Phase unwrapping based on deep learning in light field fringe projection 3D measurement*

ZHU Xinjun[1,2], ZHAO Haichuan[1,2], YUAN Mengkai[2,3], ZHANG Zhizhi[1,2], WANG Hongyi[1,2], and SONG Limei[2,3]**

1. School of Artificial Intelligence, Tiangong University, Tianjin 300387, China

2. Tianjin Key Laboratory of Intelligent Control of Electrical Equipment, Tiangong University, Tianjin 300387, China

3. School of Control Science and Engineering, Tiangong University, Tianjin 300387, China

(Received 6 January 2023; Revised 16 April 2023)

Phase unwrapping is one of the key roles in fringe projection three-dimensional (3D) measurement technology. We propose a new method to achieve phase unwrapping in camera array light filed fringe projection 3D measurement based on deep learning. A multi-stream convolutional neural network (CNN) is proposed to learn the mapping relationship between camera array light filed wrapped phases and fringe orders of the expected central view, and is used to predict the fringe order to achieve the phase unwrapping. Experiments are performed on the light field fringe projection data generated by the simulated camera array fringe projection measurement system in Blender and by the experimental 3×3 camera array light field fringe projection system. The performance of the proposed network with light field wrapped phases using multiple directions as network input data is studied, and the advantages of phase unwrapping based on deep learning in light filed fringe projection are demonstrated.

Document code: A **Article ID:** 1673-1905(2023)09-0556-7

DOI https://doi.org/10.1007/s11801-023-3002-4

Structured light three-dimensional (3D) measurement technology has attracted wide attention in recent years for their ability to provide 3D information with simplicity, accuracy, high speed and non-contact property[1-3]. The 3D information is related to the phase in the deformed fringe pattern, and the phase is recovered by phase retrieval operator. However, the retrieval wrapped phase is truncated at $(-\pi, \pi]$, which requires phase unwrapping to obtain a continuous phase[4-9].

With the successful applications of deep learning in the field of fringe projection 3D measurement[10-14], the phase unwrapping method based on convolutional neural network (CNN) has attracted more and more attention. Generally speaking, as to the deep learning based phase unwrapping methods, the first kind method is related to the single camera fringe projection 3D measurement system. For instance, for the prediction problem of wrapped phase to fringe orders, the most typical way to solve the problem with classification ideas based on deep learning is PhaseNet 2.0 proposed by SPOORTHI[15]. YIN et al[16] used the three-step phase shift method to obtain the wrapped phases corresponding to two different frequency gratings and used them as the input of the network. On contrast to the first kind of deep learning phase unwrapping methods, another kind is based on binocular stereo cameras fringe projection. QIAN et al[17] built a multi-channel input network, and the fringe patterns captured by two cameras, as well as the reference information are fed into network, and the output of network is the fringe order map of the measured object in left camera.

In addition, light filed can record the position and direction information of the light ray at the same time to realize multi-viewpoint imaging[18,19]. CAI et al[20] combined fringe projection structured light and light field camera imaging to propose a structured light field phase unwrapping method, which encodes the light field information and determines the fringe orders by comparing the phase consistency of the spatial candidate points in the light field. WANG et al[21] proposed a phase unwrapping method for the fringe projection 3D system assisted by the light field to realize the phase unwrapping of isolated complex objects.

As mentioned above, deep learning based phase unwrapping is only devoted to single camera based or binocular cameras based fringe projection 3D measurement. To the best of our knowledge, deep learning has not been used in fringe projection 3D measurement based on light field camera. We introduce the deep learning method in

* This work has been supported by the National Natural Science Foundation of China (No.61905178), the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (No.2019KJ021), and the Natural Science Foundation of Tianjin (No.18JCQNJC71100).

** E-mail: xinjunzhu@tiangong.edu.cn

the light field fringe projection phase unwrapping. In this letter, from the perspective of light field multi-view phase unwrapping, making full use of the constraint information between camera array sub-aperture images, the convolution neural network models from light field multi-view wrapped phases to fringe orders prediction are developed and studied, and a camera array light field phase unwrapping method based on deep learning is proposed.

