

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Fully automated extraction of the fringe skeletons in dynamic electronic speckle pattern interferometry using a U-Net convolutional neural network

Biyuan Li
Chen Tang
Tingyue Zheng
Zhenkun Lei

SPIE.

Biyuan Li, Chen Tang, Tingyue Zheng, Zhenkun Lei, "Fully automated extraction of the fringe skeletons in dynamic electronic speckle pattern interferometry using a U-Net convolutional neural network," *Opt. Eng.* **58**(2), 023105 (2019), doi: 10.1117/1.OE.58.2.023105.

Fully automated extraction of the fringe skeletons in dynamic electronic speckle pattern interferometry using a U-Net convolutional neural network

Biyuan Li,^a Chen Tang,^{a,*} Tingyue Zheng,^a and Zhenkun Lei^b

^aTianjin University, School of Electronic and Information Engineering, Tianjin, China

^bDalian University of Technology, State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian, China

Abstract. With the development of artificial intelligence technology, intelligent fringe processing is a goal of relevant researchers in optical interferometry. We propose an intelligent method to achieve fully automated extraction of the fringe skeletons in electronic speckle pattern interferometry (ESPI) based on U-Net convolutional neural network. In the proposed method, the network is first trained by the samples that consist of the noisy ESPI fringe patterns and the corresponding skeleton images. After training, the other multiframe ESPI fringe patterns are fed to the trained network simultaneously; the corresponding skeleton images can be obtained in batches. Using our method, it is not necessary to process fringe patterns frame by frame. Our method is especially suitable for multiframe fringe patterns processing. We apply the intelligent method to one computer-simulated and one real-dynamic ESPI measurement, respectively. For the simulated measurement, it takes just 40 s to obtain the skeleton images of 20 noisy ESPI fringe patterns using our method. Even for low-quality experimental obtained ESPI fringe patterns, our method can also give desired results. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.OE.58.2.023105]

Keywords: skeleton extraction; intelligent fringe processing; convolutional neural network; electronic speckle pattern interferometry.

Paper 181628 received Nov. 14, 2018; accepted for publication Jan. 25, 2019; published online Feb. 12, 2019.

1 Introduction

Electronic speckle pattern interferometry (ESPI) as a well-known optical metrology technique has been extensively investigated and widely used in numerous fields.¹⁻³ It is well known that accurate extraction of phase terms is of fundamental importance for the successful application of ESPI. Compared with the static measurement, the extraction of phase terms in real-time and dynamic ESPI measurement is a more challenging problem. The fringe skeleton method may be the most straightforward approach to estimate the phase terms in dynamic ESPI. In the fringe skeleton method, the phase terms can be obtained by interpolating the assigned skeletons.

In our opinion, the fringe skeleton method can be improved from three points of view. In the first point, a majority of effort is to improve the denoising algorithms. If a sufficiently high degree of quality can be achieved, then skeletonization of fringes of the resulting perfectly filtered fringe pattern can be accomplished with a common binarization thinning method.^{4,5} In the second point of view, few efforts are put into improving the accuracy (AC) of interpolation algorithms, mainly including backpropagation neural networks method⁶ and radial basis function (RBF) interpolation method.⁷ It is worth mentioning that the RBF interpolation method works well even under a seriously disconnected skeleton image. In the last point of view, some efforts are concentrated on developing the skeleton extraction algorithms. The thresholding binary-fringe thinning⁸ and fringe extreme tracking⁴ are two simple algorithms for extracting the skeletons. However, the two methods are likely to be affected by speckle noise. Recently, the skeleton extraction algorithms

based on gradient vector fields (GVFs) have been actively developed.⁹ The basic principle is that the GVFs of an original ESPI fringe pattern are calculated by a set of particular partial differential equations (PDEs). Then the skeletons of a fringe pattern are extracted by analyzing the topological structure of GVFs.¹⁰ As far as we know, there have been three kinds of governing PDEs especially proposed for calculating GVFs of ESPI fringe patterns. The coupled nonlinear governing PDEs were first proposed for ESPI fringe patterns with usual uniform density.¹⁰ The oriented couple PDEs for high-density fringe were proposed.¹¹ The anisotropic PDEs were also proposed for uniform density.¹² Recently, Tang et al.¹³ constructed the new GVFs based on variational image decomposition (VID) for skeletonization of ESPI fringe patterns with variable density.

When ESPI systems are used in dynamic measurements, large numbers of fringe patterns are usually recorded. Using all of the previous fringe skeletonization methods, one has to process the fringe patterns frame by frame, which is inconvenient and time-consuming. Therefore, it would be highly expected to find an intelligent method, which can extract the fringe skeletons in a simple and efficient way.

Recently, artificial intelligence (AI) has seen an explosion of interest and is being successfully applied across an extraordinary range of problem domains. AI enables computer processes to do work that only humans could do in the past.¹⁴ Deep learning is the foundation of AI. In the past few years, many network architectures based on deep learning have been developed. U-Net is one of the advanced convolutional neural networks (CNN) in deep learning. Ronneberger et al.¹⁵ first proposed the U-Net architecture network and

*Address all correspondence to Chen Tang, E-mail: tangchen@tju.edu.cn

applied it to the semantic segmentation on biomedical images. Since then, U-Net CNN has emerged as one of the most prominent methods for image processing problems in various domains, including medical image computing^{16–23} and computer vision.^{24,25}

In recent years, machine learning has been introduced into the optical fringe analysis domain. Sawaf and Groves²⁶ employed a neural network based on machine learning to identify phase discontinuities in the presence of random noise. Cai et al.²⁷ used extra filters and multilayer neural network to identify different noise types and levels of interference fringes. Anantrasirichai et al.²⁸ applied an Alexnet CNN to classify interferometric fringes in InSAR data.

