

Fringe pattern analysis using deep learning

Shijie Feng,^{a,b,c} Qian Chen,^{a,b,*} Guohua Gu,^{a,b} Tianyang Tao,^{a,b} Liang Zhang,^{a,b,c} Yan Hu,^{a,b,c} Wei Yin,^{a,b,c} and Chao Zuo^{a,b,c,*}

^aNanjing University of Science and Technology, School of Electronic and Optical Engineering, Nanjing, China

^bJiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing, China

^cNanjing University of Science and Technology, Smart Computational Imaging Laboratory (SCILab), Nanjing, China

Abstract. In many optical metrology techniques, fringe pattern analysis is the central algorithm for recovering the underlying phase distribution from the recorded fringe patterns. Despite extensive research efforts for decades, how to extract the desired phase information, with the highest possible accuracy, from the minimum number of fringe patterns remains one of the most challenging open problems. Inspired by recent successes of deep learning techniques for computer vision and other applications, we demonstrate for the first time, to our knowledge, that the deep neural networks can be trained to perform fringe analysis, which substantially enhances the accuracy of phase demodulation from a single fringe pattern. The effectiveness of the proposed method is experimentally verified using carrier fringe patterns under the scenario of fringe projection profilometry. Experimental results demonstrate its superior performance, in terms of high accuracy and edge-preserving, over two representative single-frame techniques: Fourier transform profilometry and windowed Fourier transform profilometry.

Keywords: fringe analysis; phase measurement; deep learning.

Received Aug. 22, 2018; accepted for publication Jan. 8, 2019; published online Feb. 28, 2019.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.1.2.025001](https://doi.org/10.1117/1.AP.1.2.025001)]

Optical measurement techniques such as holographic interferometry,¹ electronic speckle pattern interferometry,² and fringe projection profilometry³ are quite popular for noncontact measurements in many areas of science and engineering, and have been extensively applied for measuring various physical quantities, such as displacement, strain, surface profile, and refractive index. In all these techniques, the information about the measured physical quantity is stored in the phase of a two-dimensional fringe pattern. The accuracy of measurements carried out by these optical techniques is thus fundamentally dependent on the accuracy with which the underlying phase distribution of the recorded fringe patterns is demodulated.

Over the past few decades, tremendous efforts have been devoted to developing various techniques for fringe analysis. The techniques can be broadly classified into two categories: (1) phase-shifting (PS) methods that require multiple fringe patterns to extract phase information,⁴ and (2) spatial phase-demodulation methods that allow phase retrieval from a single

fringe pattern, such as the Fourier transform (FT),⁵ windowed Fourier transform (WFT),⁶ and wavelet transform (WT) methods.⁷ Compared with spatial phase demodulation methods, multiple-shot PS techniques are generally more robust and can achieve pixel-wise phase measurement with higher resolution and accuracy. Furthermore, the PS measurements are quite insensitive to nonuniform background intensity and fringe modulation. Nevertheless, due to their multishot nature, these methods are difficult to apply to dynamic measurements and are more susceptible to external disturbance and vibration. Thus, for many applications, phase extraction from a single fringe pattern is desired, which falls under the purview of spatial fringe analysis. In contrast to PS techniques where the phase map is demodulated on a pixel-by-pixel basis, phase estimation at a pixel according to spatial methods is influenced by the pixel's neighborhood, or even all pixels in the fringe pattern, which provides better tolerance to noise, yet at the expense of poor performance around discontinuities and isolated regions in the phase map.^{8,9}

Deep learning is a powerful machine learning technique that employs artificial neural networks with multiple layers of

*Address all correspondence to Qian Chen, E-mail: chenqian@njust.edu.cn; Chao Zuo, E-mail: zuochao@njust.edu.cn