In this study, the desired light field data were obtained by using the simulated camera array fringe projection measurement system designed in Blender software and the experimental setup. The light field fringe projection data acquisition system will get 12 groups projected fringe patterns containing four steps shift of three different frequencies, where the first 4 groups of patterns are used to calculate the sub-aperture wrapped phase. No preprocessing such as filtering was used in the process from fringe patterns to wrapped phase calculations. In order to show the generation process of wrapped phases and fringe order in light field fringe projection, Fig.1 takes the 9×9 camera array model as an example to illustrate the calculation process of the sub-aperture wrapped phases in the four directions. The calculation formula of wrapped phases is

$$\varphi(x,y)=\tan^{-1}\left(\frac{I_4(x,y)-I_2(x,y)}{I_1(x,y)-I_3(x,y)}\right), \tag{1}$$

where $\varphi(x, y)$ is the wrapped phase, $I_1(x, y)$, $I_2(x, y)$, $I_3(x, y)$ and $I_4(x, y)$ represent the projected fringe patterns with phase shifts of 0, $\pi/2$, π and $3\pi/2$, respectively.

The wrapped phase at the center view of the camera array is calculated using the multi-frequency phase-shifting method[22], and the corresponding fringe order of the unwrapped phase at that position is obtained and used as the label data for the network. The calculation formula is

$$\phi(x,y)=\varphi_1(x,y)+2\pi k(x,y), \tag{2}$$

where $\phi(x, y)$ is the unwrapped phase, $\varphi_1(x, y)$ is the wrapped phase, and $k(x, y)$ is the fringe order.

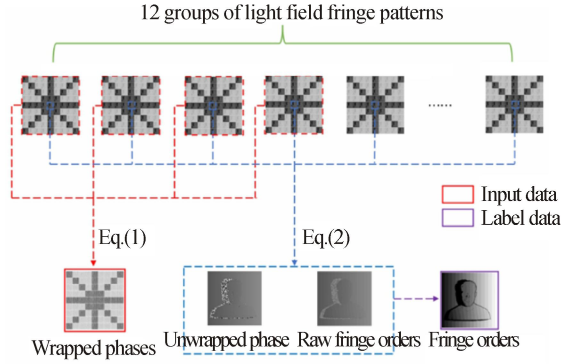

Fig.1 Data preparation of input data and label data in camera array light field fringe projection

In this article, the object is projected by fringe structured light, and the projected fringe patterns with different views are captured by a camera array based light field camera. Extracting features in different directions is advantageous for the network model to effectively learn the mapping relationship between sub-aperture wrapped phase images and central view fringe orders, given the consistency of sub-aperture images in four directions. The network model consists of two parts that are initial feature extraction network and U-shaped network feature fusion model. The initial feature extraction network fully extracts the features of the multi-direction sub-aperture wrapped phases of the camera array and summarizes them as features in the four directions of the camera array, and then outputs them to the feature fusion network. The U-shaped network feature fusion model realizes the fusion processing of the sub-aperture wrapped phases feature information of the camera array and further extracts the features. Through the proposed network model, the mapping relationship between the input multi-view wrapped phases and the output fringe orders of the center view is established, and the prediction from the sub-aperture wrapped phases to the fringe orders is realized, to achieve the phase unwrapping.

Fig.2 shows the initial feature extraction network, which utilizes multiple input branches to process the sub-aperture wrapped phase of each direction and extract feature information. Then output the phase features wrapped by sub apertures in the four directions of the camera array. Taking the network for the feature extraction of sub-aperture wrapped phase in the direction of 0° as an example, its network structure is shown in Fig.3. The detail features are extracted from the 9 light field sub-aperture wrapped phase images in the direction of 0° through 9 branches respectively, and then the features in the direction of 0° are simply fused by a convolution block. The feature extraction network for the feature extraction of sub-aperture wrapped phases in the direction of 90°, 45° and 135° has the same structure as above. Specifically, two convolution blocks are used in each branch, and 64 convolution kernels of 3×3 are used. The step size and padding are set to 1. The design of the convolutional block refers to the convolutional block structure adopted in the input branch of EPINET.