Based on deep learning, it is possible to construct an intelligent network that can conveniently process large numbers of fringe patterns without frame by frame. In this paper, we propose an intelligent method to achieve fully automated extraction of the fringe skeletons in dynamic ESPI based on U-Net CNN. In the proposed method, the network is first trained by the samples. The samples consist of the noisy ESPI fringe patterns and the corresponding skeleton images. After training, the other multiframe ESPI fringe patterns are fed to the trained network simultaneously, the corresponding skeleton image can be obtained in batches. Using our method, it is not necessary to process fringe patterns frame by frame. Our method is especially suitable for multiframe fringe pattern processing.

The paper is organized as follows. In Sec. 2, we describe the proposed method in detail. Some experimental results are demonstrated in Sec. 3. Finally, Sec. 4 concludes this paper.

2 Our Proposed Method

2.1 Architecture of U-Net Network Used in Our Method

Our network architecture, which is based on the U-Net CNN, consists of a decomposition path (encoder) and a reconstruction path (decoder) as shown in Fig. 1. The advantage of this architecture is that the deeper features of an image can be obtained by the decomposition path and the reconstruction path gradually recovers the image. The decomposition path has four convolutional blocks, which increase the

number of feature maps from 1 to 512. Every block has two convolutional layers with the filter sizes of 3×3 pixels and the convolution stride is fixed to 1 pixel. Every convolutional layer is equipped with the rectified linear unit (ReLU), which is used to resolve the vanishing gradient problem. After every block in the decomposition path, a 2×2 pixels max pooling with stride 2 is used to reduce the dimensionality, so the size of the feature maps decreases from 512×512 to 64×64 . In the reconstruction path, every block starts with a 3×3 up-convolution with stride 2. After the up-convolution step, the size of the feature maps is increased, but the number of the feature maps is halved. Then a merge operation is performed by concatenating the feature maps into the correspondingly feature maps that from the decomposition path. In the final layer, a 1×1 convolution is used to reduce the number of the feature maps to 2. Different from the original U-Net architecture, we use zero padding so that the output dimension does not decrease after each convolution. As a result, the final output of the U-Net network produces a pixel-wise probability map. In order to reduce over-fitting, there is a dropout layer after each convolutional layer in the network. The dropout is to randomly drop units (along with their connections) from the neural network during training. The dropout layer makes the network independent of some local features and prevents the parameters over-fitting. Using dropout, the generalization performance and robustness of the network are improved.

2.2 Training Process of Our Method

The training process can be divided into two parts, the forward propagation and the backpropagation. The flowchart of the training process is shown in Fig. 2.

In the forward propagation, an input ESPI image is fed to the U-Net network. In the network, we let symbols W^l and B^l be the weight parameters and deviation parameters of the l 'th convolutional layer. The parameters can reflect the relationship between the input ESPI fringe pattern and the output skeleton image. The training network actually converts into a process of finding the optimal parameters.

The problem of the training process can be formulated as a per-pixel classification problem. For an original input ESPI fringe image $I = \{\hat{I}_i, i = 1, \dots, Z\}$, where \hat{I}_i is the pixel

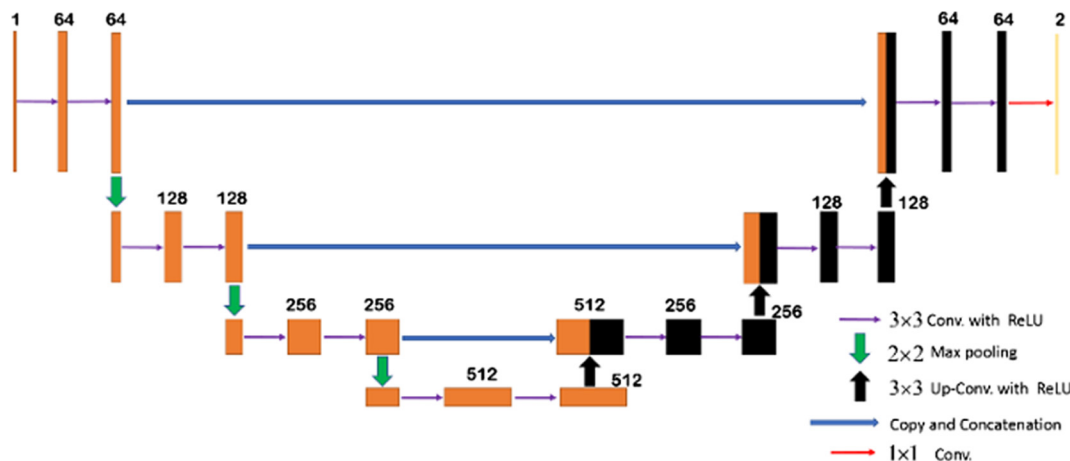


Fig. 1 Structure of the proposed deep convolutional network for fringe skeleton detection and extraction.

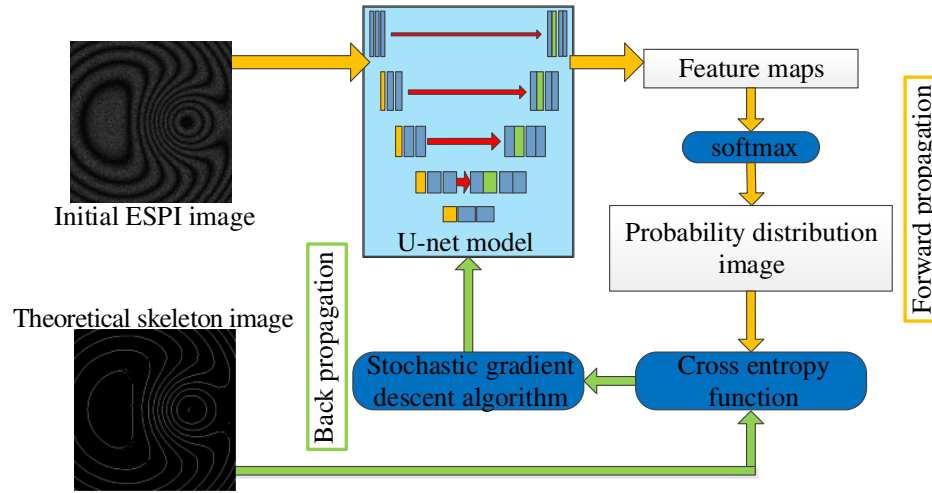


Fig. 2 Flowchart of the training process.

in the image and Z is the number of pixels in the image. Our purpose is to predict the skeleton image $S = \{\hat{S}_i, i = 1, \dots, Z\}$ of the input fringe image by the trained network. The pixel $\hat{S}_i \in \{0,1\}$ denotes the predicted label for each pixel \hat{I}_i , which means that if the pixel \hat{I}_i is predicted as a skeleton pixel, the pixel $\hat{S}_i = 1$; otherwise $\hat{S}_i = 0$.

