RECONSTRUCTION OF THE SURFACE PROFILE OF OPTICAL-MECHANICAL STRUCTURE WITH WHITE LIGHT INTERFERENCE BY DEEP LEARNING U-NET MODEL

Doan Thong Nguyen[1,*], Hoang Hai Le[2], Kim Thu Le[3],
Linh Khuong Tran[4], Van Nhu Le[1]

[1]*Faculty of Special Equipments, Le Quy Don Technical University*
[2]*Office of Science and Technology, Le Quy Don Technical University*
[3]*Institute of Techniques for Construction Technology, Le Quy Don Technical University*
[4]*Office of Logistics and Facilities Management, Le Quy Don Technical University*

Abstract

This research proposes a method to accurately determine the location of the interference signal peak in white light interferometry. This improves the accuracy and resolution of the measurement, enabling precise measurement and reconstruction of three-dimensional micro-structures on optical surfaces. By using a deep learning algorithm with a U-net neural network architecture to generate a predicted signal that closely matches the reference signal, combined with Fast Fourier Transform to filter noise and fit the proposed function, the exact location of the envelope signal peak is determined, providing information about the height of the point. This method can achieve an accuracy of around 0.9 nanometers at a noise level of 50 dB, which is over 40% more accurate than the traditional Fourier transform method at various noise levels. The persistent issues of the traditional Fourier transform method, such as determining the location of the interference signal peak at a signal position and the accuracy being heavily influenced by noise and the fitting signal, are effectively addressed.

Keywords: *White light interferometry; analysis interference fringe; Fourier transform; deep learning; U-net.*

1. Introduction

Concomitant with the rapid and widespread development of the semiconductor industry, precision engineering, optoelectronics systems, Micro-Electro-Mechanical Systems (MEMS), among others, the demand for the metrology of micro-optical components has been steadily increasing. White Light Interference (WLI) has been a widely adopted measurement method since its inception [1, 2], and it is continually being developed and improved in terms of accuracy, processing speed, and the expansion of the measurement range and objects [3, 4]. The analysis and processing of measurement signals in WLI is of great importance, as the elimination of noise and the extraction of

---

[*] Corresponding author email: thongnd@lqdtu.edu.vn

DOI: 10.56651/lqdtu.jst.v2.n02.851.pce

accurate peak position information can contribute to the improvement of measurement accuracy. Several methods have been proposed for interferometric signal processing, with the simplest approach being the determination of the intensity maxima to identify the position of the zero-order fringe and obtain the corresponding height information. This method is fast and memory-efficient, but it suffers from large errors due to sensitivity to noise, and it has limited resolution and accuracy. Alternative approaches include frequency-domain processing using Fast Fourier Transform (FFT) [5, 6] and envelope fitting methods, the latter of which can eliminate phase errors but are susceptible to noise from various sources [7]. Some researchers have proposed combining these two methods to improve the measurement quality [2, 8]. The envelope-based method with Fourier transform can enhance the accuracy, but it is still affected by noise with frequencies close to the central wavelength of the light source. The direct polynomial fitting algorithm has been suggested to increase the resolution [8], but the accuracy is not yet satisfactory. Fundamentally, these methods are influenced by noise or dependent on the average wavelength of the light source. Noise can originate from various sources, such as the optical system, the piezoelectric transducer (PZT) displacement, environmental vibrations, and the image sensor itself [5, 9].

Deep learning algorithms have been studied since the 1960s. With increasing computational power and memory capacity, recent years have witnessed the rapid development of deep learning algorithms in most fields from healthcare, defense, industry, and education, to cultural life. U-net is one of the powerful deep neural network architectures and has been applied to image segmentation problems. U-net operates as a convolutional neural network with the purpose of analyzing and diagnosing images in the medical field. It consists of two main parts: the feature extractor (encoder) comprising convolutional layers combined with max-pooling layers to process and extract features from the data, and the feature decoder comprising convolutional layers with up-sampling layers to increase the output size and reconstruct the local details. The distinctive feature of U-net is the use of skip connections between corresponding layers, which allows it to leverage spatial information and characteristics at different levels [10].

In this study, a new method using the U-net deep learning algorithm is proposed to generate an interference signal along the axis direction that is closest to the reference signal, then the Fourier transform is used to match the envelope of the signal. The results of this method will demonstrate its effectiveness in noise reduction and envelope fitting using the proposed function thereby providing more accurate height information compared to other methods.

2. Principles

2.1. White light interference

In Fig. 1, the light is transmitted from a white light source, collimated through the illumination system, and then directed to the beam splitter. At the beam splitter, the light is divided into two beams: Beam 1 is directed towards the sample surface and reflected back; Beam 2 is directed towards the reference mirror and then reflected back. These two beams recombine and interfere at the imaging lens, creating an interference signal on the camera sensor. The PZT micro-translation stage is used to move the objective lens along the optical axis (Fig. 1). This translation causes the optical path to change, and when the optical path difference between the two beams is zero (ZOPD), the maximum interference signal is obtained. This signal is recorded and processed to obtain the topography of the scanned sample. A broadband white light source is used because its short coherence length allows for easy differentiation of the interference fringe edges.