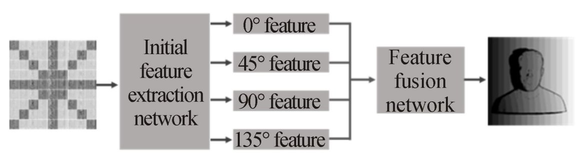

Fig.2 The method proposed for phase unwrapping in light field fringe projection

Once the initial feature extraction is done, the extracted features from the four directions of the camera array are combined and forwarded to the feature fusion network for further processing, as depicted in Fig.4. Finally, the target fringe order map is output in the form of classification.

This part of work is completed by the recurrent residual CNN based on U-Net (R2U-Net). R2U-Net is implemented by replacing U-Net's encoding and decoding convolution unit with a more complex recurrent residual convolutional unit (RRCU). After each RRCU in the encoder network, there is a max pooling layer with a stride of 2×2 to reduce the spatial resolution of the feature maps. After each RRCU in the decoder network, an upsampling (transposed convolution) layer is used to restore the input spatial resolution. The kernel size of the convolutional layers in RRCU is 3×3. The number of output channels of the network output convolutional layer is determined by the range of the fringe order, and a 1×1 convolution kernel is used.

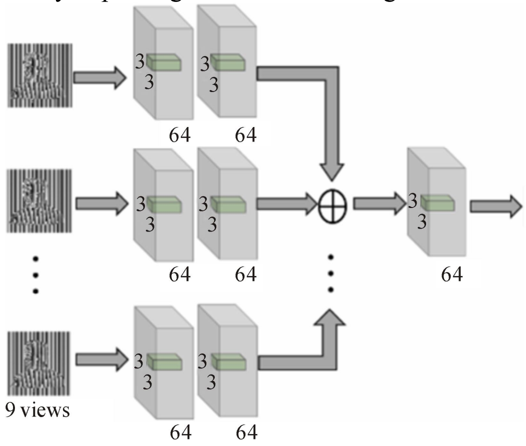

Fig.3 Diagram of the initial feature extraction network in Fig.2

As a classification task, the data output by the model with a dimension of batch 17×512×512 needs to go through a softmax operation to determine the category with the highest probability at each pixel, that is, the fringe order. In order to achieve the train of the proposed model, the cross-entropy loss function is used.

The experiment was conducted on a PC equipped with NVIDIA GeForce RTX 3090 (24 GB), Intel Core i7-10700K @ 3.80 GHz 8 cores and 64 GB RAM for network training and testing. The network architecture is implemented based on the Pytorch framework version 1.7.0+cu110 of Python 3.8.3.