In the forward propagation, the input image I will finally convert into a feature vector with two channels after all of the l convolutional layers. The feature vector of the pixel \hat{I}_i can be expressed as

$$H_i^l = \text{ReLU}(W^l \otimes H_i^{l-1} + B^l), \quad (1)$$

where \otimes is the convolutional operator, H_i^{l-1} is the output feature map of $(l-1)$ 'th convolutional layer. ReLU is the rectified linear unit.

The output feature vectors I_{fm} of I can be expressed as $I_{fm} = (H_i^l, i = 1, \dots, Z)$. Then a softmax function is used to determine whether the pixel \hat{I}_i belongs to the skeleton pixel or the nonskeleton pixel. The softmax function is defined as

$$p_c(\hat{I}_i) = \frac{e^{I_{fm}}}{\sum_{c=1}^2 e^{I_{fm}}}, \quad (2)$$

where $p_c(\hat{I}_i)$ is a probability value that the pixel \hat{I}_i belongs to category c . In our case, the pixel only has two possibilities, so $c = 1$ or 2 . Using softmax function, the output results are a series of probability values.

Then a loss function is used to quantify the difference between the probability values obtained by the forward propagation and the truth values. Here we use the cross entropy loss function, which can be expressed as

$$\text{loss} = \frac{1}{m} \sum_{i=1}^m \sum_{c=1}^2 y^i \log p_c(\hat{I}_i), \quad (3)$$

where y^i is a 1×2 vector, which denotes the truth category of the pixel \hat{I}_i . $y^i = 0$ or 1 and the judgment is depended on the corresponding skeleton pixel \hat{S}_i .

In the backpropagation, by minimization of the cross entropy loss function, the optimal parameters can be finally obtained

$$(W^l, B^l) = \arg \min(\text{loss}). \quad (4)$$

To solve Eq. (4), the stochastic gradient descent method²⁹ is used. When the parameters of the network are determined, the training process is finished.

2.3 Prediction Process of Our Method

In the prediction process, the testing images are fed to the network. Each pixel of the testing images will be determined whether it belongs to the skeleton pixel or not by the trained network. Finally, the skeleton images can be obtained.

3 Simulated and Experimental Results

In this section, one computer-simulated and one real-dynamic ESPI measurement experiments are used to evaluate the performance of our method.

The implementation of our method is based on python 2.7 under the same conditions of a personal computer equipped with TensorFlow framework with NVIDIA Quadro K 2200 as GPU, Xeon E5-1650 Dual CPU at 3.5 GHz and 16 GB RAM memory. The momentum of stochastic gradient descent method is 0.9. The initial learning rate is 0.001, and it will reduce by a factor of 2 if the validation loss does not improve for at least six epochs.

3.1 Simulated Dynamic ESPI Measurement

3.1.1 Training samples

We simulate a dynamic ESPI experiment and obtain 1000 noisy ESPI fringe patterns. The phase values of the 1000 ESPI fringe patterns are generated by the following equations:

$$\begin{aligned} \psi_{ij} = \alpha \times \left\{ \exp \left[-\frac{(2i-m)^2 + (2j - \frac{11n}{8})^2}{50000} \right] \right\} \\ + \exp \left[-\frac{(2i-m)^2 + (2j - \frac{n}{2})^2}{35000} \right] \\ + 10 \times \left[\left(\frac{6i-3m}{2m} \right)^2 + \left(\frac{3j-n}{n} \right)^2 \right], \end{aligned} \quad (5)$$

where ψ_{ij} is the phase, m, n are the image sizes, here $m = 512, n = 512$, and α is a parameter and $\alpha \in [0.01, 100]$.

α is a parameter for the simulated phase. By changing the value of α , we can obtain a series of simulated ESPI fringe patterns. We simulate 1000 images with a range of α from 0.01 to 100. The value of α increases by 0.0999 each time.

We pick 21 original ESPI fringe patterns and the corresponding skeleton images as the training samples from the whole dynamic process. The other images are used for the test. The 21 simulated fringe patterns are picked from the 1000 fringe patterns evenly. The values of α in the 21 ESPI fringe patterns, respectively, are 0.01, 5.0095, 10.009, 15.0085, 20.008, 25.0075... The interval of α is 4.9995.

To obtain the skeleton images of the training samples, we have to apply the GVF method based on VID¹³ 21 times to the 21 ESPI fringe patterns.

The details for the GVF method based on VID can be found in the published literatures in Ref. 13. Here we give a brief review of this method.

A skeleton extraction method is proposed based on the GVFs and VID.¹³ The GVFs of a whole image are the sum of the decomposed GVFs of low-density regions and high-density regions based on VID. Each part of the decomposed GVFs is described by a suitable function space. The skeletons of ESPI fringe patterns with variable density can be obtained based on the topological analysis of the GVFs of a whole image.

The GVF is defined by

$$V(x, y) = [u(x, y), v(x, y)], \quad (6)$$

where $u(x, y)$ and $v(x, y)$ are the two components of the GVF.