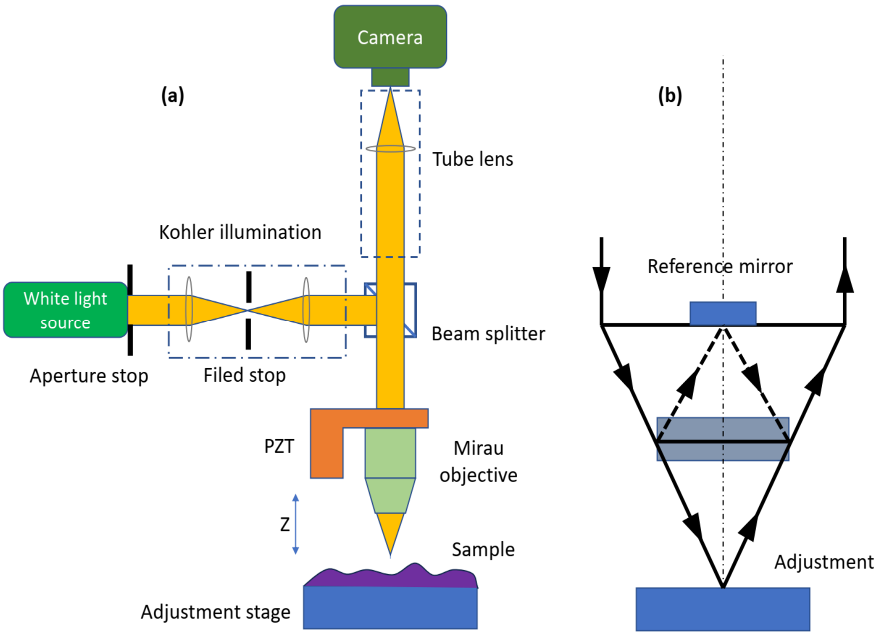

Fig. 1. Principle diagram of WLI (a); Path of light beam in Mirau objective (b).

The relationship between the light intensity distribution on the camera and the height *h* is represented by the following Eq. (1) [11]:

$$\begin{aligned} I(p) &= I_0(p) + A(p)\alpha(h-H_b)\cos\left[2k_b(h-H_b)\right] \\ &= I_0(p)\left[1+\beta c(h-H_b)\cos\left[2k_b(h-H_b)\right]\right] \\ &= I_0(p)\left[1+F\cos\left[2k_b(h-H_b)\right]\right] \end{aligned} \tag{1}$$

where $I_0(p)$ is the background intensity, h is the height of the point on the scanned samples along the z-axis. H_b is the reference distance dependent on the specific microscope objective, and β is the fringe contrast.

$$\beta = \frac{A(p)}{I_0(p)} \tag{2}$$

$\alpha(h-H_b)$ is the signal envelope coefficient, which depends on the light source's spectrum. $F=\beta\alpha\left(h-H_b\right)$ is the combined envelope function. λ_0 is the central wavelength of the light source, $\Delta\lambda$ is the spectral bandwidth, $l_c=\frac{\lambda_0{}^2}{\Delta\lambda}$ is the coherence length, $k_b=\frac{2\pi}{\lambda_0}$ is the wavenumber.

The light intensity distribution at each point can be written as [2]

$$I(z) = I_0 + \beta I_0 \exp\left[-\left(\frac{z-z_0}{l_c}\right)^2\right]\cos\left[\frac{4\pi}{\lambda_0}\left(z-z_0\right)+\theta_0\right] \tag{3}$$

where I_0 is the background intensity, β is the fringe contrast, l_c is the coherence length of the light source, z is the position scanned along the z-axis, z_0 is the position of the maximum ZOPD, θ_0 is the phase difference between the two light beams (the reference beam and the beam reflected from sample).

$$I(z) = I_0\left\{1+\beta F(z)\cos\left[\theta(z)\right]\right\} \tag{4}$$

$$F(z) = \exp\left[-\left(\frac{z-z_0}{l_c}\right)^2\right] \tag{5}$$

$$\theta(z) = \left[\frac{4\pi}{\lambda_0}\left(z-z_0\right)+\theta_0\right] \tag{6}$$

$F(z)$ is the envelope function of the signal.

2.2. Deep learning algorithms and U-net model

The U-net uses a network consisting of convolution layers to perform the task of segmenting pixels into semantic layers. The network has a symmetric architecture with an encoder that extracts spatial features from the input data and a decoder that helps build the segmentation map from the encoder features. The encoder follows the typical structure of a convolution network, including a sequence of two 3 × 3 convolution operations and a 2 × 2 max pooling operation with a stride of two. This sequence is repeated four times, and after each down-sampling, the number of filters in convolutional layers is doubled. Finally, two 3 × 3 convolution operations connect the encoder with the decoder.

In the opposite direction, the decoder first samples the feature map using a 2 × 2 transposed convolution, halving the number of feature channels, followed by a sequence of two 3 × 3 convolution operations. Similar to the encoder, this up-sampling and two-convolution sequence is repeated four times, with the number of filters in each stage halved. Finally, a 1 × 1 convolution is performed to create the final segmentation map. Except for the last one, all the convolutional layers in this architecture, use the Rectified Linear Unit (ReKU) activation function. The final convolution layer uses the Sigmoid activation function.

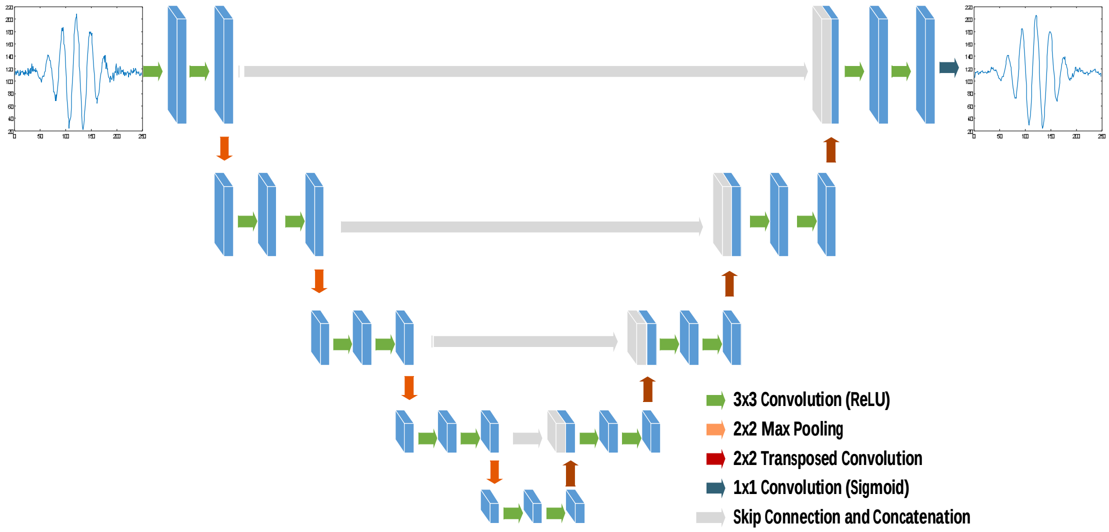

Fig. 2. The U-net architecture.

