Opto-Electronic Advances

Supplementary information

January 2024, Vol. 7, No. 1

DOI: 10.29026/oea.2024.230034

Physics-informed deep learning for fringe pattern analysis

Wei Yin[1,2,3†], Yuxuan Che[1,2,3†], Xinsheng Li[1,2,3], Mingyu Li[1,2,3], Yan Hu[1,2,3], Shijie Feng[1,2,3*], Edmund Y. Lam[4*], Qian Chen[3*] and Chao Zuo[1,2,3*]

[1]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; [2]Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing 210019, China; [3]Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing 210094, China; [4]Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR 999077, China.

[†]These authors contributed equally to this work.

*Correspondence: SJ Feng, E-mail: shijiefeng@njust.edu.cn; EY Lam, E-mail: elam@eee.hku.hk; Q Chen, E-mail: chenqian@njust.edu.cn; C Zuo, E-mail: zuochao@njust.edu.cn

This file includes:

Supplementary information for this paper is available at https://doi.org/10.29026/oea.2024.230034

Section 1: Architecture of the lightweight DNN in PI-FPA

At present, mainstream fringe analysis approaches using deep learning exploit end-to-end fully convolutional networks to establish an accurate inverse mapping between single-frame fringe and the label phase using massive network parameters. To achieve high-performance phase retrieval, these networks extract high-resolution features to encode accurate spatial information of fringes that is crucial to predict the detailed output, while enlarging the receptive field of feature maps to distill high-level semantic information for improving the accuracy of phase measurement. The U-Net and its derivatives employ multiple convolution and downsampling layers with a small stride to encode different features in both scale and level, and fuse the features of the same level to gradually increase the spatial resolution by skip connections. Due to the redundant computation introduced by the high-resolution feature extraction at the top of the network structure, U-Net improves the phase measurement accuracy at the cost of the network inference speed.

Different from fringe analysis methods using end-to-end deep learning, as shown in Fig. S1, the proposed PI-FPA consists of two main parts: a LeFTP module with the prior knowledge of FT methods and a lightweight deep neural network (DNN). Thanks to robust phase estimation of LeFTP, it not only helps PI-FPA to circumvent the requirement of collecting a large amount of high-quality data in supervised learning methods, but also relieves the burden of phase refinement for lightweight DNNs. The lightweight network, consisting of the context path and the spatial path inspired by BiSeNet[S1,S2], is utilized to further improve the phase accuracy at a low computational cost compared with universal end-to-end image transform networks (U-Net and its derivatives). The context path aims at collecting the fringe and initial phase features with a large receptive field through fast downsampling, and integrating the global context information to guide the refined features for learning. In the encoder part of the context path, a fast downsampling strategy with several ConvX blocks and the Short-Term Dense Concatenate (STDC) module is first used to extract the feature information with scalable receptive field and multi-scale information in Fig. S1(b) and S1(d). Instead of configuring more channels for higher-level layers as U-Net, low-level layers with broader channels in our blocks are used to encode more fine-grained features with small receptive field, while high-level feature tensors with large receptive field have fewer channels as the resolution decreases. The fast downsampling strategy can produce feature maps with different downsampled ratios (including 1/2, 1/4, 1/8, 1/16, and 1/32), respectively. In the decoder phase, the attention-based feature refinement (AFR) module and the fast upsampling operation based on bilinear interpolation are utilized to improve the feature resolution progressively for avoiding the increased computational complexity caused by common transposed convolution.