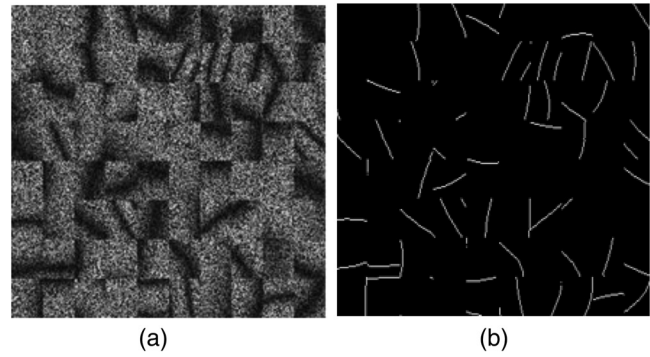


Fig. 4 Example of the input training sample patches and the corresponding skeleton image patches for training the network: (a) 32×32 pixel random portions of computer-generated noisy ESPI fringe and (b) the corresponding skeletons of (a).

The GVFs $V(x, y)$ of the ESPI fringe patterns with variable density are decomposed into three components

$$V(x, y) = V_l + V_h + V_n, \quad (7)$$

where $V_l(u_l, v_l)$, $V_h(u_h, v_h)$, and $V_n(u_n, v_n)$ are the decomposed GVFs in low-density regions, high-density regions, and noise regions, respectively.

The GVFs in low-density regions $V_l(u_l, v_l)$, high-density regions $V_h(u_h, v_h)$, and noise regions $V_n(u_n, v_n)$ are described in Beppo-Levi space, the high-density regions in Hilbert space, and the noise regions in Curvelet space, respectively.

In the VID method, the GVFs in low-density regions $V_l(u_l, v_l)$, high-density regions $V_h(u_h, v_h)$, and noise

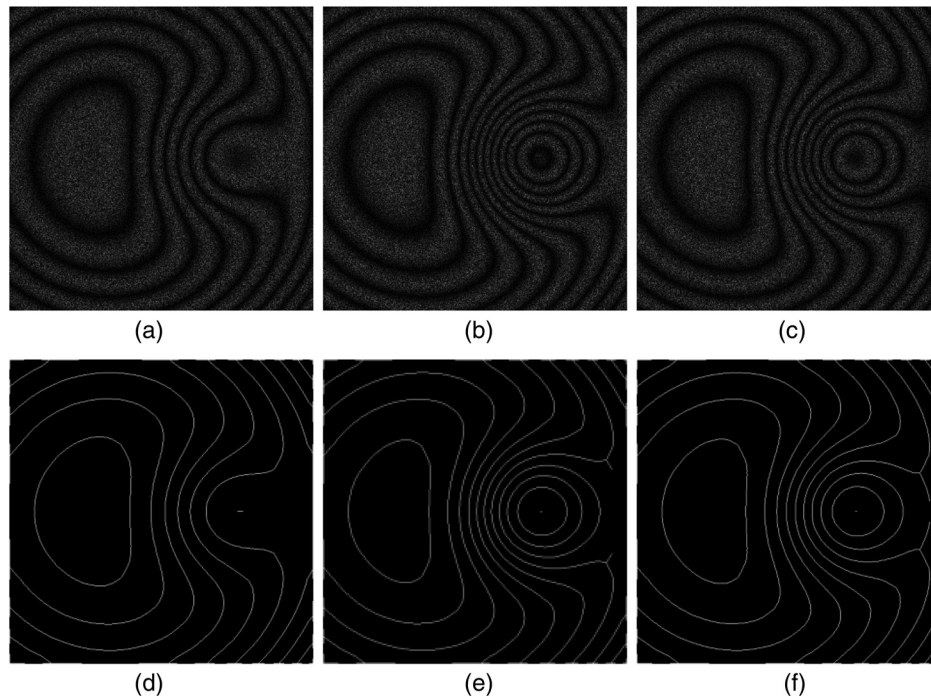


Fig. 3 Parts of the training samples and the corresponding skeleton images used for training the network: (a)–(c) initial images and (d)–(f) skeleton images of (a)–(c), respectively.

regions $V_n(u_n, v_n)$ can be obtained by solving the following minimization problem:

$$(V_l, V_h, V_n) = \arg \min_{\tilde{V}_l, \tilde{V}_h, \tilde{V}_n} \frac{1}{2} \|V - \tilde{V}_l - \tilde{V}_h - \tilde{V}_n\|_{L^2}^2 + \lambda \|V_l\|_{BL}^2 + \mu \|V_h\|_{\xi}^2 + \sigma \|V_n\|_{\text{Curvelet}}^2, \quad (8)$$

where λ, μ , and σ are three positive reals to balance each term and $\frac{1}{2} \|V - \tilde{V}_l - \tilde{V}_h - \tilde{V}_n\|_{L^2}^2$ is a fidelity term that takes into account the presence of noise in the image.

The GVF's of whole image can be calculated by adding desired results of V_l with V_h

$$u = u_l + u_h, \quad v = v_l + v_h. \quad (9)$$

After calculating the GVF's, the skeletons can be extracted based on the topological analysis of GVF's. For instance, the GVF's are first normalized, then the skeleton image of black fringes $f_{bfs}(x, y)$ can be obtained by