At all four levels, the output of the convolutional layer before the pooling operation in the encoder is passed to the decoder. These feature maps are then concatenated with the output of the up-sampling operation and the concatenated feature map is passed to the subsequent layers. These shortcut connections are a distinctive feature of the U-net architecture, allowing the recovery of spatial information lost due to the pooling operation.

2.3. The steps for signal processing

Step 1: Acquire the interference signal from the image sensor and convert it to the axial interference signal.

Step 2: Use the U-net model to output the axis interference signal that best matches the ground truth (all-zero level). The deep learning algorithm is used to process the signal, bringing any input signal as close as possible to the ground truth signal.

Step 3: Convert the axial interference signal into the frequency domain using Fourier transform.

Step 4: Use a unilateral Fourier signal window to filter the noise and convert this signal into the interference fringe envelope signal.

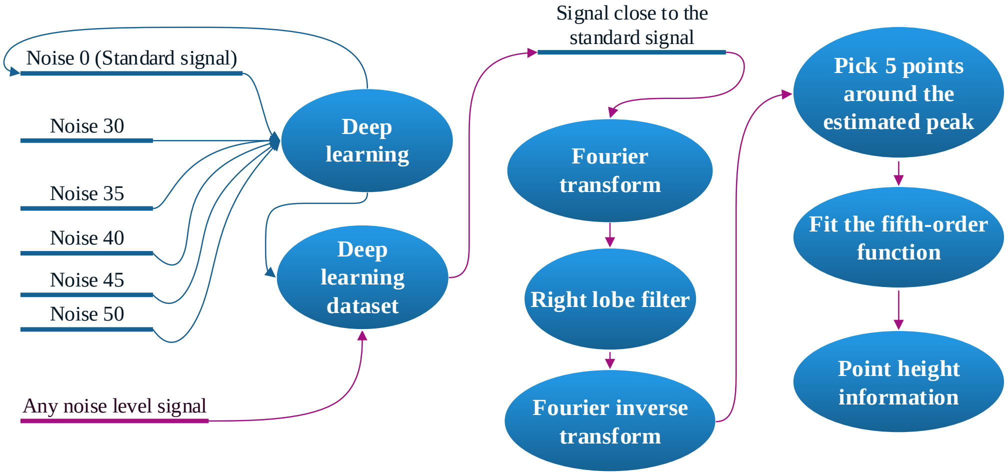

Fig. 3. Block diagram of the steps of the proposed method.

A fifth-order polynomial curve fitting function is proposed to provide fast yet accurate results. The specific steps of the curve-fitting process are shown in Fig. 4. Instead of using the preliminary peak position to calculate the height, which can result in large errors due to noise, or fitting the envelope function to find the peak position which can be very time-consuming, this research proposed a method of fitting a 5 points polynomial curve around the preliminary peak position after removing noise using the deep learning algorithm and Fourier transform.

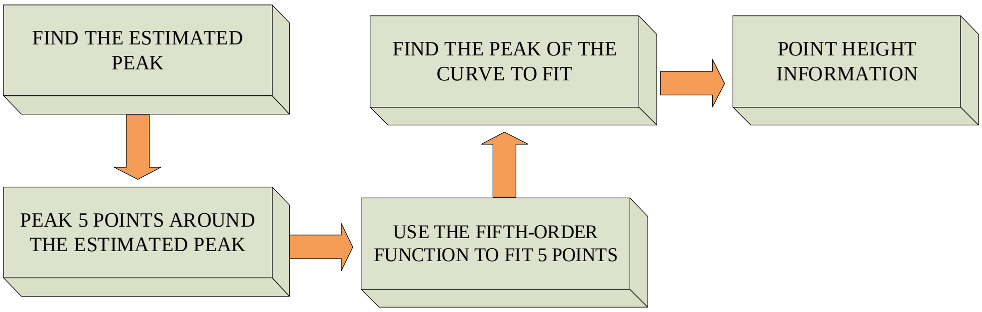

Fig. 4. Diagram of the signal fitting process.

The proposed curve fitting method helps save time compared to fitting the entire envelope function. Furthermore, the accuracy and resolution of the measurement can be improved at this step by reducing the step size of the signal being fitted. The results demonstrating this improvement will be shown in the section below.

3. Results and discussion

3.1. Simulation

The simulation system uses a microscope objective with a magnification of 20X and the pixel size is 4280 nm. The intensity distribution signal is set up in the form of expression (4), where λ_0 = 555 nm, the spectral width is 80 nm, the background intensity I_0 is 115, the fringe contrast is 0.8 and the scan step is 20 nm. Noise is introduced at different levels to test the processing algorithm and in this study, white Gaussian noise is used with the Signal-to-Noise Ratio (SNR) measured in (dB).

The simulation is performed on a desktop computer with an Intel Core i5-12400 (2.5 GHz, 8 MB cache) CPU, 16 GB RAM, and NVIDIA GeForce GTX 1050Ti.

The deep learning dataset is created as a basis for the processing algorithm, including noise levels of 0 dB, which is the standard noise level for deep learning, and 30 dB, 35 dB, 40 dB, 45 dB and 50 dB. The 0 dB noise level is the standard dataset when the image is not affected by noise, and the different noise levels are added according to Eq. (4).

$$I(z) = I_0 + \beta I_0 \exp\left[-\left(\frac{z-z_0}{l_c}\right)^2\right]\cos\left[\frac{4\pi}{\lambda_0}\left(z-z_0\right)+\theta_0\right] + \text{noise} \tag{7}$$

Figure 5 shows the intensity distribution of the signal at different noise levels at a certain scanning position. For each noise levels, different signals corresponding to randomly generated height positions are created. These signals are then used as input to the U-net deep learning algorithm to create the dataset.