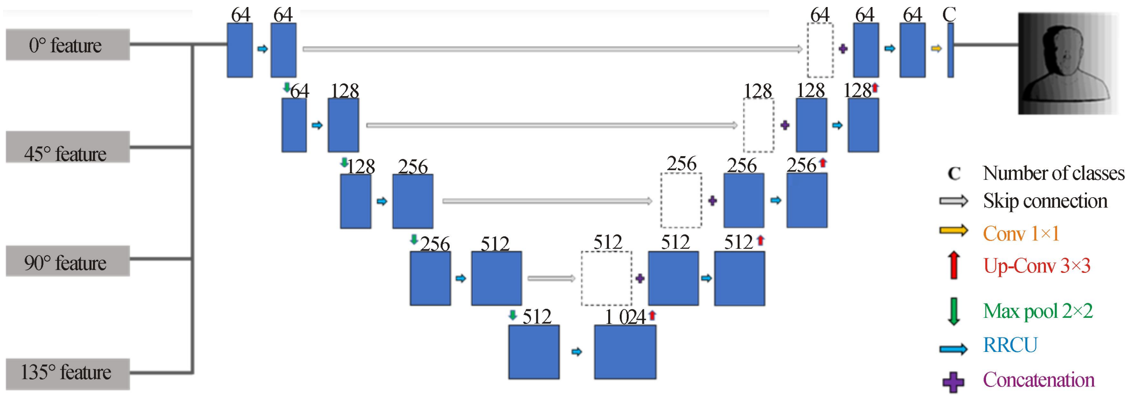

Fig.4 R2U-Net feature fusion model

For the Blender simulation, the camera pitch is 80 mm and the focal length is 100 mm in camera array. A total of 12 fringe patterns with three frequencies and four-step phase shifts for obtaining phase information are loaded into the projector. The deformed fringe patterns on test objects are captured by camera array. In the process of data generation, a total of 40 groups of data were collected from multiple angles of 13 computer aided drafting (CAD) model objects, among which 30 groups were used for training, 5 groups were used for verification and the remaining 5 groups were used for testing. According to the range of fringe order of the simulation data, the number of output channels at the last layer of the network is set as 17, in which there are 16 fringe order categories, and noise is divided into another category. These models have complex shapes and rich details to verify the performance of the proposed method.

Fig.5 shows the schematic diagram of one direction (1-direction), two directions (2-directions), and four directions (4-directions) as the network input.

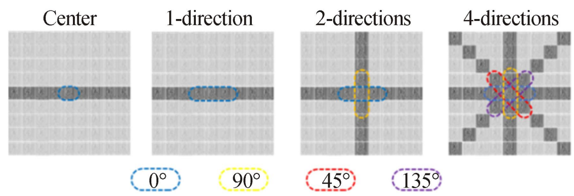

Fig.5 Schematic diagram of multi-directional data

For a more comprehensive comparison, a single-view (center view of camera array) training method is added to the comparison, that is, only the single frame wrapped phase at the center view of camera array is used for training and predicting the corresponding fringe order. Due to the number of network parameters on the computer graphics card memory demand is large, only 3×3 view in the center area of 9×9 view of optical field data is selected as the input of the network in this section.

The network training batch size is set to 2, the number of training times is 100, and the initial learning rate is set to 1×10^{-4} and attenuates to 1×10^{-5} at the 95th time of training. The data is enhanced by rotating 90° in training. After each training session of the training set, the prediction accuracy of the validation set is evaluated based on mean square error (*MSE*). The calculation method of *MSE* is

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(y_m - \hat{y}_m)^2. \quad (3)$$

Tab.1 shows the prediction result errors of the four directions, two directions, a single direction, and a single view training method on five test data. It can be seen that the sub-aperture wrapped phase input data in multiple directions can better improve the performance of the proposed multi-stream convolutional network model, and more accurate phase unwrapping can be realized.

Tab.1 Fringe order prediction error comparison under multi-directional network input on simulated camera array light filed fringe projection data

Input views	Center	1-direction	2-directions	4-directions
MSE	0.851 7	0.519 2	0.499 6	0.459 8

As shown in Fig.6, based on two test samples, the prediction results of the network under the above four input data forms are compared with the labels. It can be seen from the comparison diagram that when the sub-aperture wrapped phases in the four directions of the camera array are used as the network input data, the fringe order predicted by the network is more accurate. Compared with other schemes, the predicted values in the four directions have the highest degree of fit with the label, and basically coincide with the label curve.