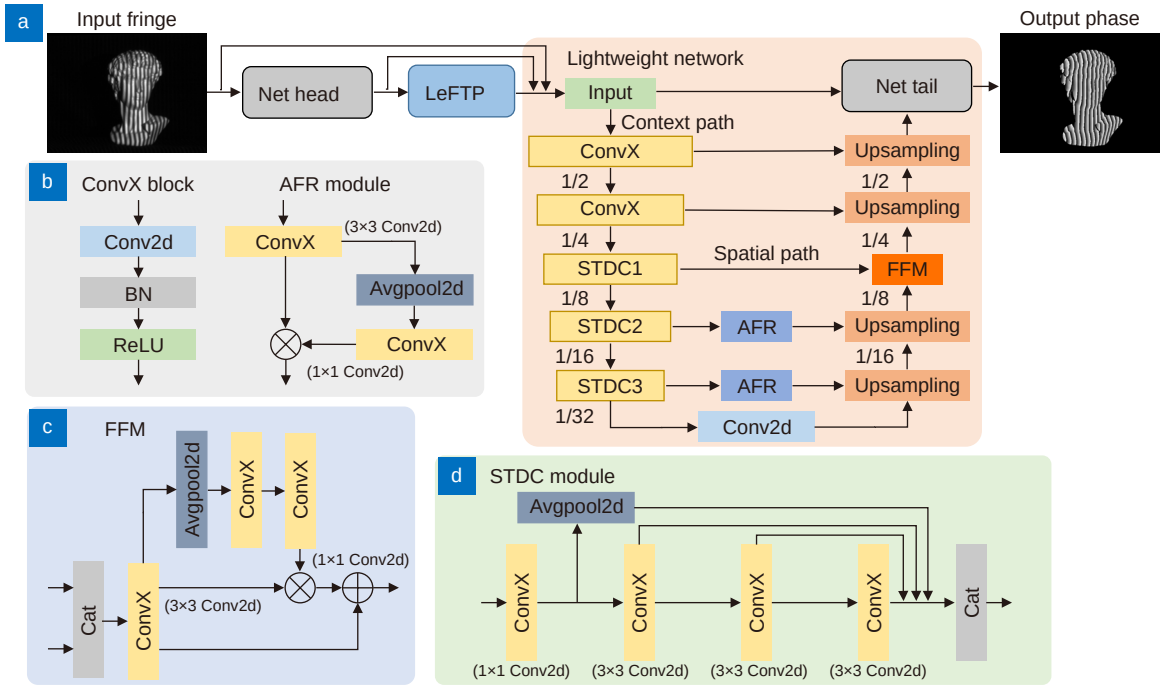


Fig. S1 | Overview of the proposed physics-informed deep learning method for fringe pattern analysis (PI-FPA). (**a**) PI-FPA including a LeFTP module and a lightweight network. (**b–d**) ConvX block, AFR, FFM, and STDC modules of the lightweight network.

For the AFR module in Fig. S1(b), the channel confidences of high-level features are estimated using a global average pooling and a 1×1 ConvX block, and act on themselves to guide the refined feature learning. The bilinear interpolation is applied to the refined tensor after it is residually connected to the inputs of each fast upsampling module.

In the spatial path, its encoder part shares the same parameters with the context path, and captures the spatial information encoding rich detail information and outputs low-level features, while the context path encodes the global context information and outputs higher-level features. The features with 1/8 downsampled ratio from the context path and the spatial path are concatenated by Feature Fusion module (FFM), and upsampled to output final phases using the predicted $M(x, y)$ and $D(x, y)$ in Eq. (7).

Section 2: Single-shot 3D imaging method using PI-FPA and stereo phase unwrapping

Benefiting from the proposed PI-FPA, we can successfully achieve high-performance single-frame phase retrieval. Although PI-FPA almost perfectly extracts the wrapped phase from a 64-period fringe image, it introduces more phase ambiguities, leading to the low reliability of phase unwrapping for absolute 3D measurements. In Fringe projection profilometry (FPP), phase unwrapping methods can be grouped into three main classes in aspects of the operating domain: spatial phase unwrapping[S3,S4], temporal phase unwrapping[S5–S9], and stereo phase unwrapping[S10–S14]. Spatial phase unwrapping is highly suited for dynamic 3D acquisition and can provide a relatively absolute phase map using only a single wrapped phase[S3]. However, the continuity of the phase is an essential prerequisite for the successful implementation of spatial phase unwrapping, making it impossible for measuring discontinuous surfaces or abrupt depth with step heights greater than 2π. In order to solve the problem above, temporal phase unwrapping methods are proposed to realize absolute phase unwrapping with the aid of additional wrapped phase maps with different frequencies[S5], but at the cost of measurement efficiency.