The U-net network structure consists of 34 layers divided into three parts, as shown in Fig. 6:

- Encoding part: Includes the input data, convolutional layers to extract features from the input signal, and layers to reduce the spatial dimensions of the signal.

- Skip connections: Used to connect detailed information from the encoding part to the decoding part. Addition is commonly used to combine the information.

- Decoding part: Helps to restore the size of the information and generate output signals with the same size as the original signal. The output is reconstructed with the same size and contains segmentation information.

The signal obtained from the deep learning algorithm is used to perform a Fourier transform to the frequency domain. From this signal, a threshold values around the preliminary peak position. The preliminary peak position is quickly determined and then five nearby points around this position are extracted. A curve-fitting algorithm is then used to find the precise maximum position.

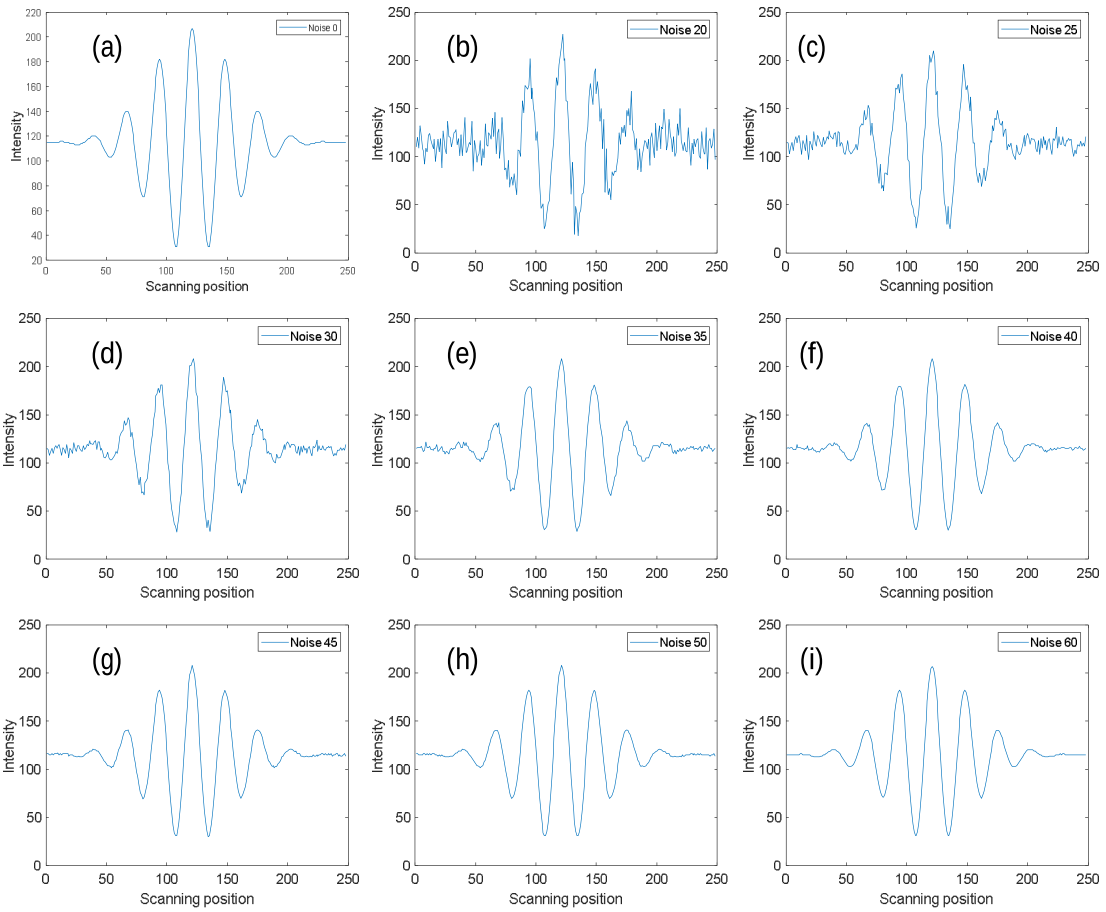

Fig. 5. Signal types correspond to different noise levels
(a, b, c, d, e, f, g, h, i correspond to noise levels 0, 20, 25, 30, 35, 40, 45, 50, 60 dB).