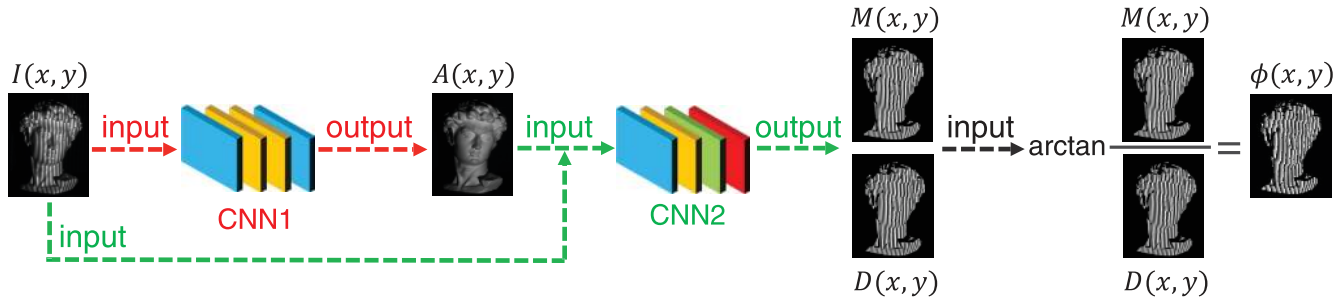


Fig. 1 Flowchart of the proposed method where two convolutional networks (CNN1 and CNN2) and the arctangent function are used together to determine the phase distribution. For CNN1 (in red), the input is the fringe image $I(x, y)$, and the output is the estimated background image $A(x, y)$. For CNN2 (in green), the inputs are the fringe image $I(x, y)$ and the background image $A(x, y)$ predicted by CNN1, and the outputs are the numerator $M(x, y)$ and the denominator $D(x, y)$. The numerator and denominator are then fed into the arctangent function to calculate the phase $\phi(x, y)$.

increasingly richer functionality and has shown great success in numerous applications for which data are abundant.^{10,11} In this letter, we demonstrate experimentally for the first time, to our knowledge, that the use of a deep neural network can substantially enhance the accuracy of phase demodulation from a single fringe pattern. To be concrete, the networks are trained to predict several intermediate results that are useful for the calculation of the phase of an input fringe pattern. During the training of the networks, we capture PS fringe images of various scenes to generate the training data. The training label (ground truth) of each training datum is a pair of intermediate results calculated from the PS algorithm. After appropriate training, the neural network can blindly take only one input fringe pattern and output the corresponding estimates of these intermediate results with high fidelity. Finally, a high-accuracy phase map can be retrieved through the arctangent function with the intermediate results estimated through deep learning. Experimental results on fringe projection profilometry confirm that this deep-learning-based method is able to substantially improve the quality of the retrieved phase from only a single fringe pattern, compared to state-of-the-art methods.

Here, the network configuration is inspired by the basic process of most phase demodulation techniques, which is briefly recalled as follows. The mathematical form of a typical fringe pattern can be represented as

$$I(x, y) = A(x, y) + B(x, y) \cos \phi(x, y), \quad (1)$$

where $I(x, y)$ is the intensity of the fringe pattern, $A(x, y)$ is the background intensity, $B(x, y)$ is the fringe amplitude, and $\phi(x, y)$ is the desired phase distribution. Here, x and y refer to the pixel coordinates. In most phase demodulation techniques, the background intensity $A(x, y)$ is regarded as a disturbance term and should be removed from the total intensity. Then a wrapped phase map is recovered from an inverse trigonometric function whose argument is a ratio for which the numerator characterizes the phase sine [$\sin \phi(x, y)$] and the denominator characterizes the phase cosine [$\cos \phi(x, y)$]:

$$\phi(x, y) = \arctan \frac{M(x, y)}{D(x, y)} = \arctan \frac{cB(x, y) \sin \phi(x, y)}{cB(x, y) \cos \phi(x, y)}, \quad (2)$$

where c is a constant dependent on the phase demodulation algorithm (e.g., in FT $c = 0.5$, in N -step PS $c = N/2$), and $M(x, y)$ and $D(x, y)$ represent the shorthand for the numerator and denominator terms, respectively. Note that the signs of $M(x, y)$ and $D(x, y)$ can be further used to uniquely define a quadrant for each calculation of $\phi(x, y)$. With the four-quadrant phasor space, the phase values at each point can be determined modulo 2π .

In order to emulate the process above, two different convolutional neural networks (CNN) are constructed, which are connected cascadedly according to Fig. 1. The first convolutional neural network (CNN1) uses the raw fringe pattern $I(x, y)$ as input and estimates the background intensity $A(x, y)$ of the fringe pattern. With the estimated background image $A(x, y)$ and the original fringe image $I(x, y)$, the second convolutional neural network (CNN2) is trained to predict the numerator $M(x, y)$ and the denominator $D(x, y)$ of the arctangent function, which are fed into the subsequent arctangent function [Eq. (2)] to obtain the final phase distribution $\phi(x, y)$.

