Wavelet based deep learning for depth estimation from single fringe pattern of fringe projection profilometry*

ZHU Xinjun, HAN Zhiqiang, SONG Limei, WANG Hongyi, and WU Zhichao**
School of Artificial Intelligence, Tiangong University, Tianjin 300387, China

(Received 19 May 2022; Revised 7 August 2022)

Depth estimation from single fringe pattern is a fundamental task in the field of fringe projection three-dimensional (3D) measurement. Deep learning based on a convolutional neural network (CNN) has attracted more and more attention in fringe projection profilometry (FPP). However, most of the studies focus on complex network architecture to improve the accuracy of depth estimation with deeper and wider network architecture, which takes greater computational and lower speed. In this letter, we propose a simple method to combine wavelet transform and deep learning method for depth estimation from the single fringe pattern. Specially, the fringe pattern is decomposed into low-frequency and high-frequency details by the two-dimensional (2D) wavelet transform, which are used in the CNN network. Experiment results demonstrate that the wavelet-based deep learning method can reduce the computational complexity of the model by 4 times and improve the accuracy of depth estimation. The proposed wavelet-based deep learning models (UNet-Wavelet and hNet-Wavelet) are efficient for depth estimation of single fringe pattern, achieving better performance than the original UNet and hNet models in both qualitative and quantitative evaluation.
Document code: A **Article ID:** 1673-1905(2022)11-0699-6
DOI https://doi.org/10.1007/s11801-022-2082-x

Depth estimation from the single fringe pattern of fringe projection profilometry (FPP) is still a challenging problem in many practical applications including machine vision, three-dimensional (3D) printing, medicine, etc[1-8]. In recent years, deep learning based on a convolutional neural network (CNN) has attracted more and more attention in FPP. YU et al[9] proposed a new geometric constraint-based phase unwrapping (GCPU) method that enabled an untrained deep learning-based FPP for the first time. SPOORTHI et al[10] formulated the phase unwrapping problem as a dense classification problem and proposed a fully convolutional network PhaseNet 2.0 trained to predict the wrap-count at each pixel from the wrapped phase maps. SAM et al[11] proposed a CNN model to depth estimation from single frame fringe pattern, where simulated fringe projection data are used to extract height information from single deformed fringe pattern. NGUYEN et al[12] proposed a robust method integrating the structured light technique with the CNN with experimental data to estimate depth from fringe or speckle data. Later, NGUYEN et al[13] proposed a global guidance network path with multi-scale feature fusion introduced into the CNN model to estimate the depth of a single fringe pattern. YUAN et al[14] enhanced the ability to capture the global context of a complex object by using a recurrent residual network. JIA et al[15] proposed a novel depth measurement method based on a CNN, which was cast as a pixel-wise classification-regression task without matching to estimate speckle structured light images. WANG et al[16] proposed a dual-path hybrid network based on UNet, which fuses the CNN path and a swin transformer path to improve the global perception of traditional CNN based networks.

Most of the studies focus on complex network architectures to improve the accuracy of depth estimation, which takes greater computational and lower speed. As a traditional image processing technique, wavelet has been explored in deep learning based computer vision tasks, such as image super-resolution[17,18], denoising[19], demoiréing[20], etc. XUE et al[18] proposed a wavelet-based residual attention network for image super-resolution, which reduces the training difficulty by explicitly decomposing low-frequency and high-frequency details into four channels. LIU et al[19] proposed a densely self-guided wavelet network for real world image denoising, which can efficiently incorporate multi-scale information and extract good local features to recover clean images. LIU et al[20] proposed a wavelet-based dual-branch network for image demoireing, which removes Moire patterns in the wavelet domain. As mentioned above, wavelet transform is efficient to depict contextual and textural information, which inspires us to introduce wavelet transform to FPP for deep estimation.