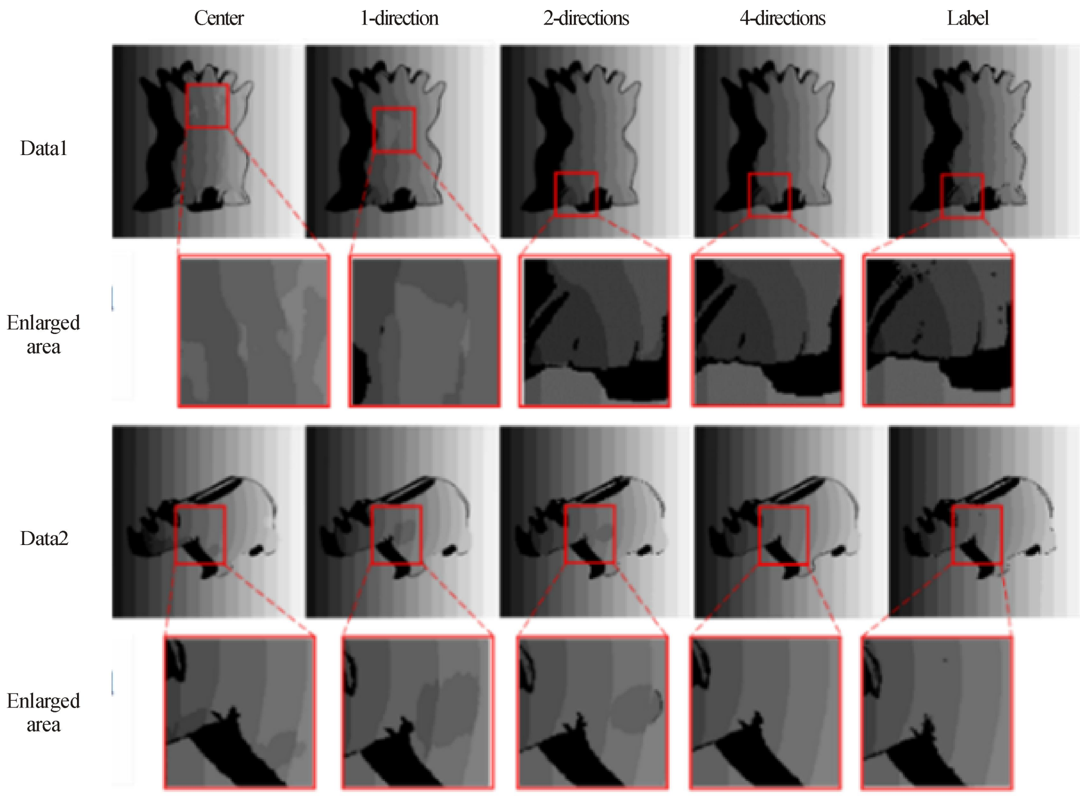

Fig.6 Comparison of multi-directional data prediction results

In order to further verify the proposed method and apply it to measurement tasks in the real environment, this section conducts experiments on the fringe pattern data collected by the 3×3 camera array fringe projection 3D measurement system built in the laboratory environment. The established camera array light field fringe projection 3D measurement system is shown in Fig.7. In the proposed system, the 3×3 cameras simultaneously triggered by the projector can capture the deformed fringe pattern on the tested objects projected by the projector.

In the process of data generation, 57 groups of data were collected from multiple angles of 7 plaster model objects, among which 32 groups were used for training, 7 groups were used for verification, and the remaining 18 groups were used for testing. According to the range of fringe order of experimental data, the number of output channels at the last layer of the network is set as 23, which means that there are 22 fringe order categories, and background and noise are divided into another category. The fringe patterns of some plaster models collected are shown in Fig.8.

In the training of this section, in order to form a significant contrast in the experimental results and improve the experimental efficiency, the feature fusion network

R2U-Net was simplified. The simplification was realized by reducing the number of encoders and corresponding decoders of R2U-Net from the original four to two. During the network training, the batch size was set as 2, the training times as 60, and the initial learning rate was set as 1×10^{-4} and attenuated to 1×10^{-5} in the 50th training.

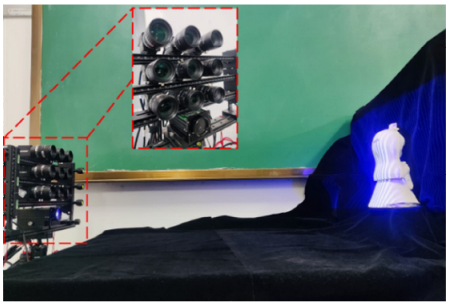

Fig.7 Real camera array light field fringe projection 3D system

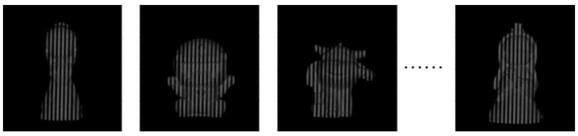

Fig.8 Part of the plaster model fringe patterns

Tab.2 shows the fringe order prediction result errors of 4-directions, 2-directions, 1-direction, and a single view (center view) training method on 18 test data. The test error of the 4-directionsas input is the smallest, which is consistent with the experimental conclusions of the simulation scene data camera array multi-direction sub-aperture images analysis section.