To generate the ground truth data used as the label to train the two convolutional neural networks, the phase retrieval is achieved by using the N -step PS method. The corresponding N PS fringe patterns acquired can be represented as

$$I_n(x, y) = A(x, y) + B(x, y) \cos[\phi(x, y) - \delta_n], \quad (3)$$

where the index $n = 0, 1, \dots, N-1$, and δ_n is the phase shift that equals $\frac{2\pi n}{N}$. With the orthogonality of trigonometric functions, the background intensity can be obtained as

$$A(x, y) = \frac{1}{N} \sum_{n=0}^{N-1} I_n(x, y). \quad (4)$$

With the least square method, the phase can be calculated as

$$\phi(x, y) = \arctan \frac{\sum_{n=0}^{N-1} I_n(x, y) \sin \delta_n}{\sum_{n=0}^{N-1} I_n(x, y) \cos \delta_n}. \quad (5)$$

Thus, the numerator and the denominator of the arctangent function in Eq. (2) can be expressed as

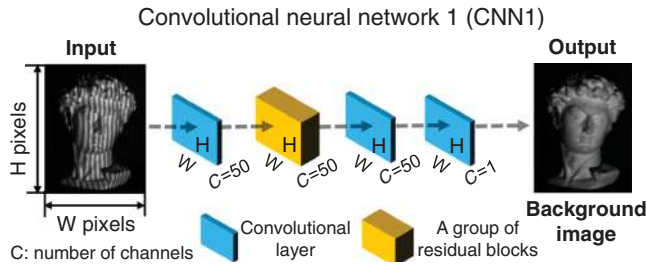


Fig. 2 Schematic of CNN1, which is composed of convolutional layers and several residual blocks.

$$M(x, y) = \sum_{n=1}^{N-1} I_n(x, y) \sin \delta_n = \frac{N}{2} B(x, y) \sin \phi(x, y), \quad (6)$$

$$D(x, y) = \sum_{n=0}^{N-1} I_n(x, y) \cos \delta_n = \frac{N}{2} B(x, y) \cos \phi(x, y). \quad (7)$$

The expressions above show that the numerator $M(x, y)$ and the denominator $D(x, y)$ are closely related to the original fringe pattern in Eq. (1) through a quasilinear relationship with the background image $A(x, y)$. Thus, $M(x, y)$ and $D(x, y)$ can be learned by deep neural networks with ease given the knowledge of $A(x, y)$, which justifies our network. It should be noted that the simple input–output network structure [linking fringe pattern $I(x, y)$ to phase $\phi(x, y)$ directly] performs poorly in our case since it is difficult to follow the phase wraps (2π jumps) in the phase map precisely. Therefore, instead of estimating the phase directly, our deep neural networks are trained to predict the intermediate results, i.e., the numerator and the denominator of the arctangent function in Eq. (2), to obtain a better phase estimate. To further validate the superiority of the proposed method, an ablation analysis is presented in Sec. 6 of the Supplementary Material, in which three methods that (1) estimate the phase $\phi(x, y)$ directly; (2) predict $D(x, y)$ and $M(x, y)$ without $A(x, y)$; and (3) calculate $A(x, y)$, $D(x, y)$, and $M(x, y)$ simultaneously are compared experimentally. The comparative

results indicate that our method is more advantageous in phase reconstruction accuracy than others.

To further reveal the internal structure of the two networks, the diagrams of the two convolutional neural networks are shown in Figs. 2 and 3. The labeled dimensions of the layers or the blocks show the size of their output data. The input of CNN1 is a raw fringe pattern with $W \times H$ pixels. It is then successively processed by a convolutional layer, a group of residual blocks (containing four residual blocks) and two convolutional layers. The last layer estimates the gray values of the background image. With the predicted background intensity and the raw fringe pattern, as shown in Fig. 3, CNN2 calculates the numerator and denominator terms. In CNN2, the input data having two channels are downsampled by $\times 1$ and $\times 2$ in two different paths. In the second path, the data are first downsampled for a high-level perception and then upsampled to match the original dimensions. With the two-scale data flow paths, the network can perceive more surface details for both the numerator and the denominator. We provide additional details about the architectures of our networks in Supplementary Sec. 3.