In this letter, our work focuses on utilizing wavelet

* This work has been supported by the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (No.2019KJ021).
** E-mail: xinjunzhu@tiangong.edu.cn

transform to enhance the CNN-based model to extract low-frequency structure and high-frequency details for depth estimation from the single fringe pattern. We built a plug-and-play two-dimensional (2D) wavelet transform module and 2D inverse wavelet transform module. In the modules, 2D wavelet transform can be easily inserted into various convolutional models to explicitly decompose the fringe pattern into coarse content and sharp details and the 2D inverse wavelet transform be easily inserted to merge the output features into a depth map, so that the training cost of the model can be reduced. The proposed method is validated by simulated and real fringe patterns dataset. Additional experimental results show that the wavelet transform can improve on depth estimation of a single fringe pattern in a CNN-based model.

In fringe projection, a fringe pattern captured by a CCD can be expressed as

$$I(x,y)=a(x,y)+b(x,y)\cos\left(\varphi(x,y)+2\pi f_0 x\right), \quad (1)$$

where $a(x, y)$ is the background, $b(x, y)$ and $\varphi(x, y)$ are the modulation intensity and the optical phase, and f_0 is the carrier frequency. The depth map $D(x, y)$ of the measured object can be obtained from the system calibration with the estimated phase $\varphi(x, y)$. However, it is a great challenge to obtain $\varphi(x, y)$ from the single fringe pattern $I(x, y)$ which requires the necessary steps of phase retrieval and phase unwrapping. Compared with the traditional models, the model based on deep learning can learn a function $\Re$ to estimate a depth map D as

$$D(x,y)=\Re\left(I(x,y)\right). \quad (2)$$

To validate the wavelet-based deep learning method, we choose two classical CNN models (UNet[12] and hNet[13]) for depth estimation in the field of FPP. The UNet model is mainly made up of two components: encoder and decoder. The encoder includes convolution layers and pooling layers that detect essential features and downsampling the features. The decoder contains transpose convolution layers and unpooling layers using bilinear interpolation operation that can stack and concatenate lower resolution features to form higher resolution features. In addition, the key is that in the UNet the local context information from the encoder is concatenated with the upsampled output by skip connection or residual connection, which is shown to be beneficial in reducing information loss in encoder-decoder architecture.

The hNet architecture is similar to the UNet architecture. However, it comprises three components: encoder, decoder, and the global guidance path. Although the encoder and decoder are the autoencoder-based UNet, a typical autoencoder-based UNet only reconstructs the output feature map with fine-level contextual information that is the same as the input fringe pattern. The hNet architecture proposes using a global guidance path to provide extra global or coarse features to the highest fine-level feature map.

The architectures of UNet and hNet models are shown in Fig.1. To learn the mapping between the input fringe pattern and output depth estimation, a back-propagation algorithm is adopted to minimize the loss and update the model parameters[21]. We compute the difference between the estimated target map D and the ground truth D^* to train this CNN network for depth estimation. We adopt ℓ_2 loss function for supervised learning as

$$loss=\frac{1}{H\times W}\sum_{x\in H}\sum_{y\in W}\left|D(x,y)-D^*(x,y)\right|^2, \quad (3)$$

where H and W are the height and width of the input fringe pattern, respectively.