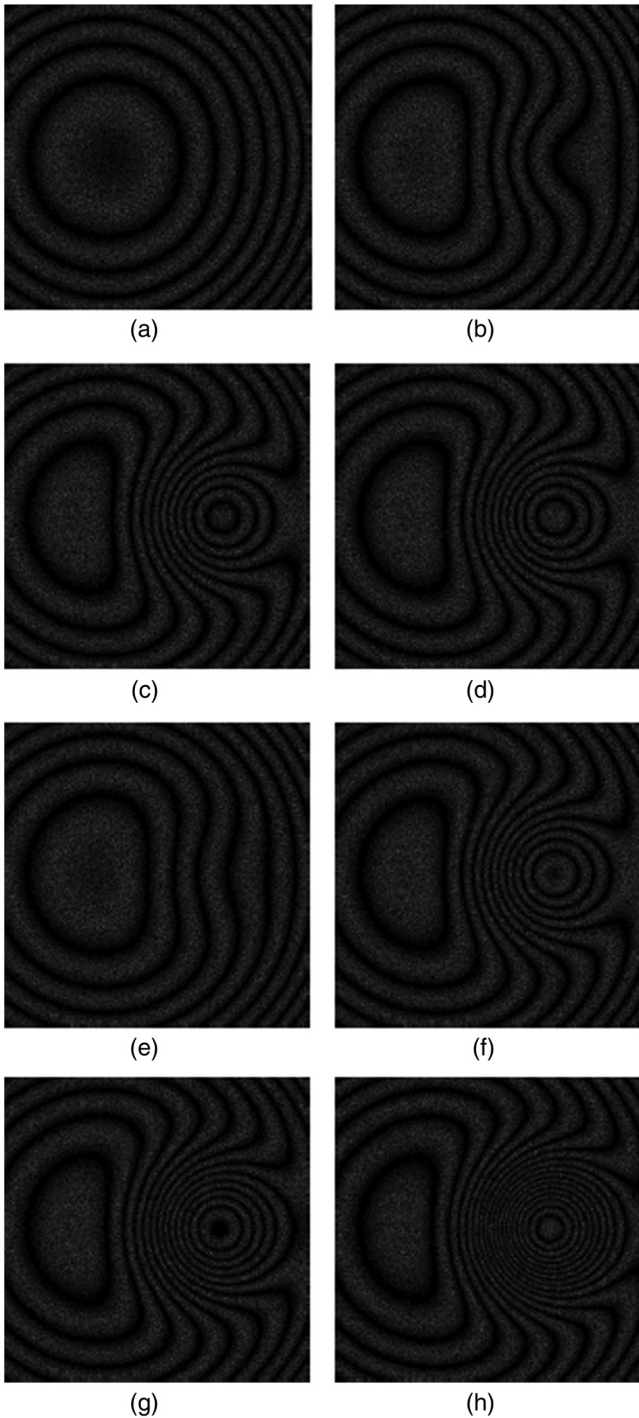


Fig. 5 Parts of the testing samples: (a)–(h) the initial ESPI images.

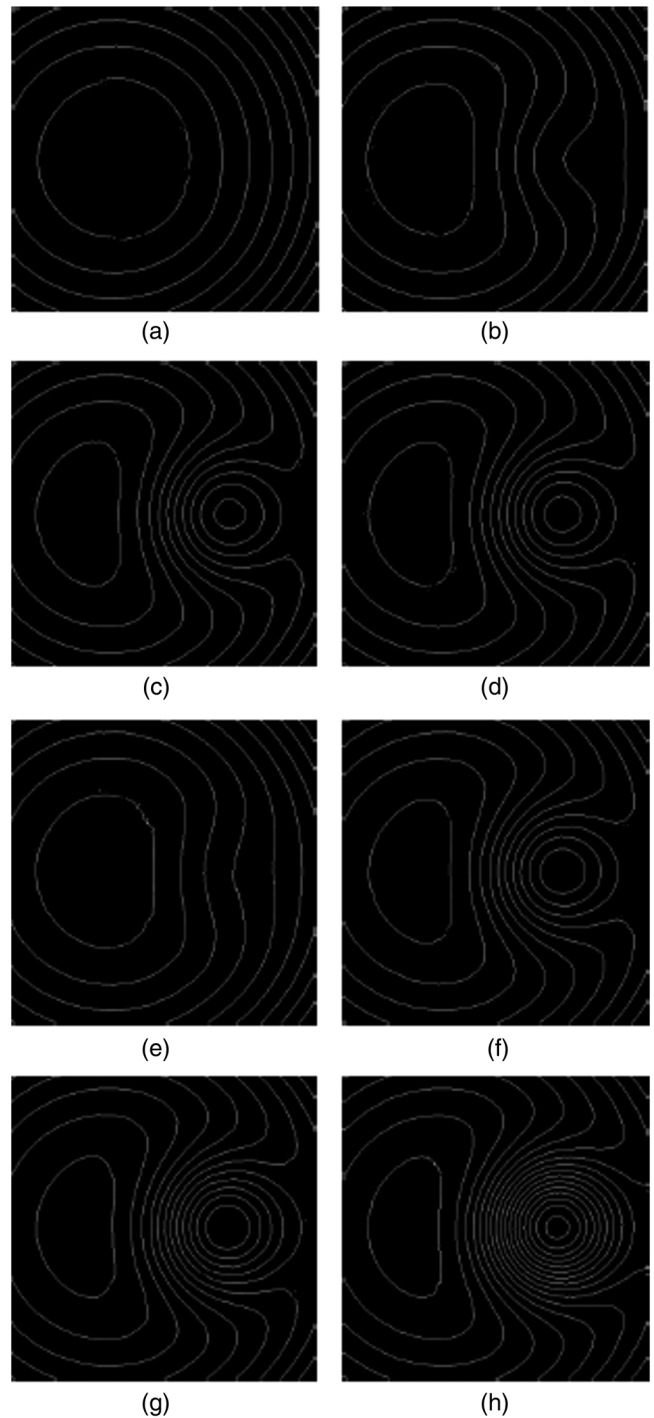


Fig. 6 The skeleton extraction results by the proposed method: (a)–(h) the skeleton images of Figs. 5(a)–5(h), respectively.

$$f_{bfs}(x, y) = \begin{cases} 1 & \text{skeleton pixel} & \text{if } S_{x,y} > T, JM_{x,y} > 0 \\ 0 & \text{nonskeleton pixel} & \text{otherwise,} \end{cases} \quad (10)$$

where T is a preset threshold. $S_{x,y}$ and $JM_{x,y}$ can be expressed as

$$S_{x,y} = \frac{\left(\frac{\partial u_{xy}^N}{\partial x}\right)^2 + \left(\frac{\partial v_{xy}^N}{\partial y}\right)^2}{\max\left[\left(\frac{\partial u_{xy}^N}{\partial x}\right)^2 + \left(\frac{\partial v_{xy}^N}{\partial y}\right)^2\right]}, \quad (11)$$

$$JM_{x,y} = \frac{\left(\frac{\partial u_{xy}^N}{\partial x} + \frac{\partial v_{xy}^N}{\partial y}\right) \operatorname{real} \sqrt{\left(\frac{\partial u_{xy}^N}{\partial x}\right)^2 - 2\left(\frac{\partial u_{xy}^N}{\partial x}\right)\left(\frac{\partial v_{xy}^N}{\partial y}\right) + \left(\frac{\partial v_{xy}^N}{\partial y}\right)^2} + 4\left(\frac{\partial u_{xy}^N}{\partial x}\right)\left(\frac{\partial v_{xy}^N}{\partial y}\right)}{\tau} \quad (12)$$

Figure 3 shows some noisy ESPI fringe patterns and the corresponding skeleton images, which are used as the training samples.