The performance of the proposed approach was demonstrated under the scenario of fringe projection profilometry. The experiment consisted of two steps: training and testing. In order to obtain the ground truth of training data, 12-step PS patterns with spatial frequency $f = 160$ were created and projected by our projector (DLP 4100, Texas Instruments) onto various objects. The fringe images were captured simultaneously by a CMOS camera (V611, Vision Research Phantom) of 8-bit pixel depth and of resolution 1280×800 . Training objects with different materials, colors, and reflectivity are preferable to enhance the generalization capability of the proposed method. Also, analogous to traditional approaches of fringe analysis that require fringes with enough signal-to-noise ratio or without saturated pixels, the proposed method prefers objects without very dark or shiny surfaces. Our training dataset is collected from 80 scenes. It consists of 960 fringe patterns and the corresponding ground truth data that are obtained by a 12-step PS method (see Supplementary Secs. 1 and 2 for details about the optical setup and the collection of training data). Since one of the inputs of CNN2 is the output of CNN1, CNN1 was trained first and CNN2 was trained with the predicted background intensities and captured fringe patterns. These two

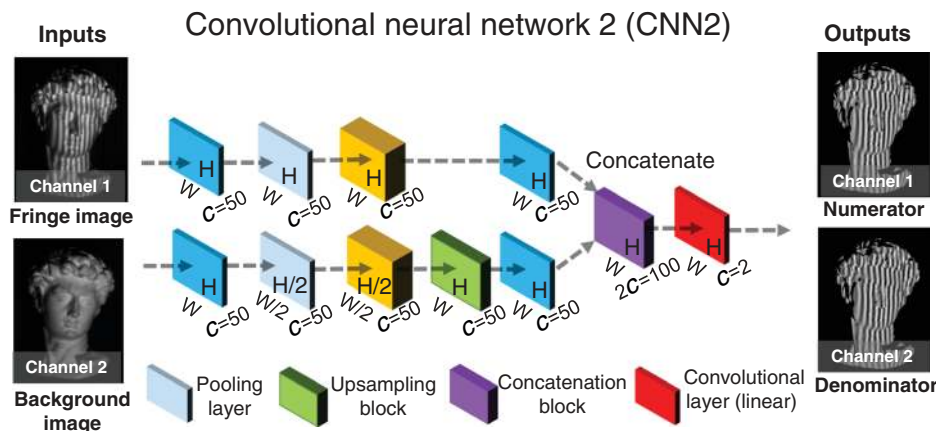


Fig. 3 Schematic of CNN2, which is more sophisticated than CNN1 and further includes two pooling layers, an upsampling layer, a concatenation block, and a linearly activated convolutional layer.

neural networks were implemented using the TensorFlow framework (Google) and were computed on a GTX Titan graphics card (NVIDIA). To monitor during training the accuracy of the neural networks on data that they have never seen before, we created a validation set including 144 fringe images from 12 scenes that are separate from the training scenarios. Additional details on the training of our networks are provided in [Supplementary Sec. 3](#).

To test the trained neural networks versus classic single-frame approaches (i.e., FT⁵ and WFT⁶), we measured a scene containing two isolated plaster models, as shown in Fig. 4(a). Compared with the right model, the left one has a more complex surface, e.g., the curly hair and the high-bridged nose. Note that this scenario was never seen by our neural networks during the training stage. The trained CNN1 using Fig. 4(a) as an input predicted a background intensity as shown in Fig. 4(b). From the enlarged views, we can see that the fringes have been removed completely through the deep neural network. Then, the trained CNN2 took the fringe pattern and the predicted background intensity as inputs and estimated the numerator $M(x, y)$ and the denominator $D(x, y)$; results are shown in Figs. 4(c) and 4(d), respectively. The phase was calculated by Eq. (2) and is shown in Fig. 4(e). In order to evaluate the quality of the estimated phase more easily, we unwrapped it by multifrequency temporal phase unwrapping,¹² in which additional phase maps of fringe patterns of different frequencies were computed with PS algorithm and were then used to unwrap the phase obtained through deep learning. To demonstrate the accuracy of the unwrapped phase, the phase error was calculated against a reference phase map, which was obtained by the 12-step PS method and was unwrapped with the same strategy.