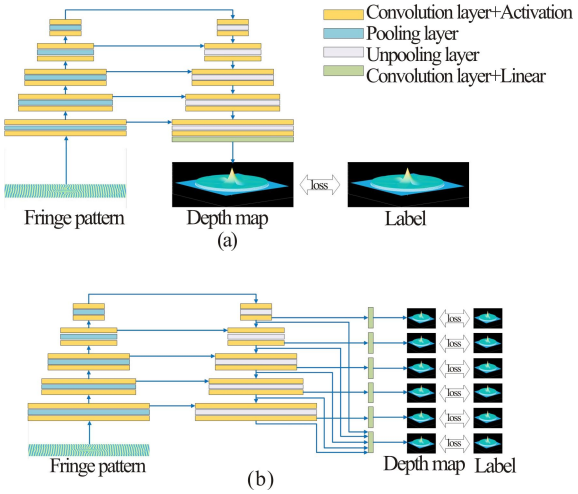


Fig.1 The UNet and hNet network architectures: (a) UNet; (b) hNet

In order to utilize wavelet transform to enhance the CNN-based models to estimate fringe pattern depth, the wavelet transform is introduced in the UNet and hNet models. In this letter, we choose the simplest Haar wavelet as the basis for the wavelet transform to decompose the fringe pattern into a sequence of wavelet coefficients of different frequency contents. The transformer iteratively applies low-pass and high-pass decomposition filters along with downsampling to compute the wavelet coefficients, where the low-pass filter (LF) is $(1/\sqrt{2}, 1/\sqrt{2})$ and the high-pass filter (HF) is $(1/\sqrt{2}, -1/\sqrt{2})$. In each level of the wavelet transform, we use the low-pass filter and the high-pass filter along the columns to transform a fringe pattern into two fringe patterns, and the same filters are used along the rows of these two fringe patterns to generate four fringe patterns as shown in Fig.2. Finally, the output is four coefficients, denoted as {A, V, H, D}. The equations to derive the four coefficients can be found in Ref.[17].

Fig.3 shows the overall structure of our proposed method. The wavelet transform is embedded into the deep learning model to improve the ability to extract high and low-frequency features. The input and output of the model are fringe pattern and depth map respectively. The Conv-bn-LeakyReLU module represents a combination

of convolution operation, an effective technique name batch normalization to decrease inter-covariate shift in networks, and a nonlinear activation function name leaky rectified linear unit (LeakyReLU). The Conv-linear module which represents a 1×1 convolution with linear activation is applied at the final layer to bring the information of the vector feature map to the corresponding 3D label. The max-pooling layers with a 2×2 window and a stride of 2 are applied to downsample the feature maps by extracting only the max value in each window. The transpose Conv layers are applied to transform the lower feature input back to a higher resolution.

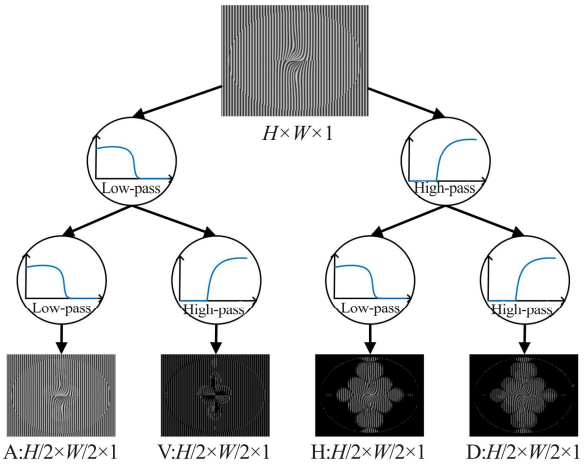


Fig.2 Illustration of the wavelet transform (The fringe pattern is decomposed into four coefficients, A (average), H (horizontal), V (vertical), and D (diagonal))

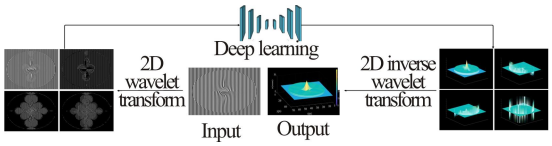


Fig.3 The overall structure of the wavelet-based deep learning method (Input: fringe pattern; Output: depth map)

To demonstrate the effectiveness of the proposed wavelet-based models for depth estimation, we choose our built simulated dataset and one real fringe pattern dataset[13] in the experiments and take the original UNet and hNet depth models for comparisons. For the simulation dataset, simulated projection fringe pattern data with the image sizes of 640×480 pixels can be obtained according to Eq.(1). In the simulation, the parameters in Eq.(1) are set as follows: frequency of the fringe pattern is f_0=1/12, the background illumination is $a(x, y)$=0.02* $\varphi(x, y)$, and the modulation intensity. The phase is simulated using Zernike functions with different Zernike polynomial parameters to generate the different shapes of the fringe pattern. For simplicity, the linear phase-height model is used, where depth data is $h(x, y)=k\varphi(x, y)$ and k=1 in this study.