As we all know, the precision of the prediction produced by the CNN can be improved by increasing the number of training examples. In our case, it is not easy to obtain annotated ESPI skeleton images. In order to tackle the deficiency of the training samples, we construct the training samples for patch-based pixel-wise segmentation and sampled overlapped patches.

In our training dataset, each noisy ESPI fringe pattern and each skeleton image are both cropped into 4800 patches, respectively. The size of each patch is 32×32 pixels. In order to maintain AC, the patches extracted from the training samples are overlapped and the total number of patches

sampled from the training dataset is 100,800. This effectively reduces GPU memory usage and takes advantage of efficient batch processing.

Figure 4 shows an example of the ESPI fringe patterns patches and the corresponding skeleton patches. Then all of the patches are as input fed to the U-Net CNN. After 200 iterations, the network training process is completed.

3.1.2 Testing results

The other 979 ESPI fringe patterns as the testing images are fed to the trained network. In each test, 20 noisy ESPI fringe patterns are simultaneously fed to the network. The corresponding skeleton images can be obtained simultaneously. Here we provide eight testing images shown in Fig. 5 and the corresponding skeleton extraction results by our method are shown in Fig. 6.

Subsequently, we apply this trained network to a group of difform ESPI fringe patterns. The formulas of phase in these images are different from those of phase in Figs. 3 and 4. The shapes of the testing images are different from those of the training samples. The skeleton extraction results are still reliable. The testing ESPI fringe patterns and the corresponding skeleton images are shown in Fig. 7.

AC is an extensively used index for the measure of the ratio of pixels that are precisely classified across all pixels. The AC³⁰ is given by the following equation:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (13)$$

where TP is the number of classifications where pixels are classified into skeleton pixels correctly and TN is the number of classifications where pixels are identified as nonskeleton pixels correctly in the outputs of the network. FP is the number of classifications where pixels are misclassified into

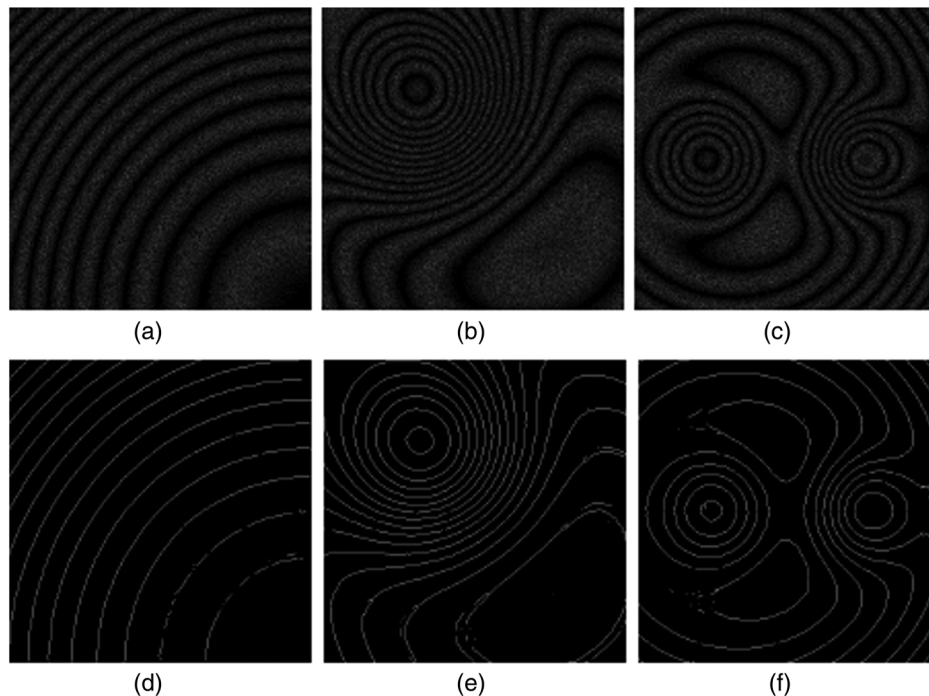


Fig. 7 Three difform ESPI fringe images and the corresponding skeleton results: (a)–(c) initial ESPI images with different density and (d)–(f) skeleton images of (a)–(c), respectively.

skeleton pixels in the outputs of the network. FN is the number of classifications where pixels are misclassified into nonskeleton pixels in the outputs of the network.