Figures 5(a)–5(c) show the overall absolute phase errors of these approaches, and the calculated mean absolute error (MAE) of each method is listed in Table 1. Note that the adjustable parameters (e.g., the window size) in FT and WFT have been carefully tuned in order to get the best results possible. The result of FT shows the most prominent phase distortion as well as the largest MAE of 0.20 rad. WFT performed better than FT, with fewer errors for both models (MAE 0.19 rad). Among these approaches, the proposed deep-learning-based method demonstrates the least error, which is 0.087 rad. Furthermore, after the training stage, our method becomes fully automatic and does not require a manual parameter search to optimize its performance. To compare the error maps in detail, the phase errors of two complex areas are presented in Fig. 5(d): the hair of the left model and the skirt of the right one. From Fig. 5(d), obvious errors can be observed in the results of FT and WFT, which are mainly concentrated in the boundaries or abrupt depth-changing regions. By contrast, our approach greatly reduced the phase distortion, demonstrating its significantly improved performance in measuring objects with discontinuities and isolated complex surfaces. To further test and compare the performance of our technique with FT and WFT, [Sec. 7 of the Supplementary Material](#) details the measurements of more kinds of objects, which also shows that our method is superior to FT and WFT in terms of phase reconstruction accuracy.

For a more intuitive comparison, we converted the unwrapped phase into 3-D rendered geometries through stereo triangulation,¹³ as shown in Fig. 6. Figure 6(a) shows that the reconstructed result from FT features many grainy distortions, which are mainly due to the inevitable spectral leakage and overlapping in the frequency domain. Compared with FT, the

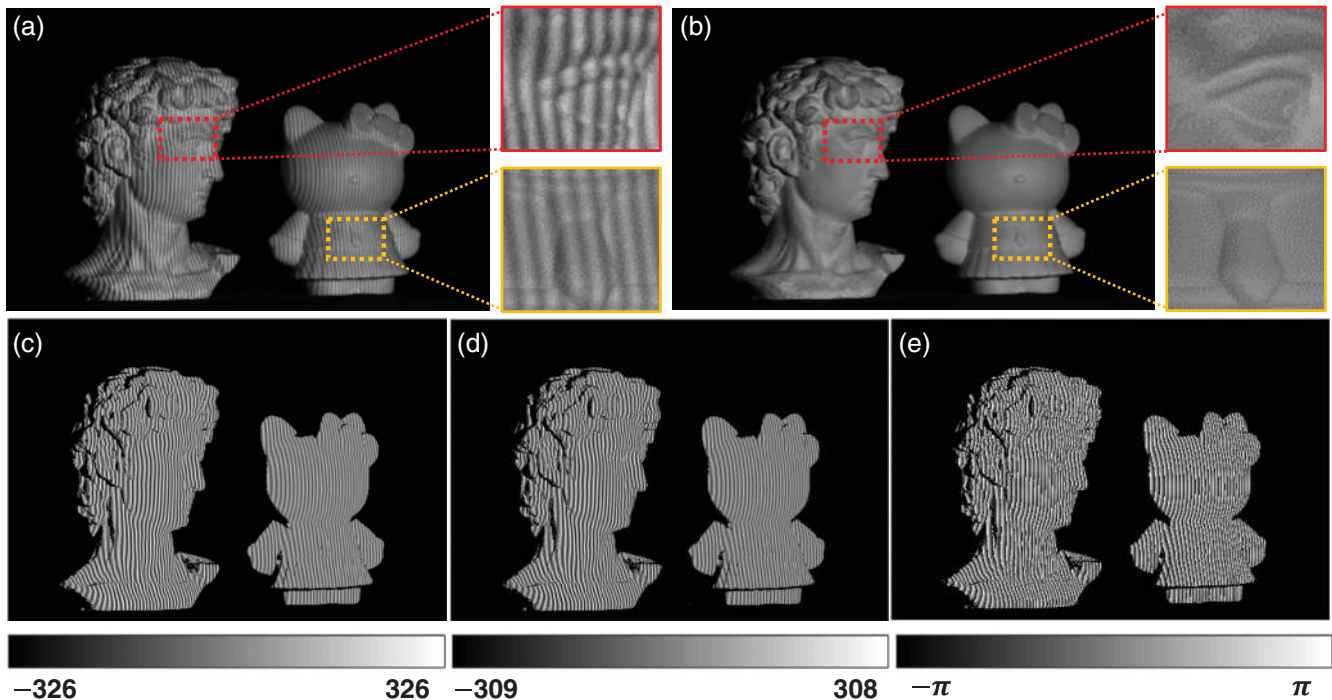


Fig. 4 Testing using the trained networks on a scene that is not present in the training phase. (a) Input fringe image $I(x, y)$, (b) background image $A(x, y)$ predicted by CNN1, (c) and (d) numerator $M(x, y)$ and denominator $D(x, y)$ estimated by CNN2, (e) phase $\phi(x, y)$ calculated with (c) and (d).