The proposed model is implemented in PyTorch. The experiments are conducted on NVIDIA GeForce RTX3090 graphics processing unit (GPU) with 24 GB memory. We use a cosine annealing warm restarts learning rate decay policy with an AdamW optimizer. The initial learning rate is 0.005, batch size is 4, and epochs are 200. The evaluation is carried out by calculating the mean absolute error (*MAE*) and the mean squared error (*MSE*) of the reconstructed 3D shapes.

The performances of the UNet, wavelet-based UNet, hNet, and wavelet-based hNet are evaluated with the simulated fringe pattern dataset, the simulated fringe pattern dataset with noise, and the real fringe pattern dataset as shown in Fig.4. It can be seen from the validation loss plot on these datasets that the models based on wavelet transform have lower errors.

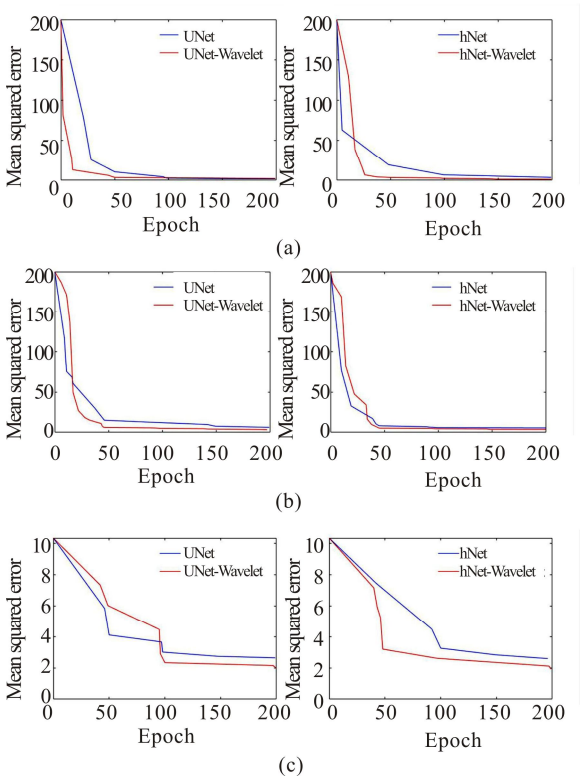


Fig.4 Validation loss plots acquired during the learning process: (a) In the simulated fringe pattern dataset; (b) In the simulated fringe pattern dataset with noise; (c) In the real fringe pattern dataset

We use the other data not included in the training or the validation datasets as the test dataset to evaluate the parameters, computational complexity, and error of these models. The cost of the 2D wavelet transform module is taken into account when computing giga floating-point operations per second (GFLOPS) and computing speed. It can be seen from Tab.1 that the computational complexity of the wavelet-based model is reduced by about four times with almost no increase in the number of parameters. The reduction of computational complexity

shortens the training period of the wavelet-based model by about 4 times, while the increased number of parameters slightly reduces the inference speed. It is well known that downsampling an image will lose part information of the image. This also works for convolution neural networks, especially for dense prediction tasks, the higher the resolution, the better the results. Unfortunately, the computational complexity of CNNs is proportional to the square of the resolution. The model based on wavelet transform can reduce the resolution of the fringe pattern by 1/2 through decomposition to reduce the computational complexity, and the high-frequency and low-frequency details generated by the decomposition can make up for the loss of image information.

Tab.1 Parameters and computational complexity of four models

Model	Params (M)	GFLOPs	Speed (ms/image)
UNet	8.64	76.88	**50.82**
UNet-Wavelet	**8.64**	**19.29**	51.75
hNet	**8.64**	77.01	**52.49**
hNet-Wavelet	8.65	**19.77**	58.23

Tab.2 shows the comparison of the base model and our method in terms of *MAE* and *MSE*. This suggests that models based on wavelet transform can predict and reconstruct 3D shapes from the fringe pattern more accurately than UNet and hNet. Since the fringe pattern has obvious high and low-frequency information, the wavelet-based deep learning method takes the four sub-bands decomposed by the wavelet transform as input and can significantly improve the speed and accuracy of 3D reconstruction. For instance, for the simulated dataset with noise, the *MSE* of the UNet-Wavelet is 4.934 3, while the the *MSE* of the UNet is 9.000 5.