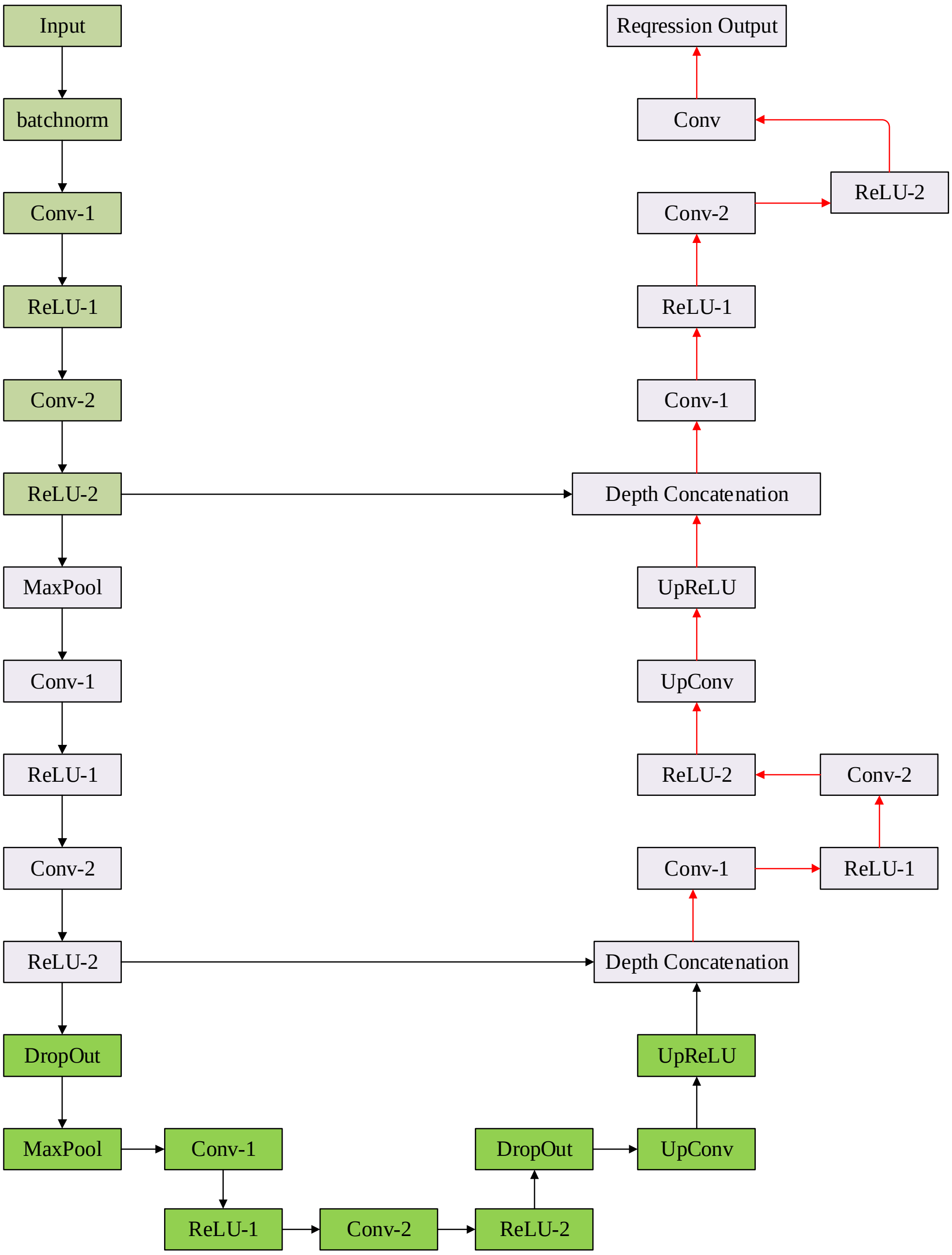

Fig. 6. Layer diagram of the U-net deep learning algorithm used in this study.

Figure 7 shows the result of the curve fitting process using the proposed fifth-order polynomial function at several arbitrary points. The results show that the maximum value of the simulated reference after the Fourier transform and curve fitting matches the simulated reference position. Compared to the preliminary OPD position, the OPD position is determined more accurately after the curve fitting.

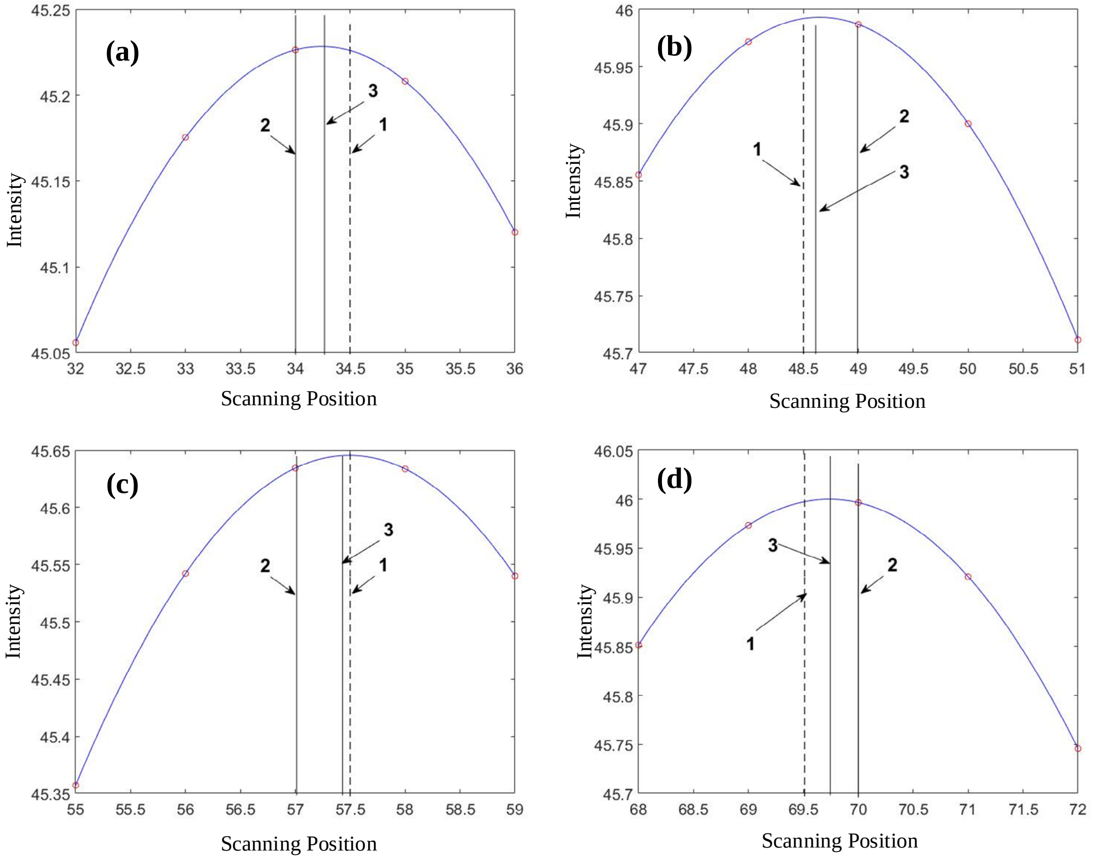

Fig. 7. The result of fitting the signal using a fifth-order polynomial function
1- standard ZOPD position; 2 - peak position after Fourier transform; 3 - peak position after fitting using the proposed function (a, b, c, d are any positions).