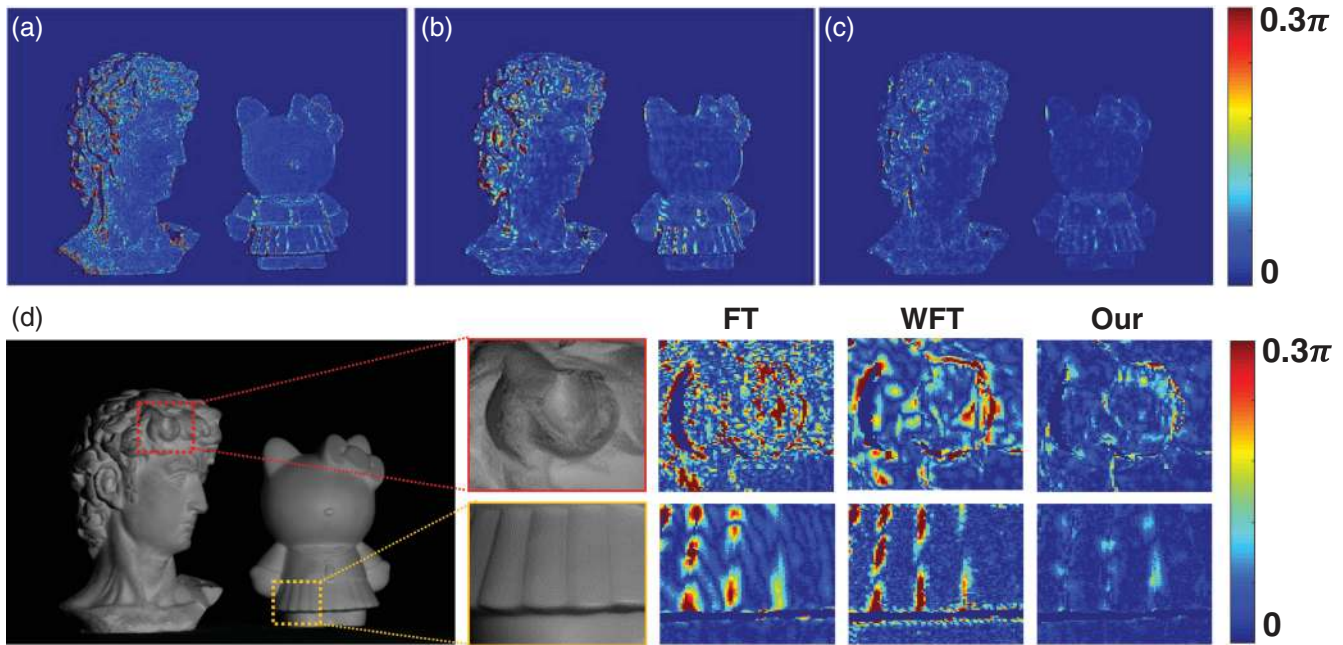


Fig 5 Comparison of the phase error of different methods: (a) FT, (b) WFT, (c) our method, and (d) magnified views of the phase error for two selected complex regions.

Table 1 Phase error of FT, WFT, and our method.

Method	FT	WFT	Our
MAE (rad)	0.20	0.19	0.087

WFT reconstructed the objects with more smooth surfaces but failed to preserve the surface details, e.g., the eyes of the left model and the wrinkles of the skirt of the right model, as can be seen in Fig. 6(b). Among these reconstructions, the deep-learning-based approach yielded the highest-quality 3-D reconstruction [Fig. 6(c)], which almost visually reproduced

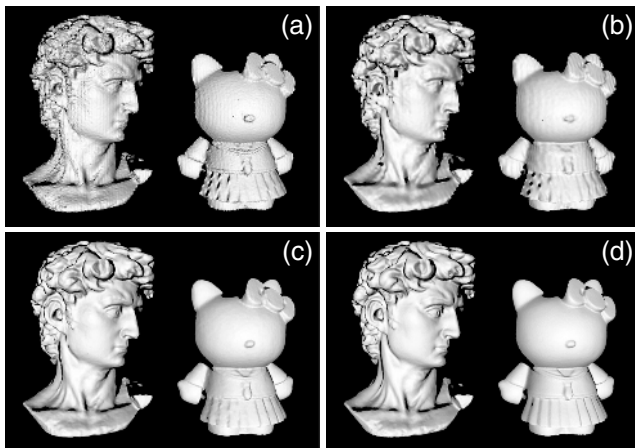


Fig 6 Comparison of the 3-D reconstruction results for different methods: (a) FT, (b) WFT, (c) our method, and (d) ground truth obtained by the 12-step PS profilometry.

the ground truth data [Fig. 6(d)] where 12-step PS fringe patterns were used.