Tab.2 Fringe order prediction error comparison under multi-directional network input on real camera array light filed fringe projection data

Input views	Center	1-direction	2-directions	4-directions
MSE	0.531 7	0.467 8	0.423 8	0.392 1

As shown in Fig.9, the prediction results of the network of two tested samples under these above four input data forms are compared with the labels. It can be seen that the fringe order predicted by the network in the 4-directions input form is the most accurate, giving the least wrong fringe order areas. In addition, a certain line of the fringe orders (the line marked by dotted line in label image) is shown in Fig.10, where the predicted fringe order values of 1-direction, 2-directions, and 4-directions and the label are plotted. Compared with other form as input, the predicted values in the 4-directions as input have the higher degree of fitness with the label, which is closer with the label curve. Therefore, the real scene data results demonstrate the effectiveness of the method, and that the more directions light field wrapped phases as input of the network improve the accuracy of the fringe order prediction results.

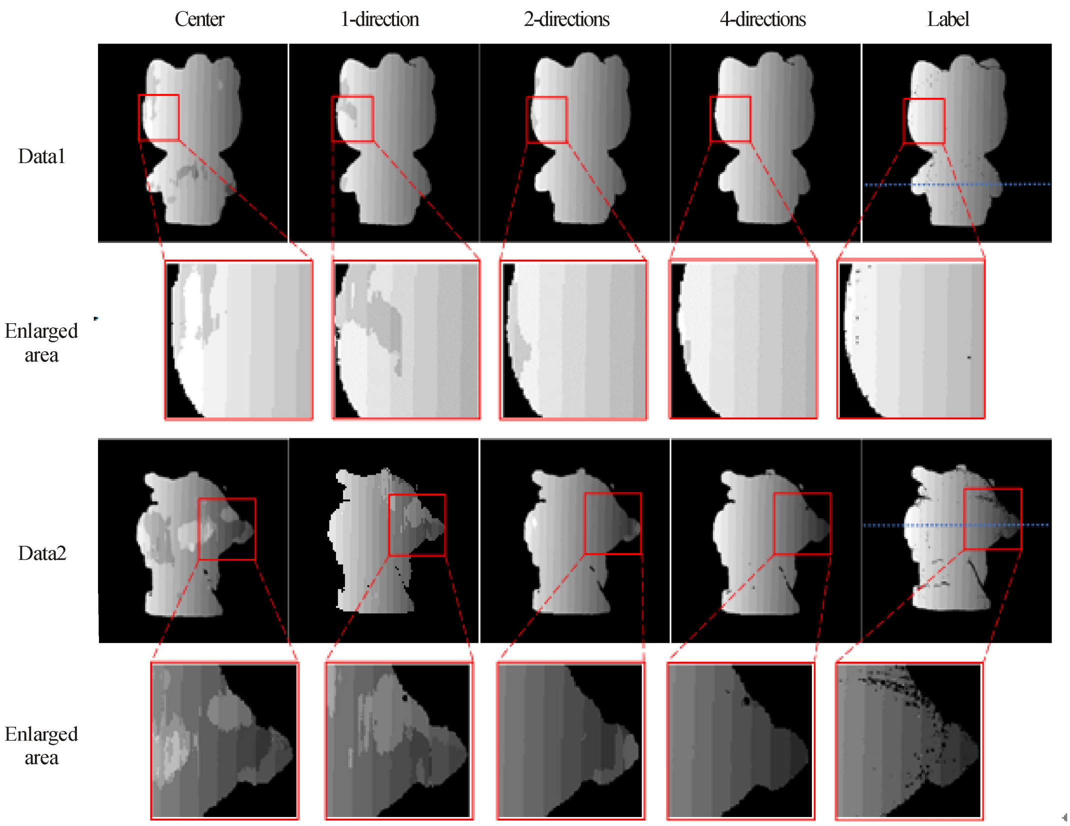

Fig.9 Fringe order prediction results on real data

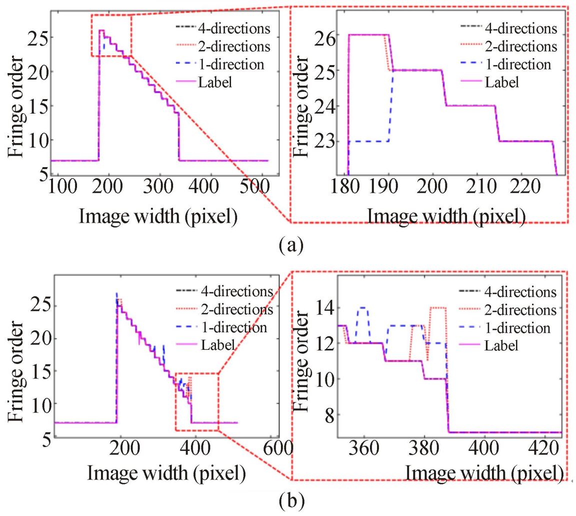

Fig.10 Comparison of prediction results of a row in the real dataset: (a) Data1; (b) Data2

Fig.11 shows the comparison results of the unwrapped phase obtained by the proposed multi-stream CNN and traditional quality guided methods. It can be seen that the results obtained using our proposed method are closer to the true unwrapped phase (multi-frequency phase-shifting method as ground truth), compared with the quality guided method. Due to the inability of traditional quality guided phase unwrapping methods to hand with multiple isolated object phase unwrapping, the unwrapped phase obtained by quality guided method is higher than the true value.