Tab.2 Evaluation metrics of four approaches on three datasets (Dataset 1: simulated fringe dataset without noise; Dataset 2: simulated fringe dataset with noise; Dataset 3: real fringe dataset)

Datasets	Dataset 1		Dataset 2		Dataset 3	
Model	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*
UNet	2.327 4	9.000 5	9.000 5	0.885 0	2.983 6	0.603 9
UNet-Wavelet	**2.261 5**	**4.934 3**	**4.934 3**	**0.822 4**	**2.792 1**	**0.549 9**
hNet	2.553 5	9.844 5	9.844 5	0.927 0	2.718 7	0.536 2
hNet-Wavelet	**1.568 7**	**5.643 1**	**5.643 1**	**0.767 7**	**2.533 7**	**0.534 4**

Fig.5(a), (b), and (c) show the reconstructed depth by UNet and wavelet-based UNet approaches on the above three datasets respectively. Fig.6 shows the enlarged parts of prediction depth results for fringe pattern in red box of Fig.5(c). The depth maps estimated by our method are better than other models in visual quality. Our proposed method preserved more accurate details of the depth map, especially for complex regions, such as the disc part of the simulated fringe pattern dataset and the cat-butt part of the real fringe pattern dataset.

We further validate the performance of the hNet and wavelet-based hNet models on the above three datasets. Fig.7 shows the reconstructed depth by hNet and wavelet-based hNet approaches respectively. Fig.8 shows the enlarged parts of prediction depth results for fringe pattern in red box of Fig.7(c). Like wavelet-based UNet, the wavelet-based hNet also achieves better depth results compared with hNet model.

Further, the plug-and-play 2D wavelet transform module is devoted to DPH model. We verify the proposed DPH-Wavelet with the comparison with the DPH. The DPH achieves an *MSE* of 7.546 6 on the above noisy dataset, while DPH-Wavelet achieves an *MSE* of 5.085 4. One of the samples is shown in Fig.9.

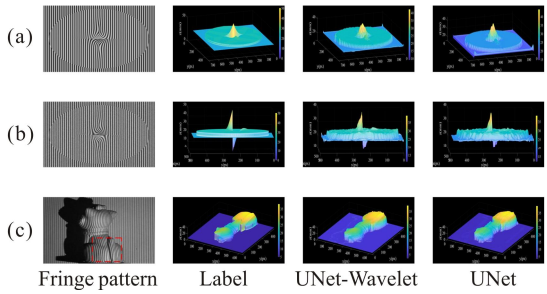
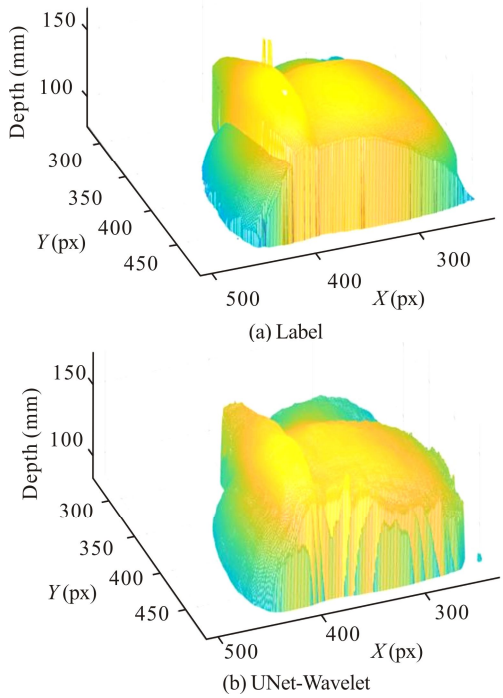