It should be further mentioned that, in the above experiment, the carrier frequency of the fringe pattern is an essential factor affecting the performance of FT and WFT, which was set sufficiently high ($f = 160$) in order to yield results with reasonable accuracy and spatial resolution. However, it can be troublesome for them to analyze the fringe patterns where the carrier frequency is relatively low. As shown in [Sec. 4 of the Supplementary Material](#), the reconstruction quality of FT and WFT degraded to 0.28 and 0.26 rad when the carrier frequency was reduced to 60. By contrast, our method produced a consistently more accurate phase reconstruction with the phase error of 0.10 rad. In addition, to find appropriate patterns, we suggest choosing a fringe with high frequency and adequate density, but which will not affect the contrast of captured patterns. [Section 5 of the Supplementary Material](#) provides detailed information on the selection of the optimal frequency for the network training.

Finally, to quantitatively determine the accuracy of the learned phase after converting to the desired physical quantity, i.e., 3-D shape of the object, we measured a pair of standard ceramic spheres whose shapes have been calibrated based on a coordinate measurement machine. Figure 7(a) shows the tested ceramic spheres. Their radii are 25.398 and 25.403 mm, respectively, and their center-to-center distance is 100.069 mm. We calculated the 3-D point cloud from the phase obtained by the proposed method and then fitted the 3-D points into the sphere model. The reconstructed result is shown in Fig. 7(b), where the “jet” colormap is used to represent the data values of reconstruction errors. The radii of reconstructed spheres are 25.413 and 25.420 mm, with deviations of 15 and 17 μm , respectively. The measured center-to-center distance is 100.048 mm, with an error of $-21 \mu\text{m}$. As the measured dimensions are very close to the ground truth, this experiment

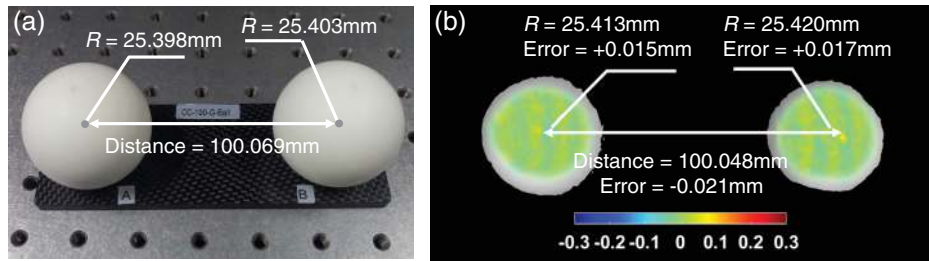


Fig 7 Quantitative analysis of the reconstruction accuracy of the proposed method. (a) Measured objects: a pair of standard spheres and (b) 3-D reconstruction result showing the measurement accuracy.

demonstrates that our method not only provides reliable phase information using only a single fringe pattern but also facilitates high-accuracy single-shot 3-D measurements.

In this letter, we have demonstrated how deep learning significantly improves the accuracy of phase demodulation from a single fringe pattern. Compared with existing single-frame approaches, this deep-learning-based technique provides a framework in fringe analysis by rapidly predicting the background image and estimating the numerator and the denominator for the arctangent function, resulting in high-accuracy edge-preserving phase reconstruction without any human intervention. The effectiveness of the proposed method has been verified using carrier fringe patterns under the scenario of fringe projection profilometry. We believe that, after appropriate training with different types of data, the proposed network framework or its derivation should also be applicable to other forms of fringe patterns (e.g., exponential phase fringe patterns or closed fringe patterns) and other phase measurement techniques for immensely promising applications.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (61722506, 61705105, and 11574152), the National Key R&D Program of China (2017YFF0106403), the Outstanding Youth Foundation of Jiangsu Province (BK20170034), the China Postdoctoral Science Foundation (2017M621747), and the Jiangsu Planned Projects for Postdoctoral Research Funds (1701038A).