In order to verify the network's noise resistance performance, Gaussian noise with zero mean and variance of 0.02 is added on the original dataset and the noisy dataset is evaluated. Fig.12 shows the test results of the proposed network's anti-noise performance. It can be concluded that our proposed network has good noise resistance in phase unwrapping.

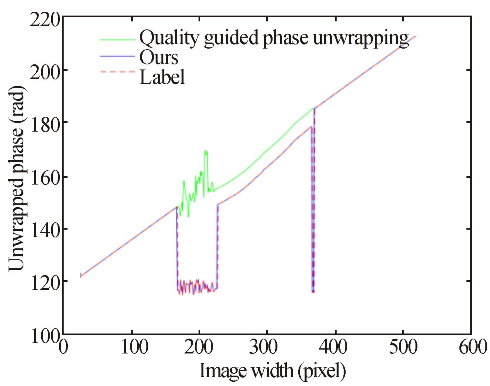

Fig.11 Unwrapped phase result comparison with the quality guided phase unwrapping method

In this letter, a new method based on deep learning is proposed to phase unwrapping in camera array light field fringe projection 3D measurement. The advantages of phase unwrapping based on deep learning to learn the mapping relationship between camera array light filed wrapped phases and fringe orders of the expected central view are demonstrated. Experimental results show that when using the wrapped phases in four directions as network input simultaneously, the better performance of the proposed multi-stream CNN model can be achieved and effective phase unwrapping can be realized. Our proposed method complements the existing deep learning phase unwrapping methods with only one camera or two cameras and paves a way to the deep learning based phase unwrapping.

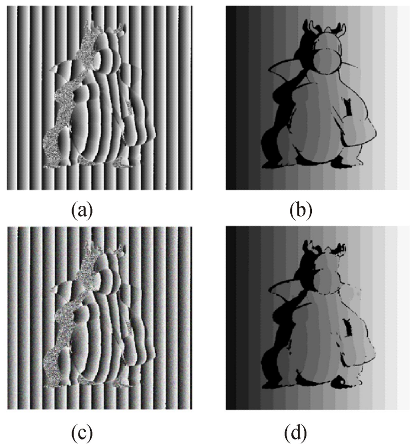

Fig.12 Network input data and prediction results: (a) Input without noise; (b) Output from (a); (c) Input with noise; (d) Output from (c)

Ethics declarations

Conflicts of interest

The authors declare no conflict of interest.

References

[1] CHEN R, XU J, ZHANG S. Comparative study on 3D optical sensors for short range applications[J]. Optics and lasers in engineering, 2022, 149：106763.