3.2. Results

To demonstrate the effectiveness of the proposed method, various three-dimensional (3D) surface profiles are generated, with the addition of different noise levels. The developed deep learning U-net algorithm processes the data and obtains the resulting interference signal. The results are compared with the traditional FFT method, without the use of the deep learning algorithm combined with the fifth-order polynomial curve fitting.

Figure 8 shows two simulated 3D surface profiles, which include step surfaces and tilted planar surfaces with different heights. The lower part shows the 2D cross-sections of the surface profiles. The interference fringes generated when scanning the surfaces are simulated and the interference signals along the axis are obtained. The proposed method is then used to evaluate the results of reconstructing the detailed 3D surface profiles, allowing the assessment of the accuracy and effectiveness of each method.

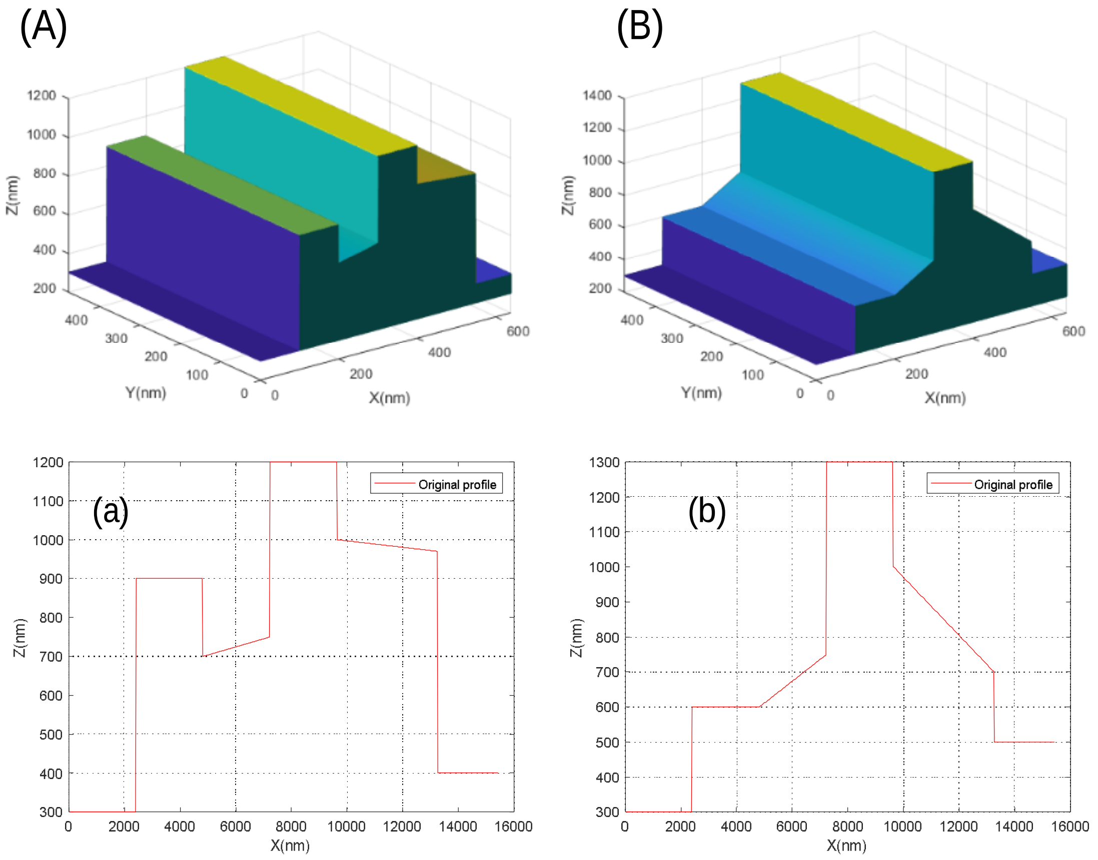

Fig. 8. Any three-dimensional surface profiles are simulated as A and B, respectively (a), (b) are the corresponding 2D cross sections.

Figure 9 below shows the 2D cross-sections of two simulated arbitrary profiles A and B. The results obtained using the proposed method are closer to the simulated profiles compared to the traditional FFT method, clearly demonstrating the effectiveness of the proposed approach. With high noise levels in the range of 30 dB to 40 dB, the proposed method shows clear and superior effectiveness in noise removal and provides results closer to the reference profiles. To provide a more comprehensive evaluation, the root mean square error (RMSE) is calculated to assess the error between the obtained results and the reference values for both methods under different noise levels for the two simulated surface profiles.