References

1. T. Kreis, *Handbook of Holographic Interferometry: Optical and Digital Methods*, John Wiley & Sons, Hoboken, New Jersey (2006).
2. P. K. Rastogi, *Digital Speckle Pattern Interferometry & Related Techniques*, John Wiley & Sons, Hoboken, New Jersey (2000).
3. S. S. Gorthi and P. Rastogi, "Fringe projection techniques: whither we are?," *Opt. Lasers Eng.* **48**(2), 133–140 (2010).
4. C. Zuo et al., "Phase shifting algorithms for fringe projection profilometry: a review," *Opt. Lasers Eng.* **109**, 23–59 (2018).
5. X. Su and Q. Zhang, "Dynamic 3-D shape measurement method: a review," *Opt. Lasers Eng.* **48**(2), 191–204 (2010).
6. Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: principles, applications and implementations," *Opt. Lasers Eng.* **45**(2), 304–317 (2007).
7. J. Zhong and J. Weng, "Spatial carrier-fringe pattern analysis by means of wavelet transform: wavelet transform profilometry," *Appl. Opt.* **43**(26), 4993–4998 (2004).
8. L. Huang et al., "Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase extraction from a single fringe pattern in fringe projection profilometry," *Opt. Lasers Eng.* **48**(2), 141–148 (2010).
9. Z. Zhang et al., "Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase calculation at discontinuities in fringe projection profilometry," *Opt. Lasers Eng.* **50**(8), 1152–1160 (2012).
10. A. Sinha et al., "Lensless computational imaging through deep learning," *Optica* **4**(9), 1117–1125 (2017).
11. Y. Rivenson et al., "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Sci. Appl.* **7**, 17141 (2018).
12. C. Zuo et al., "Temporal phase unwrapping algorithms for fringe projection profilometry: a comparative review," *Opt. Lasers Eng.* **85**, 84–103 (2016).
13. C. Zuo et al., "High-speed three-dimensional profilometry for multiple objects with complex shapes," *Opt. Express* **20**(17), 19493–19510 (2012).

Shijie Feng received his PhD in optical engineering at Nanjing University of Science and Technology. He is an associate professor at Nanjing University of Science and Technology. His research interests include phase measurement, high-speed 3D imaging, fringe projection, machine learning, and computer vision.

Qian Chen received his BS, MS, and PhD degrees from the School of Electronic and Optical Engineering, Nanjing University of Science and Technology. He is currently a professor and a vice principal of Nanjing University of Science and Technology. He has been selected as Changjiang Scholar Distinguished Professor. He has broad research interests around photoelectric imaging and information processing, and has authored more than 200 journal papers. His research team develops novel technologies and systems for mid-/far-wavelength infrared thermal imaging, ultrahigh sensitivity low-light-level imaging, noninterferometric quantitative phase imaging, and high-speed 3D sensing and imaging, with particular applications in national defense, industry, and bio-medicine. He is a member of SPIE and OSA.

Guohua Gu received his BS, MS, and PhD degrees at Nanjing University of Science and Technology. He is a professor at Nanjing University of Science and Technology. His research interests include optical 3D measurement, fringe projection, infrared imaging, and ghost imaging.

Tianyang Tao received his BS degree at Nanjing University of Science and Technology. He is a fourth-year PhD student at Nanjing University of Science and Technology. His research interests include multiview optical 3D imaging, computer vision, and real-time 3D measurement.

Liang Zhang received his BS and MS degrees at Nanjing University of Science and Technology. He is a fourth-year PhD student at Nanjing

University of Science and Technology. His research interests include high-dynamic-range 3D imaging and computer vision.

Yan Hu received his BS degree at Wuhan University of Technology. He is a fourth-year PhD student at Nanjing University of Science and Technology. His research interests include microscopic imaging, 3D imaging, and system calibration.

Wei Yin is a second-year PhD student at Nanjing University of Science and Technology. His research interests include deep learning, high-speed 3D imaging, fringe projection, and computational imaging.

Chao Zuo received his BS and PhD degrees from Nanjing University of Science and Technology (NJUST) in 2009 and 2014, respectively. He was a research assistant at Centre for Optical and Laser Engineering, Nanyang Technological University from 2012 to 2013. He is now a professor at the Department of Electronic and Optical Engineering and the principal investigator of the Smart Computational Imaging Laboratory (www.scilaboratory.com), NJUST. He has broad research interests around computational imaging and high-speed 3D sensing, and has authored over 100 peer-reviewed journal publications. He has been selected into the Natural Science Foundation of China (NSFC) for Excellent Young Scholars and the Outstanding Youth Foundation of Jiangsu Province, China. He is a member of SPIE, OSA, and IEEE.