, we proposed fpPSC to minimize the ρ for a projection sequence. Then OPSR is proposed to produce more different combinations of projecting light and surface reflectivity, thus augmenting the local intensity distribution types $\bar{\omega}$ in each projection sequence. This could maximize Loss1&2's guidance. Chances are that OPSR provides high-quality LS fitting results (see **Supplement 1**(Section S2)), strengthening Loss0's guiding effect. In Section 2.2, we adopted a weight-sharing network architecture, constructed physics-based loss functions to give DNN more careful guidance, which

is the enhancement for Ω. To sum up, the fringe sequence is encoded for maximizing neural network's ability, and the network architecture is designed to fully utilize the captured fringe patterns. This is a self-consistent view that must be understood from the overall optimization perspective under the circumstance of dual-frequency composite fringe analysis in this paper.

Actually, this view is compatible with single-frequency fringe analysis and can be extended to other applications. In single-frequency fringe phase retrieval, the sinusoidal PS fringes perfectly satisfy the OPSR. Using weight-sharing strategy with constraints proposed in Section 2.2 to conduct DNN training process, the wrapped phase quality can also be improved considerably in testing stage. This method combined with SPU may push single-frame fringe analysis to a new limitation. In other scenarios that traditional fitting algorithms restrict some specific applications (like dynamic measurement in FPP) while deep learning is faced with underdetermination (ill-posed question), researchers may explore a way to fully utilize the raw data and try to integrate the mathematical/physical rules into deep learning to achieve better outcomes.

In this paper, we introduce a universal deep learning fringe analysis framework for phase-shifting fringe sequence. Taking dual-frequency composite fringes as an example, we present the phase-shifting guided fringes encoding strategy, network architecture and loss functions. Through optimal design for the overall process, our FPP system can achieve unprecedented enhancement compared with previous label-based methods, realizing higher-accuracy and more robust 3D shape measurement. The possible explanation for DL phase retrieval is also discussed, which indicates that although DL performs single-frame fringe analysis well in fixed measurement environment, one trained network is not always reliable in all application scenarios. Due to the essence of present deep neural networks, we would not expect one network to deal with all scenarios in the future. Instead, different expert subnetworks which tell their own pixel-wise confidence levels would be tried to construct, in order to perform self-adaptive accountable high-quality pixel selection, thus further pushing the precision and reliability of deep learning fringe analysis.

Funding. Content in the funding section will be generated entirely from details submitted to Prism. Authors may add placeholder text in the manuscript to assess length, but any text added to this section in the manuscript will be replaced during production and will display official funder names along with any grant numbers provided. If additional details about a funder are required, they may be added to the Acknowledgments, even if this duplicates information in the funding section. See the example below in Acknowledgements. For preprint submissions, please include funder names and grant numbers in the manuscript.

Acknowledgments. The section title should not follow the numbering scheme of the body of the paper. Additional information crediting individuals who contributed to the work being reported, clarifying who received funding from a particular source, or other information that does not fit the criteria for the funding block may also be included; for example, "K. Flockhart thanks the National Science Foundation for help identifying collaborators for this work."

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.

Supplemental document. See Supplement 1 for supporting content.

References

1. C. Zuo, X. Zhang, Y. Hu, *et al.*, "Has 3D finally come of age? ——An introduction to 3D structured-light sensor," Infrared Laser Eng. **49**, 303001–303001 (2020).
2. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," Appl. Opt. **22**, 3977 (1983).
3. Q. Kemao, "Windowed Fourier transform for fringe pattern analysis," Appl. Opt. **43**, 2695 (2004).
4. Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations," Opt. Lasers Eng. **45**, 304–317 (2007).