[2] GENG J. Structured-light 3D surface imaging：a tutorial[J]. Advances in optics and photonics, 2011, 3(2)：128-160.

[3] MARRUGO A G, GAO F, ZHANG S. State-of-the-art active optical techniques for three-dimensional surface metrology：a review[J]. Journal of the Optical Society of America A, 2020, 37(9)：B60-B77.

[4] ZUO C, FENG S J, HUANG L, et al. Phase shifting algorithms for fringe projection profilometry：a review[J]. Optics and lasers in engineering, 2018, 109：23-59.

[5] ZHANG S. High-speed 3D shape measurement with structured light methods：a review[J]. Optics and lasers in engineering, 2018, 106：119-131.

[6] YIN W, ZUO C, FENG S J, et al. High-speed three-dimensional shape measurement using geometry-constraint-based number-theoretical phase unwrapping[J].

Optics and lasers in engineering, 2019, 115：21-31.

[7] PISTELLATO M, BERGAMASCO F, ALBARELLI A, et al. Robust phase unwrapping by probabilistic consensus[J]. Optics and lasers in engineering, 2019, 121：428-440.

[8] ZHANG S. Absolute phase retrieval methods for digital fringe projection profilometry：a review[J]. Optics and lasers in engineering, 2018, 107：28-37.

[9] AN H, CAO Y, WU H, et al. Spatial-temporal phase unwrapping algorithm for fringe projection profilometry[J]. Optics express, 2021, 29(13)：20657-20672.

[10] FENG S J, CHEN Q, GU G, et al. Fringe pattern analysis using deep learning[J]. Advanced photonics, 2019, 1(2)：025001-025001.

[11] ZHENG Y, WANG S D, LI Q, et al. Fringe projection profilometry by conducting deep learning from its digital twin[J]. Optics express, 2020, 28(24)：36568-36583.

[12] SHI J S, ZHU X J, WANG H Y, et al. Label enhanced and patch based deep learning for phase retrieval from single frame fringe pattern in fringe projection 3D measurement[J]. Optics express, 2019, 27(20)：28929-28943.

[13] NGUYEN H, WANG Y Z, WANG Z Y. Single-shot 3D shape reconstruction using structured light and deep convolutional neural networks[J]. Sensors, 2020, 20(13)：3718.

[14] MACHINENI R C, SPOORTHI G E, VENGALA K S, et al. End-to-end deep learning-based fringe projection framework for 3D profiling of objects[J]. Computer vision and image understanding, 2020, 199：103023.

[15] SPOORTHI G E, GORTHI R K S S, GORTHI S. PhaseNet 2.0：phase unwrapping of noisy data based on deep learning approach[J]. IEEE transactions on image processing, 2020, 29：4862-4872.

[16] YIN W, CHEN Q, FENG S J, et al. Temporal phase unwrapping using deep learning[J]. Scientific reports, 2019, 9(1)：1-12.

[17] QIAN J M, FENG S J, TAO T Y, et al. Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3D shape measurement[J]. APL photonics, 2020, 5(4)：046105.

[18] ORTH A, CROZIER K B. Light field moment imaging[J]. Optics letters, 2013, 38(15)：2666-2668.

[19] TAO T Y, CHEN Q, ZHANG Y Z, et al. Multi-view phase unwrapping with composite fringe patterns[C]//International Conference on Optical and Photonics Engineering (icOPEN 2016), September 26-30, 2016, Chengdu, China. Washington：SPIE, 2017, 10250：240-245.

[20] CAI Z W, LIU X L, CHEN Z Z, et al. Light-field-based absolute phase unwrapping[J]. Optics letters, 2018, 43(23)：5717-5720.

[21] WANG Z W, YANG Y, LIU X L, et al. Light-field-assisted phase unwrapping of fringe projection profilometry[J]. IEEE access, 2021, 9：49890-49900.

[22] SONG L M, DONG X X, XI J T, et al. A new phase unwrapping algorithm based on three wavelength phase shift profilometry method[J]. Optics & laser technology, 2013, 45：319-329.