As shown in Figs. 3 and 5, the original ESPI fringe patterns are obviously very noisy and of limited visibility. Using the previous methods, it is necessary to process the 1000 ESPI fringe patterns for one thousand times. However, using our method, we only need to process few numbers of the training samples. Once the network is trained well, the

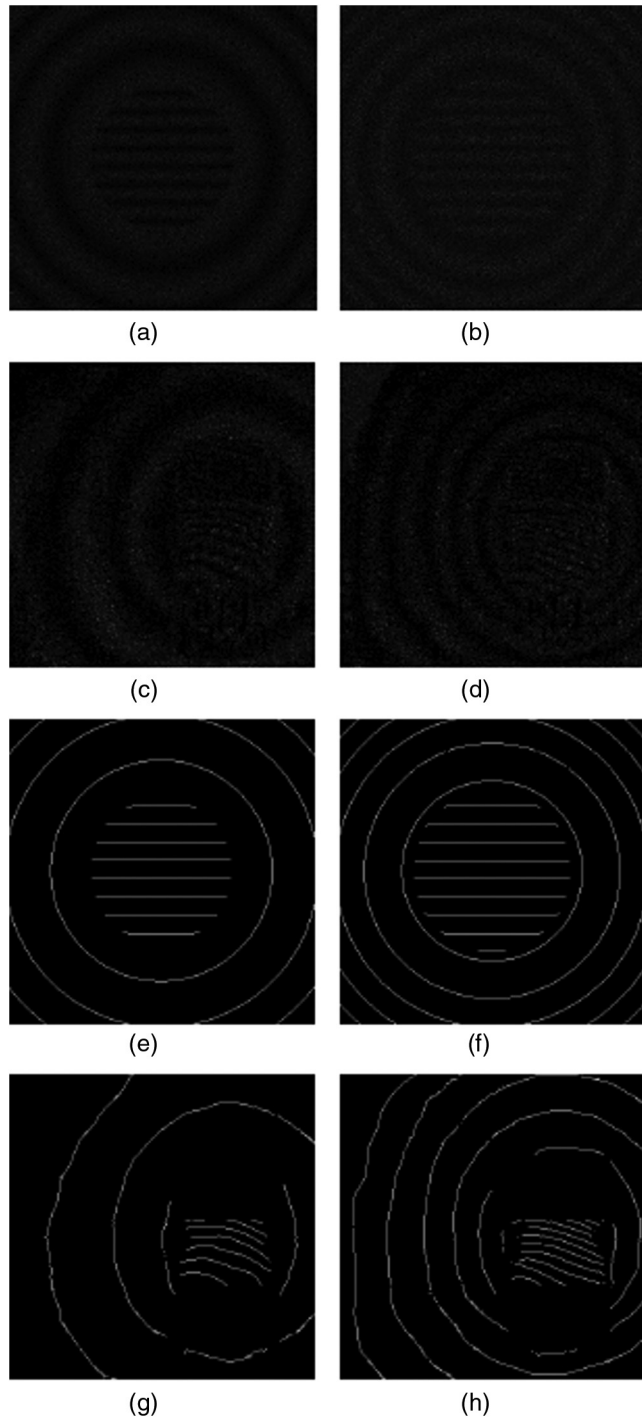


Fig. 8 Parts of the training samples for the experimental dynamic ESPI measurement: (a) and (b) two simulated ESPI fringe images with low quality; (c) and (d) two real ESPI fringe images with low quality; and (e)–(h) the corresponding skeleton images of (a)–(d), respectively.

skeleton images can be obtained in batches. In our trained network, we can simultaneously process 20 noisy ESPI fringe patterns and it takes just 40 s to obtain 20 skeleton images. As can be seen in Fig. 6, the skeleton extraction results by the trained network are desired.

In addition, as shown in Fig. 7, we are surprised to find that even though the shapes of the testing images are slightly different from the training samples, our method also gives desired results. In the simulation test, the AC of the trained network is 0.9727.

3.2 Experimental Dynamic ESPI Measurement

In this part, we apply our network to a real dynamic measurement of a printed circuit board with a chip under normal working conditions by ESPI.

Figures 8(a) and 8(b) show the two experimentally obtained original ESPI fringe images with sizes of 390×390 pixels, which depict the out-of-plane displacements of the board at 1.80 and 2.04 s. As can be seen in Figs. 8(a) and 8(b), the quality of the two experimentally obtained ESPI fringe patterns is very poor because of high noise and low contrast. It takes a lot of effort with specialized technology to obtain the skeleton images. Extracting skeleton from this type of ESPI image is particularly challenging using the previously existing methods.

For the experimentally obtained ESPI fringe patterns, it is very difficult to extract the perfect skeletons. We laboriously obtain six fringe skeleton images of the original noisy ESPI fringe images using the GVF methods based on VID.¹³ The six training samples (six experimentally obtained ESPI fringe patterns and the corresponding skeleton images) are not enough to train the U-Net CNN. Compared with the experimentally obtained ESPI fringe patterns, it is relatively easy to obtain the skeletons of the computer-simulated ESPI fringe patterns. Therefore, we add 10 simulated ESPI fringe patterns to the training samples. The criteria are that: first, the sizes of the simulated ESPI fringe patterns should be same as those of the experimentally obtained ESPI fringe patterns; second, the contrasts and the shapes of the simulated ESPI fringe patterns should be similar to those of the testing images as much as possible. The corresponding skeleton images of the 10 simulated ESPI fringe patterns are obtained using the GVF methods based on VID. The training samples in the test are composed of six experimentally obtained ESPI fringe patterns and 10 simulated ESPI fringe patterns.

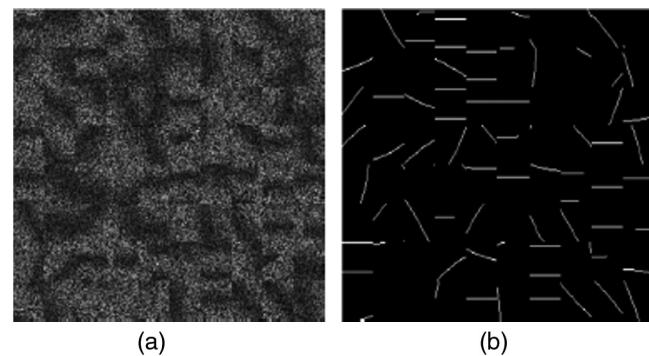


Fig. 9 Example of the input training sample patches and the corresponding skeleton image patches for training the network: (a) 32×32 pixel random portions of noisy ESPI fringe and (b) the corresponding skeletons of (a).

This strategy can be extended to other real-dynamic ESPI measurements.