Different from the two methods above, stereo phase unwrapping utilizes multi-view geometric constraints to retrieve absolute phases using high-frequency wrapped phases from different perspectives as shown in Fig. S2, potentially overcoming the respective bottlenecks of spatial and temporal phase unwrapping. Here, we continue to use the previous work[S11] proposed by our team, which builds a position-optimized multi-camera system to guarantee the reliability of stereo phase unwrapping. Specifically, a multi-view structured light system including a projector and three cameras and a weighted phase consistency check strategy are adopted to achieve single-shot 3D imaging. Due to the truncation effect of the arctangent function, the obtained phase $\phi(x, y)$ is wrapped within the range of $(-\pi, \pi]$, and its relationship with the absolute ones $\Phi(x, y)$ is:

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y) \ , \tag{S1}$$

where $k(x, y)$ represents the fringe order of $\Phi(x, y)$, and its value range is from 0 to $N-1$. For the wrapped phase $\phi^{c1}(x, y)$ of N-period fringes from the main camera, the fringe order of each valid pixel exists N possibilities. Taking each possible order $k^{c1}(x, y)$ into consideration, the corresponding depth value $Z_k^{c1}(x, y)$ is calculated using system parameters between the main camera and the projector. It is easy to find that some possible depth values $Z_k^{c1}(x, y)$ are beyond the pre-defined depth range based on depth constraint, and the corresponding orders $k(x, y)$ can be excluded from the candidates:

$$Z_{\min} \le Z_k^{c1}(x, y) \le Z_{\max} \ . \tag{S2}$$

By setting an appropriate depth range $[Z_{\min}, Z_{\max}]$, it is theoretically possible to exclude all wrong candidate points and directly output the absolute phase, while is infeasible in most applications for a large 3D measurement volume. And then, a weighted phase consistency check strategy is used to eliminate remaining false candidates after depth constraint:

$$\Delta\phi_k^{c1c2c3}(x^{c1}, y^{c1}) = \gamma_1 \Delta\phi_k^{c1c2}(x^{c2}, y^{c2}) + \gamma_2 \Delta\phi_k^{c1c3}(x^{c3}, y^{c3}) \ , \tag{S3}$$

where (x^{c2}, y^{c2}) and (x^{c3}, y^{c3}) are the corresponding coordinates of the remaining candidates reprojected in the auxiliary cameras C2 and C3, $\Delta\phi_k^{c1c2}(x^{c2}, y^{c2})$ and $\Delta\phi_k^{c1c3}(x^{c3}, y^{c3})$ are the phase difference between the main camera C1 and the auxiliary cameras C2 and C3. The weighted coefficients γ_1 and γ_2 depend on the baseline between the main camera and the auxiliary camera, and the large baseline corresponds to a small weighted coefficient. Here, we set a small baseline between C1 and C2 and a small threshold of γ_1 to efficiently decrease the number of candidates by the phase consistency

check. And a large baseline between C1 and C3 and a large threshold of γ_2 are set to further reject all the remaining false candidates by the second phase consistency check, and convert the absolute phase to the 3D coordinates using system parameters between the cameras C1 and C3, enabling high-precision single-shot 3D imaging.

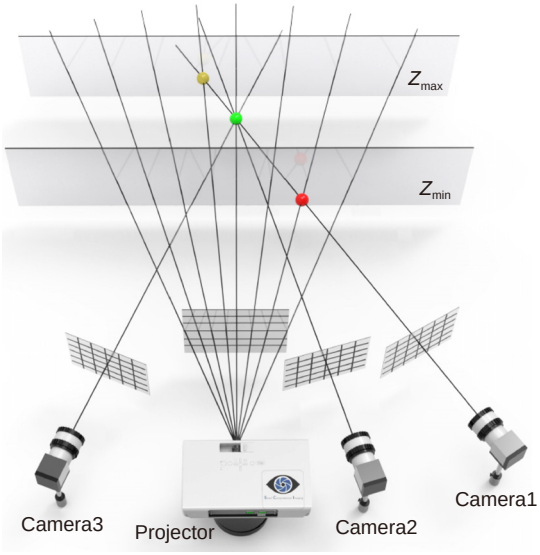


Fig. S2 | The diagram of stereo phase unwrapping based on multi-view geometric constraints.