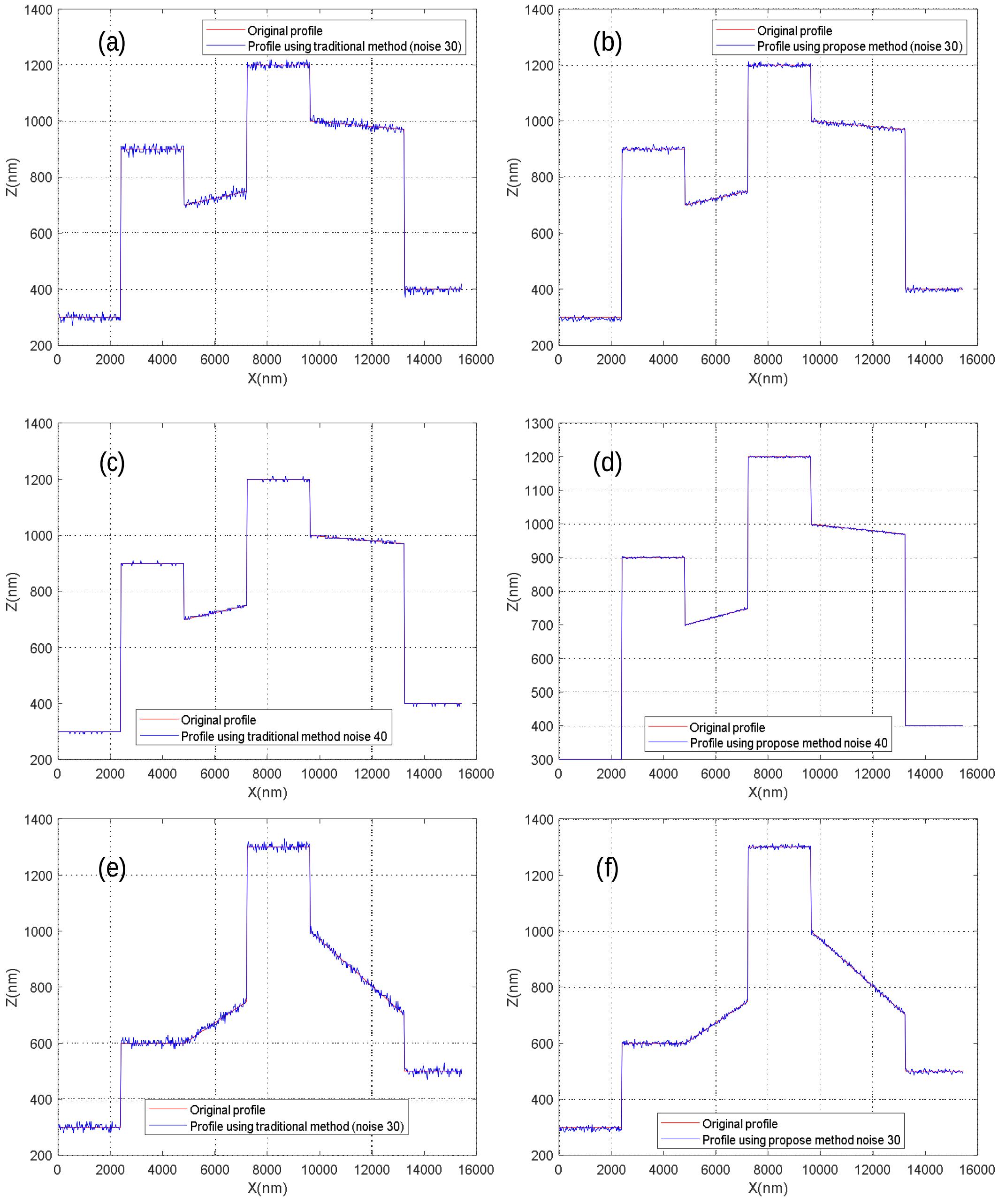

Fig. 9. The results of the proposed method are compared with the results of the traditional FFT method (a, c are cross sections of profile A using the traditional FFT method at 30 dB and 40 dB noise levels, respectively; b, d are cross sections of profile A using the proposed method at the above noise levels; e, f are the cross-section of profile B using the traditional FFT method and the proposed method respectively, at a noise level of 30 dB).

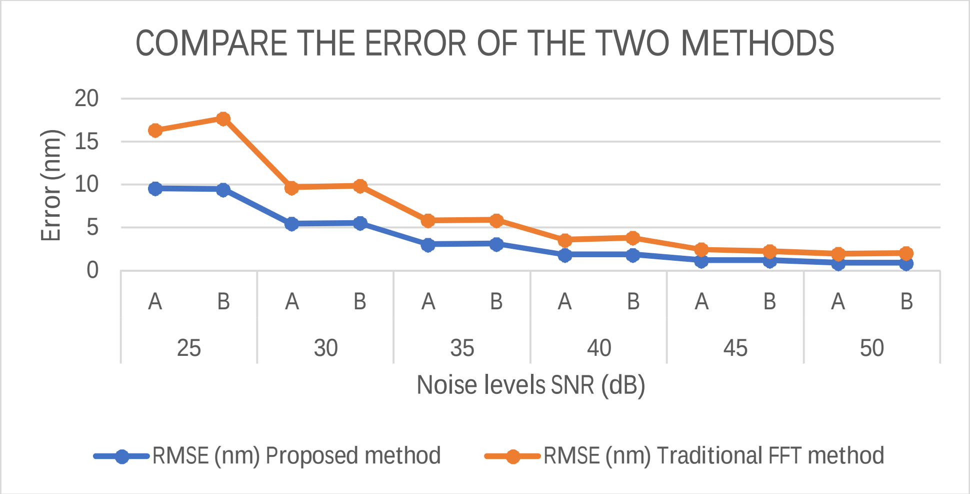

Fig. 10. Graph comparing the accuracy of the methods with two profiles A and B at different noise levels.

Figure 10 compares the error of the proposed method, which combines deep learning with a fifth-order polynomial fitting function, against the traditional FFT method for the two profiles A and B under different noise levels from high to low. In general, as the noise level decreases from 25 dB to 50 dB, the error of both methods decreases, but the proposed method has a significantly lower error compared to the traditional method across all noise levels and for both profiles.

At the high noise level of 25 dB, the error of the traditional method is 16.3496 nm for profile A and 17.7314 nm for profile B, while the proposed method has a much lower error of 9.5695 nm for profile A and 9.4414 nm for profile B, a reduction of over 40%. This demonstrates the effectiveness of the proposed method, especially in handling high-noise conditions.

Table. 1. RMSE error results of the two methods with profiles at different noise levels.

Noise levels (dB)	**25**	**30**	**35**	**40**	**45**	**50**
			Profile A			
Proposed method	9.5695	5.4617	3.0569	1.8513	1.1734	0.8919
Traditional FFT method	16.3496	9.661	5.8348	3.5678	2.4712	1.9488
			Profile B			
Proposed method	9.4414	5.547	3.1254	1.8638	1.1633	0.8781
Traditional FFT method	17.7314	9.8732	5.8733	3.8401	2.2491	2.0377

At a noise level of 50 dB, the accuracy of the proposed method can reach 0.8781 nm, while the traditional FFT method can only achieve 1.9488 nm.

4. Conclusion

In this paper, a method using the U-net deep learning algorithm with suitable parameters has been developed and proven effective in processing interference signals to be very close to the standard waveform. Subsequently, a fifth-order polynomial function is used to fit the signal in order to determine the ZOPD peak position, thereby improving the accuracy in reconstructing the 3D surface profile of the optical component using WLI.