Fig.5 Reconstructed depths by UNet and UNet-Wavelet approaches on the three datasets: (a) Simulated fringe pattern; (b) Simulated fringe pattern with noise; (c) Real fringe pattern

The wavelet image processing algorithm can help the deep learning model to reduce the computational cost. We construct a plug-and-play 2D wavelet transform layer that can be easily inserted into any deep learning



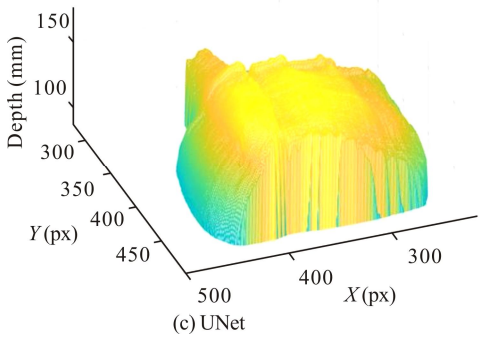


Fig.6 Enlarged parts of prediction results for fringe pattern in red box of Fig.5(c)

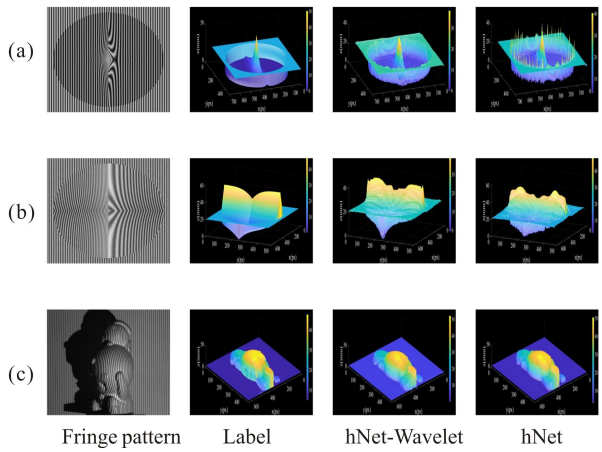
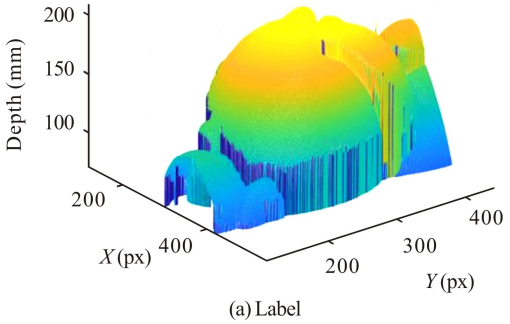
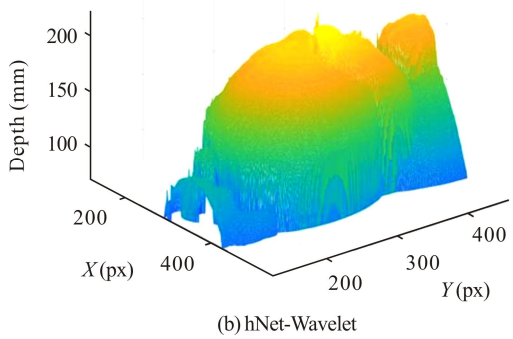


Fig.7 Reconstructed depths by hNet and hNet-Wavelet approaches on the three datasets: (a) Simulated fringe pattern; (b) Simulated fringe pattern with noise; (c) Real fringe pattern





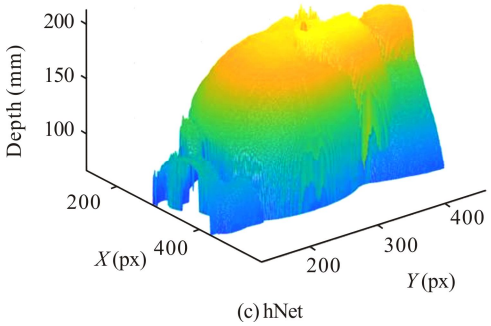


Fig.8 Enlarged parts of prediction results for fringe pattern in red box of Fig.7(c)