Section 3: Experimental setup and data process

To prepare datasets for the deep neural network, a multi-view structured light system is set up including a projector (LightCrafter 4500Pro, Texas Instruments) and three cameras (acA640-750um, Basler) as shown in Fig. S3. To collect fringe data for training, the projector projects three sets of PS patterns with different periods (including 1, 8, and 64) onto the test objects. The captured 64-period fringe image is the input of PI-FPA, and the label phase is obtained by 12-step PS. In the experiment, we collected the dataset for training, validation, and testing from 1200 different scenes including the random combination of 30 simple and complex objects. The whole dataset has 1200 image pairs, which are divided into 800 image pairs for training, 200 image pairs for validation, and 200 image pairs for testing. For the training images, all samples were randomly combined, rotated, and placed in the measured area. The fringe images collected by the camera are directly input into the network without any image preprocessing and data augmentation. During training, to monitor the performance of the neural networks for samples that they have never seen, the scenes in these training, verification, and testing datasets are separate from each other. The proposed PI-FPA is implemented using Pytorch framework (Facebook) and is computed on a desktop computer equipped with an Intel i7-9700K CPU (8 cores, 8 threads) and an NVIDIA GeForce RTX2080Ti graphics card (4352 CUDA cores, 11 GB VRAM). The composite loss function consists of mean square error (MSE) and mean absolute error (MAE) in Eqs. (12–14). To maximize the single-frame fringe analysis performance of PI-FPA, α_1 and β_1 are set as 1 and 0.5 after an exhaustive empirical search. Since

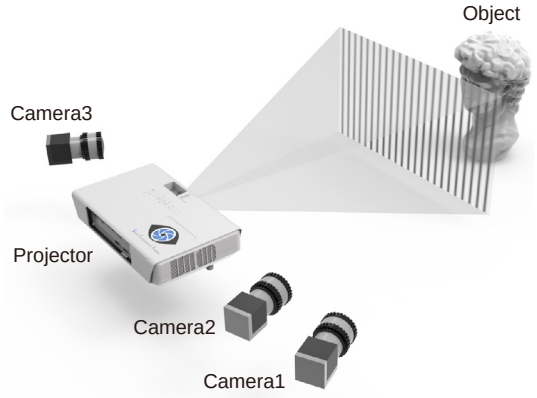


Fig. S3 | The diagram of the multi-view structured light system consisting of a projector and three cameras.

the initial phase from LeFTP guides the lightweight DNN to enhance its phase recovery ability, in the early stage of network training, α_2 and β_2 are set as 0.1, and then gradually reduced to 0.02. The optimizer is Adam, and the training epoch is set as 300. The loss curves of the training and validation dataset for LeFTP and PI-FPA are shown in Fig. S4.

Section 4: Fringe analysis results of PI-FPA and U-Net using different amounts of training images

Here, we present single-shot fringe pattern analysis results of U-Net and PI-FPA using different amounts of training images as shown in Fig. S5. Compared with U-Net with 800 training image pairs, PI-FPA reduces the MAE of the phase errors by about 12.55% while requiring only 400 training image pairs, which demonstrates its good generalization.