Simulations of two arbitrary 3D profiles were conducted as the basis for evaluating the effectiveness of the proposed method. The accuracy of the method can reach 0.8781 nanometers at low noise levels. The accuracy can be further improved by increasing the input data as well as selecting optimal parameters for the U-net deep learning algorithm, optimizing the layers, and tuning the network parameters.

References

[1] N. N. Phan, H. H. Le, and C. D. Duong, "Surface curvature measurement of microlenses using a white-light interference microscope and fast geometric fit algorithm", *Optical Engineering*, Vol. 58, Iss. 12, 2019. DOI: 10.1117/1.OE.58.12.124105

[2] Q. Vo, F. Fang, X. Zhang, and H. Gao, "Surface recovery algorithm in white light interferometry based on combined white light phase shifting and fast Fourier transform algorithms", *Applied Optics*, Vol. 56, Iss. 29, pp. 8174-8185, 2017. DOI: 10.1364/AO.56.008174

[3] R. Ma, X. Li, X. Dong, and Y. Xia, "Wavelength scanning distance interferometry using inflection point retrieval for phase unwrapping", *Optics Communications*, Vol. 410, pp. 292-296, 2018. DOI: 10.1016/j.optcom.2017.10.030

[4] L. Zhu, Y. Dong, Z. Li, and X. Zhang, "A novel surface recovery algorithm for dual wavelength white LED in vertical scanning interferometry (VSI)", *Sensors (Basel, Switzerland)*, Vol. 20, Iss. 18, 2020. DOI: 10.3390/s20185225

[5] W. Chen, Y. Xiong, J. Chen, and S. Chen, "Accurate white light phase-shifting interferometry under PZT scanning error", *Optics and Lasers in Engineering*, Vol. 169, 2023. DOI: 10.1016/j.optlaseng.2023.107728

[6] X. Li, "Wavelet transform for detection of partial fringe patterns induced by defects in nondestructive testing of holographic interferometry and electronic speckle pattern interferometry", *Optical Engineering*, Vol. 39, Iss. 10, pp. 2821-2827, 2000. DOI: 10.1117/1.1308485

[7] Y. Li, Y. Yang, J. Zhang, Q. Yuan, and Y. Liang, "Envelope peak detection algorithm based on the CEEMDAN in white light interferometry", *Optics Continuum*, Vol. 2, Iss. 8, pp. 1875-1888, 2023. DOI: 10.1364/OPTCON.496459

[8] M. C. Park and S. W. Kim, "Direct quadratic polynomial fitting for fringe peak detection of white light scanning interferograms", *Optical Engineering*, Vol. 39, Iss. 4, pp. 952-959, 2000. DOI: 10.1117/1.602445

[9] M. Konnik and J. Welsh, "High-level numerical simulations of noise in CCD and CMOS photosensors: Review and tutorial", *arXiv*, 2014. https://arxiv.org/pdf/1412.4031

[10] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation", *Neural Networks*, Vol. 121, pp. 74-87, 2020. DOI: 10.1016/j.neunet.2019.08.025

[11] L. Ma, J. Jia, X. Pei *et al.*, "A robust surface recover algorithm based on random phase noise correction for white light interferometry", *Optics and Lasers in Engineering*, Vol. 128, 2020. DOI: 10.1016/j.optlaseng.2020.106016

TÁI TẠO BIÊN DẠNG BỀ MẶT CHI TIẾT QUANG-CƠ VỚI GIAO THOA ÁNH SÁNG TRẮNG BẰNG MÔ HÌNH HỌC SÂU U-NET

Nguyễn Doãn Thông[1], Lê Hoàng Hải[2], Lê Kim Thư[3],
Trần Linh Khương[4], Lê Văn Nhu[1]
[1]*Khoa Thiết bị đặc biệt, Trường Đại học Kỹ thuật Lê Quý Đôn*
[2]*Phòng Khoa học và Công nghệ, Trường Đại học Kỹ thuật Lê Quý Đôn*
[3]*Viện Kỹ thuật công trình đặc biệt, Trường Đại học Kỹ thuật Lê Quý Đôn*
[4]*Phòng Hậu cần - Kỹ thuật, Trường Đại học Kỹ thuật Lê Quý Đôn*

Tóm tắt: Nghiên cứu này đề xuất một phương pháp để xác định chính xác vị trí đỉnh của tín hiệu giao thoa trong giao thoa ánh sáng trắng. Qua đó nâng cao độ chính xác cũng như độ phân giải của phép đo, ứng dụng trong đo lường chính xác và tái tạo vi cấu trúc ba chiều bề mặt chi tiết quang cơ. Bằng việc sử dụng thuật toán học sâu với kiến trúc mạng nơron U-net để đưa ra tín hiệu dự đoán sát với tín hiệu chuẩn nhất, kết hợp với biến đổi nhanh Fourier để lọc nhiễu và làm khớp với hàm số đề xuất, vị trí chính xác của đỉnh đường bao tín hiệu được xác định từ đó đưa ra thông tin về độ cao của điểm. Phương pháp này có thể đạt độ chính xác cỡ 0,9 nanomet ở mức nhiễu 50 dB, độ chính xác tăng hơn 40% ở các mức nhiễu so với phương pháp biến đổi Fourier truyền thống. Các vấn đề còn tồn tại của phương pháp biến đổi Fourier truyền thống trước đây như việc xác định vị trí đỉnh của tín hiệu giao thoa tại một vị trí, độ chính xác còn bị ảnh hưởng nhiều bởi nhiễu và tín hiệu làm khớp được khắc phục một cách hiệu quả và rõ rệt.

Từ khóa: *Hiển vi giao thoa ánh sáng trắng; phân tích vân giao thoa; biến đổi Fourier; thuật toán học sâu; U-net.*

Received: 05/08/2024; Revised: 30/09/2024; Accepted for publication: 30/10/2024

❑