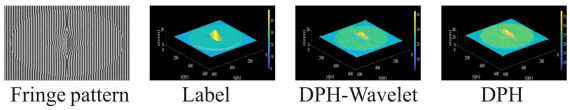


Fig.9 The performance on the simulated fringe pattern with noise by DPH and DPH-Wavelet approaches

model, which reduces the hardware computational cost by decomposing the image with lower resolution in the network model. The experimental results on simulated and real datasets demonstrate that the wavelet based depth estimation model using the 2D wavelet transform layer reduces the training time to about 1/4 times, and the accuracy is also improved. The wavelet transform modules in the proposed UNet-Wavelet, hNet-Wavelet and DPH-Wavelet models for fringe projection depth estimation can also be extended to other deep learning models in future.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] GORTHI S S, RASTOGI P. Fringe projection techniques: whether we are?[J]. Optics and lasers in engineering, 2009, 48(2): 133-140.

[2] SONG L M, LI X Y, YANG Y G, et al. Structured-light based 3D reconstruction system for cultural relic packaging[J]. Sensors (Basel), 2018, 18(9): 2981.

[3] LI B, AN Y, CAPPELLERI D, et al. High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics[J]. International journal of intelligent robotics and applications, 2017, 1(1): 86-103.

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

[5] 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.

[6] ZHANG S. High-speed 3D shape measurement with

structured light methods：a review [J]. Optics and lasers in engineering, 2019, 106：119-131.

[7] SONG L M, GAO Y, ZHU X J, et al. A 3D measurement method based on multi-view fringe projection by using a turntable[J]. Optoelectronics letters, 2016, 12(6)：389-394.

[8] JI Y, CHEN Y, SONG L M, et al. 3D defect detection of connectors based on structured light[J]. Optoelectronics letters, 2021, 17(2)：107-111.

[9] YU H T, HAN B W, BAI L F, et al. Untrained deep learning-based fringe projection profilometry[J]. APL photonics, 2022, 7：016102.

[10] 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.

[11] SAM V D J, JORIS J J D. Deep neural networks for single shot structured light profilometry[J]. Optics express, 2019, 27：17091-17101.

[12] 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.

[13] NGUYEN H, LY K L, TRAN T, et al. hNet：single-shot 3D shape reconstruction using structured light and h-shaped global guidance network[J]. Results in optics, 2021, 4：100104.

[14] YUAN M, ZHU X, HOU L. Depth estimation from single frame fringe projection pattern based on R2U-Net[J]. Laser and optoelectronics progress, 2021：1-19. (in Chinese)

[15] JIA T, et al. Depth measurement based on a convolutional neural network and structured light[J]. Measurement science and technology, 2022, 33：025202.

[16] WANG L, LU D Q, QIU R W, et al. 3D reconstruction from structured-light profilometry with dual-path hybrid network[J]. Eurasip journal on advances in signal processing, 2022, 2022：14.

[17] HUANG H B, HE R, SUN Z N, et al. Wave-let-SRNet：a wavelet-based CNN for multi-scale face super resolution[C]//Proceedings of 2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York：IEEE, 2017：1698-1706.

[18] XUE S K, QIU W Y, LIU F, et al. Wavelet-based residual attention network for image super-resolution[J]. Neurocomputing, 2020, 382：116-126.

[19] LIU W, YAN Q, ZHAO Y Z. Densely self-guided wavelet network for image denoising[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 14-19, 2020, Seattle, WA, USA. New York：IEEE, 2020：1742-1750.

[20] LIU L, LIU J Z, YUAN S X, et al. Wavelet-based dual-branch network for image demoiréing[C]//Proceedings of 2020 European Conference on Computer Vision, November 28, 2020, Glasgow, UK. Springer：Cham, 2020：86-102.

[21] ZUO C, QIAN J M, FENG S J, et al. Deep learning in optical metrology：a review[J]. Light：science and applications, 2022, 11(1)：39.