5. J. Zhong and J. Weng, "Spatial carrier-fringe pattern analysis by means of wavelet transform: Wavelet transform profilometry," Appl. Opt. **43**, 4993 (2004).
6. L. Huang, Q. Kemao, 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**, 141–148 (2010).
7. X. Su and W. Chen, "Fourier transform profilometry:," Opt. Lasers Eng. **35**, 263–284 (2001).
8. 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**, 1152–1160 (2012).
9. C. Zuo, S. Feng, L. Huang, *et al.*, "Phase shifting algorithms for fringe projection profilometry: A review," Opt. Lasers Eng. **109**, 23–59 (2018).
10. C. Zuo, L. Huang, M. Zhang, *et al.*, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," Opt. Lasers Eng. **85**, 84–103 (2016).
11. Z. Wu, W. Guo, and Q. Zhang, "Two-frequency phase-shifting method vs. Gray-coded-based method in dynamic fringe projection profilometry: A comparative review," Opt. Lasers Eng. **153**, 106995 (2022).
12. R. R. Garcia and A. Zakhor, "Consistent stereo-assisted absolute phase unwrapping methods for structured light systems," IEEE J. selected topics Signal Process. **6**, 411–424 (2012).
13. T. Tao, Q. Chen, S. Feng, *et al.*, "High-precision real-time 3D shape measurement based on a quad-camera system," J. Opt. **20**, 014009 (2017).
14. K. Liu, Y. Wang, D. L. Lau, *et al.*, "Dual-frequency pattern scheme for high-speed 3-D shape measurement," Opt. Express **18**, 5229 (2010).
15. C. Zuo, Q. Chen, G. Gu, *et al.*, "High-speed three-dimensional profilometry for multiple objects with complex shapes," Opt. express **20**, 19493–19510 (2012).
16. T. Tao, Q. Chen, J. Da, *et al.*, "Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system," Opt. Express **24**, 20253 (2016).
17. S. Feng, Q. Chen, and C. Zuo, "Graphics processing unit–assisted real-time three-dimensional measurement using speckle-embedded fringe," Appl. optics **54**, 6865–6873 (2015).
18. S. Feng, C. Zuo, T. Tao, *et al.*, "Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry," Opt. Lasers Eng. **103**, 127–138 (2018).
19. W. Guo, Z. Wu, Y. Li, *et al.*, "Real-time 3D shape measurement with dual-frequency composite grating and motion-induced error reduction," Opt. Express **28**, 26882 (2020).
20. L. Lu, V. Suresh, Y. Zheng, *et al.*, "Motion induced error reduction methods for phase shifting profilometry: A review," Opt. Lasers Eng. **141**, 106573 (2021).
21. S. Feng, Q. Chen, G. Gu, *et al.*, "Fringe pattern analysis using deep learning," Adv. Photonics **1**, 1 (2019).
22. H. Yu, X. Chen, Z. Zhang, *et al.*, "Dynamic 3-D measurement based on fringe-to-fringe transformation using deep learning," Opt. Express **28**, 9405 (2020).
23. Z. Qi, X. Liu, J. Pang, *et al.*, "PSNet: A Deep Learning Model-Based Single-Shot Digital Phase-Shifting Algorithm," Sensors **23**, 8305 (2023).
24. J. Qian, S. Feng, T. Tao, *et al.*, "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement," APL Photonics **5**, 046105 (2020).
25. Y. Li, J. Qian, S. Feng, *et al.*, "Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement," Opto-Electronic Adv. **5**, 210021–210021 (2022).
26. Y. Li, J. Qian, S. Feng, *et al.*, "Composite fringe projection deep learning profilometry for single-shot absolute 3D shape measurement," Opt. Express **30**, 3424 (2022).
27. P. Yao, S. Gai, Y. Chen, *et al.*, "A multi-code 3D measurement technique based on deep learning," Opt. Lasers Eng. **143**, 106623 (2021).
28. C. Wang, P. Zhou, and J. Zhu, "Deep learning-based end-to-end 3D depth recovery from a single-frame fringe pattern with the MSUNet++ network," Opt. Express **31**, 33287 (2023).
29. G. E. Spoorthi, R. K. Sai Subrahmanyam Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase Unwrapping of Noisy Data Based on Deep Learning Approach," IEEE Trans. on Image Process. **29**, 4862–4872 (2020).
30. H. Nguyen, Y. Wang, and Z. Wang, "Single-shot 3D shape reconstruction using structured light and deep convolutional neural networks," Sensors **20**, 3718 (2020).
31. S. Fan, S. Liu, X. Zhang, *et al.*, "Unsupervised deep learning for 3D reconstruction with dual-frequency fringe projection profilometry," Opt. Express **29**, 32547 (2021).
32. H. Yu, B. Han, L. Bai, *et al.*, "Untrained deep learning-based fringe projection profilometry," APL Photonics **7**, 016102 (2022).
33. H. Yu, X. Chen, R. Huang, *et al.*, "Untrained deep learning-based phase retrieval for fringe projection profilometry," Opt. Lasers Eng. **164**, 107483 (2023).
34. W. Yin, Y. Che, X. Li, *et al.*, "Physics-informed deep learning for fringe pattern analysis," Opto-Electronic Adv. **0**, 230034–230034 (2024).
35. Y. Zheng, S. Wang, Q. Li, and B. Li, "Fringe projection profilometry by conducting deep learning from its digital twin," Opt. Express **28**, 36568 (2020).