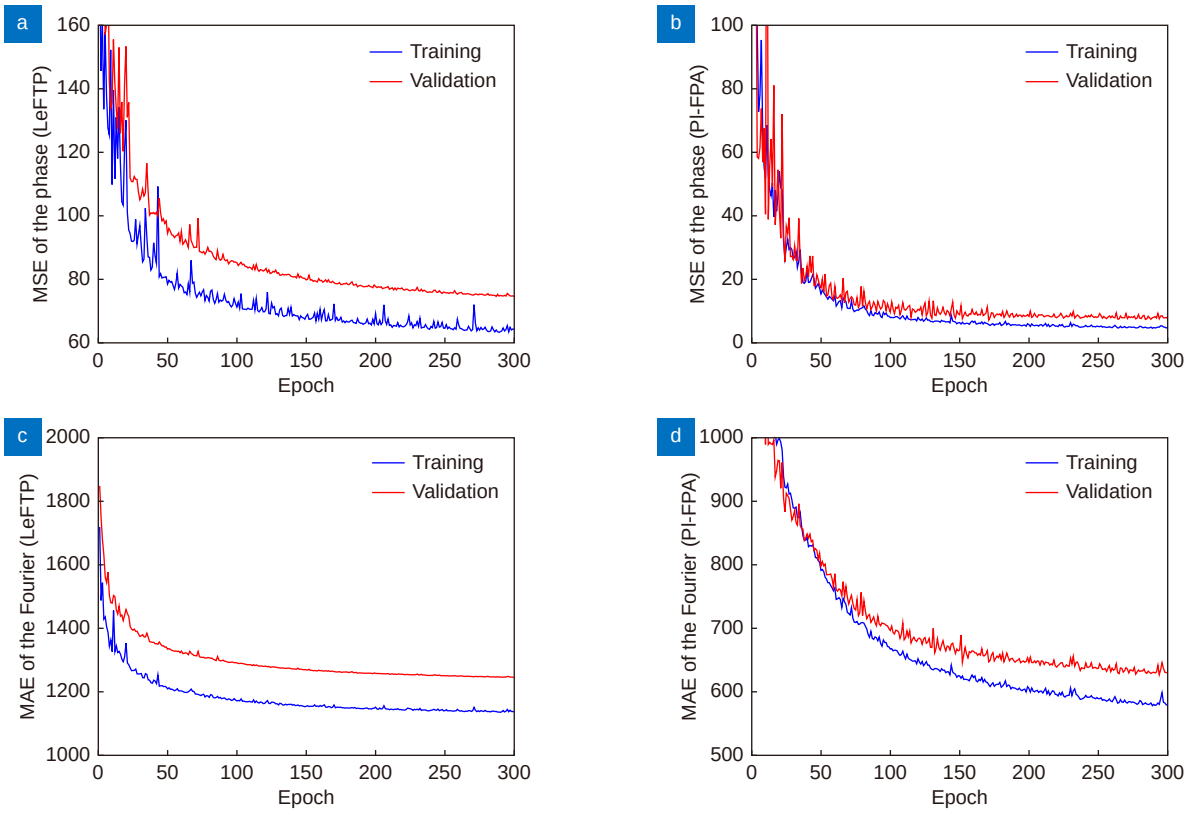


Fig. S4 | Loss curves of the training and validation dataset for LeFTP and PI-FPA. (**a**) MSE of the phase outputted by LeFTP. (**b**) MSE of the phase outputted by PI-FPA. (**c**) MAE of the Fourier outputted by LeFTP, (**d**) MAE of the Fourier outputted by PI-FPA.

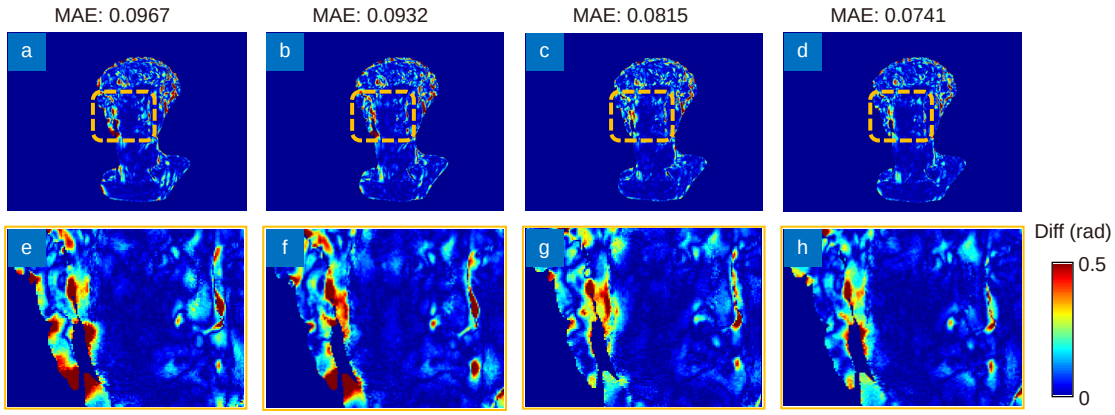
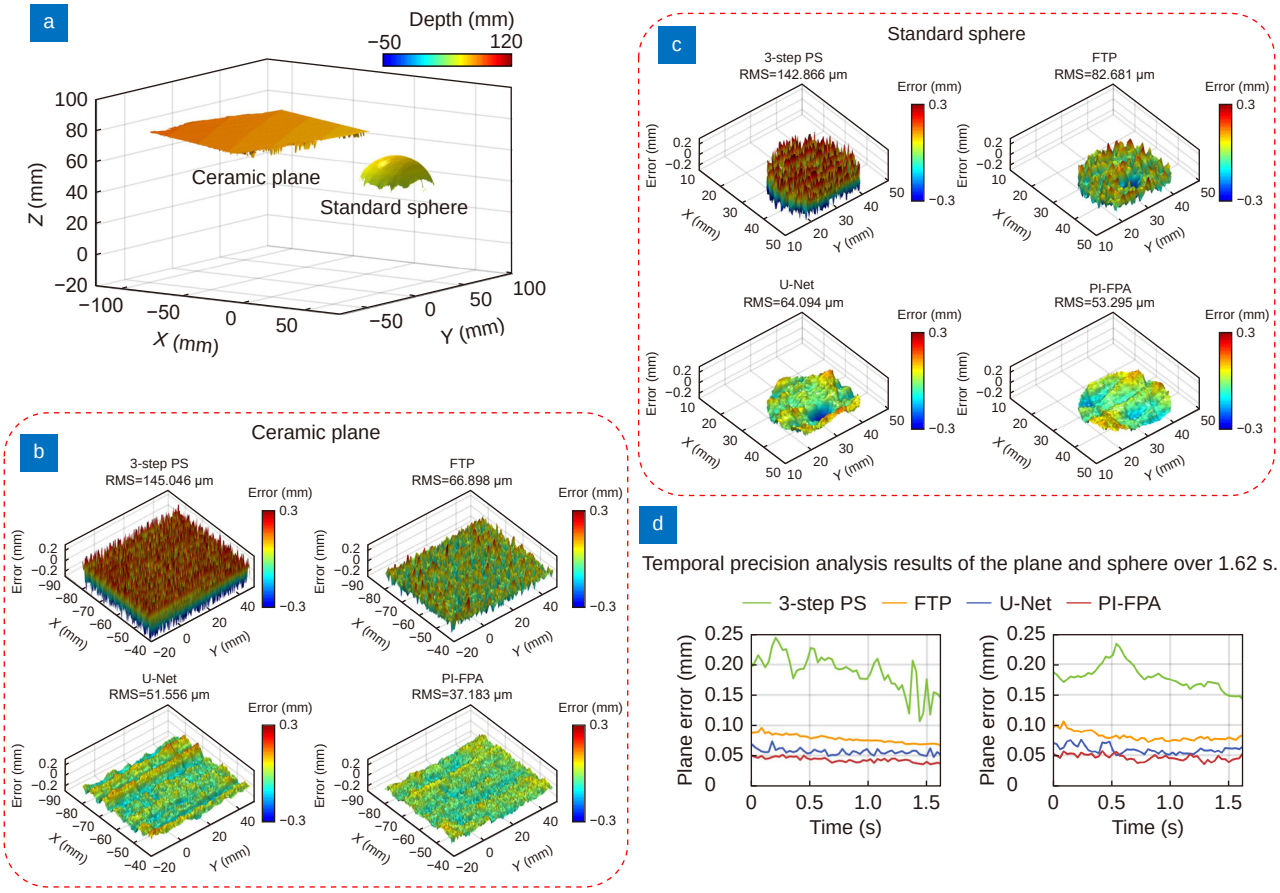
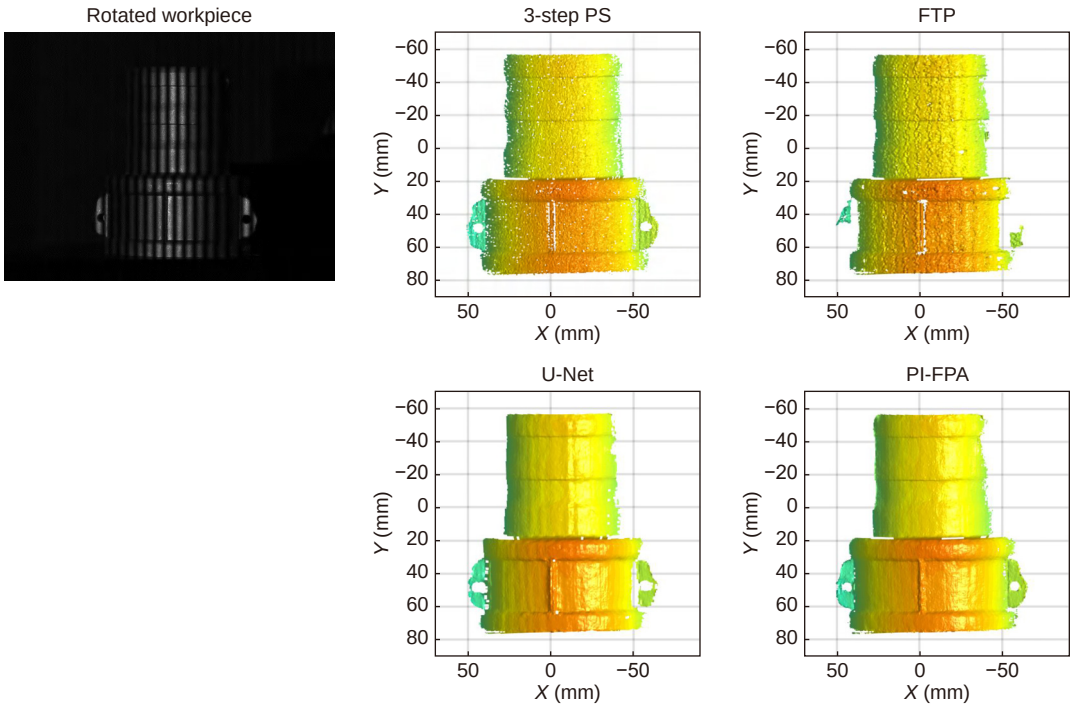


Fig. S5 | Comparative results for single-shot fringe pattern analysis of the David model using different amounts of training images. (**a**, **b**) The phase errors using U-Net with 400 training image pairs and 800 training image pairs. (**c**, **d**) The phase errors using PI-FPA with 400 training image pairs and 800 training image pairs. (**e**, **h**) the corresponding magnified views of (a, d).

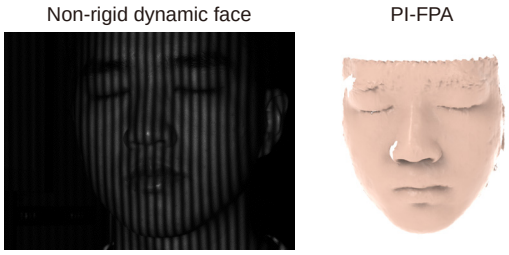
Section 5: Supplementary Videos



Video S1 | Precision analysis results for measuring a ceramic plane and a standard sphere moving along the *Z* axis using 3-step PS, FTP, U-Net, and PI-FPA.



Video S2 | 3D measurement results of the rotated industrial part using 3-step PS, FTP, U-Net, and PI-FPA.



Video S3 | 3D measurement results of non-rigid dynamic face using PI-FPA.

References

S1. Yu CQ, Wang JB, Peng C et al. BiSeNet: bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the 15th European Conference on Computer Vision (ECCV)* 334–349 (Springer, 2018); http://doi.org/10.1007/978-3-030-01261-8_20.

S2. Fan MY, Lai SQ, Huang JS et al. Rethinking BiSeNet for real-time semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9711–9720 (IEEE, 2021); http://doi.org/10.1109/CVPR46437.2021.00959.

S3. Su XY, Chen WJ. Reliability-guided phase unwrapping algorithm: a review. *Opt Lasers Eng* **42**, 245–261 (2004).

S4. Zhao M, Huang L, Zhang QC et al. Quality-guided phase unwrapping technique: comparison of quality maps and guiding strategies. *Appl Opt* **50**, 6214–6224 (2011).

S5. Zuo C, Huang L, Zhang ML et al. Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review. *Opt Lasers Eng* **85**, 84–103 (2016).

S6. Wang YJ, Zhang S. Novel phase-coding method for absolute phase retrieval. *Opt Lett* **37**, 2067–2069 (2012).

S7. Wu ZJ, Guo WB, Li YY et al. High-speed and high-efficiency three-dimensional shape measurement based on gray-coded light. *Photonics Res* **8**, 819–829 (2020).

S8. Yin W, Zuo C, Feng SJ et al. High-speed three-dimensional shape measurement using geometry-constraint-based number-theoretical phase unwrapping. *Opt Lasers Eng* **115**, 21–31 (2019).

S9. Yin W, Chen Q, Feng SJ et al. Temporal phase unwrapping using deep learning. *Sci Rep* **9**, 20175 (2019).

S10. Zhong K, Li ZW, Shi YS et al. Fast phase measurement profilometry for arbitrary shape objects without phase unwrapping. *Opt Lasers Eng* **51**, 1213–1222 (2013).

S11. Tao TY, Chen Q, Feng SJ et al. High-precision real-time 3D shape measurement based on a quad-camera system. *J Opt* **20**, 014009 (2018).

S12. Yin W, Feng SJ, Tao TY et al. High-speed 3d shape measurement using the optimized composite fringe patterns and stereo-assisted structured light system. *Opt Exp* **27**, 2411–2431 (2019).

S13. Yin W, Zhong JX, Feng SJ et al. Composite deep learning framework for absolute 3d shape measurement based on single fringe phase retrieval and speckle correlation. *J Phys Photonics* **2**, 045009 (2020).

S14. Lohry W, Chen V, Zhang S. Absolute three-dimensional shape measurement using coded fringe patterns without phase unwrapping or projector calibration. *Opt Express* **22**, 1287–1301 (2014).