Each image used in the training process is also cropped into 4800 patches with size of 32×32 pixels. The total number of patches sampled from the training dataset is 76,800. Figure 8 shows the parts of the training samples. Then all of the patches are as input fed to the U-Net CNN, and after 150 iterations, the training process is finished. The sizes of the patch will affect the experimental results. In order to ensure

the number of positive samples, the sizes of the patch should be larger than the width of the fringes. Figure 9 shows an example of the ESPI fringe patterns patches and the corresponding skeleton patches. We apply our trained network to the other real ESPI fringe patterns. Parts of the testing results of the real ESPI fringe patterns are shown in Fig. 10. The AC of the trained network in the test is 0.9514.

For better illustration of our method, we give a comparison of our method with the GVF's method based on VID.¹³

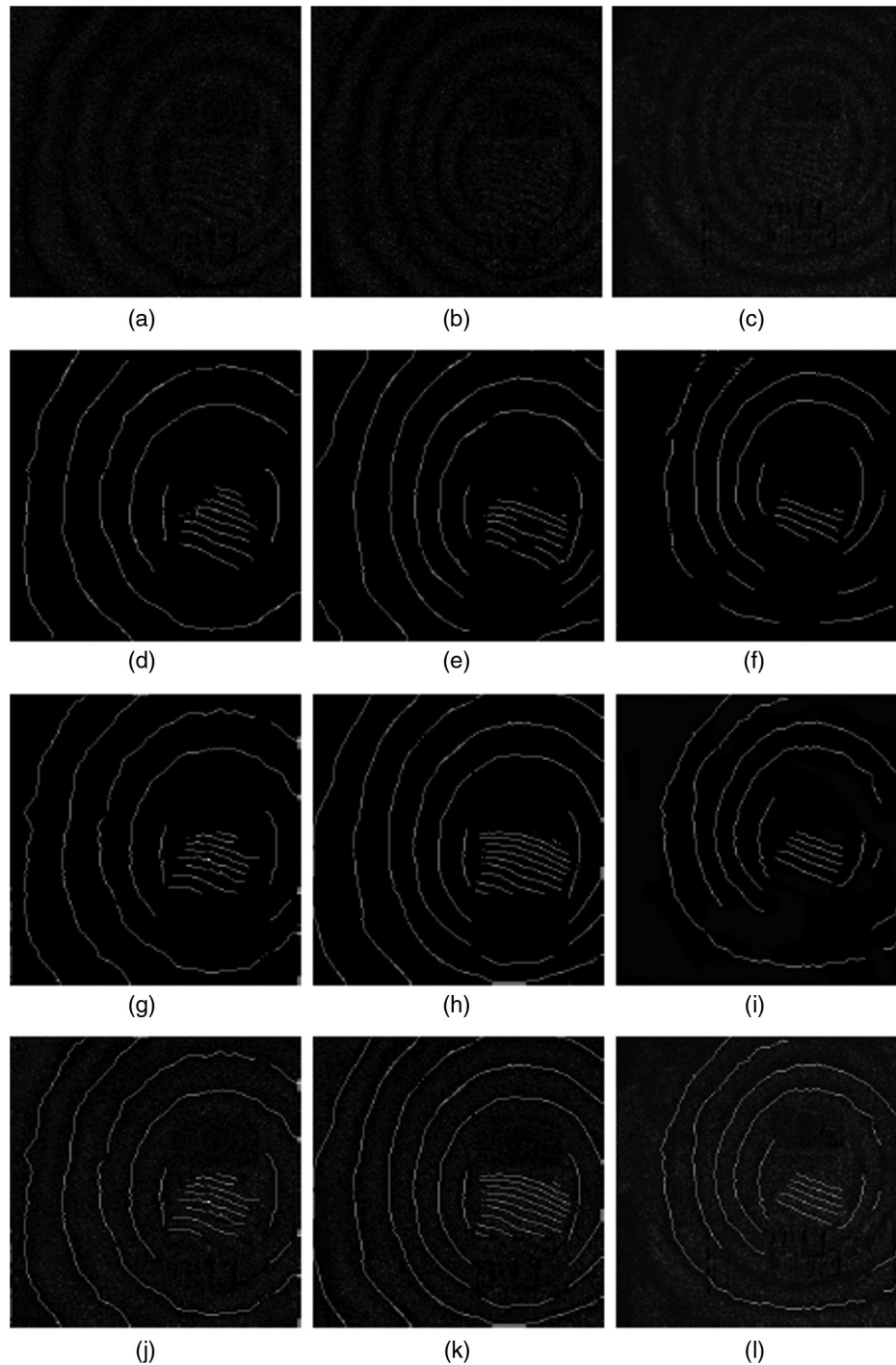


Fig. 10 Skeleton extraction results on the experimental obtained ESPI fringe patterns: (a)–(c) three real ESPI fringe images with low quality; (d)–(f) the corresponding skeleton images of (a)–(c) by GVF's-VID, respectively; (g)–(i) the corresponding skeleton images of (a)–(c) by our method, respectively; (j)–(l) the superimposition of (g)–(i) onto enhanced (a)–(c).

Figures 10(a)–10(c) show the experimentally obtained original ESPI fringe patterns; Figs. 10(d)–10(f) show the skeleton images, which are obtained by the GVF's method based on VID; and Figs. 10(g)–10(i) show the skeleton images, which are obtained by our method. In addition, we enhance the contrast of Fig. 10(a) and superimpose Fig. 10(g) on Fig. 10(a), shown in Fig. 10(j). The similar results of Figs. 10(h) and 10(i) are shown in Figs. 10(k) and 10(l), respectively.

From Figs. 8 and 10, we can observe that the qualities of initial ESPI images are considerably low, because of variable density, high noise, and low contrast. Extracting the skeletons in this case is very challenging. As shown in Figs. 10(d)–10(f), the skeletons obtained by the GVF's method based on VID are satisfactory. But the good results of the GVF's method based on VID are relative to the parameters. The parameters need to be adjusted carefully for each image. By employing our trained network, we can obtain the skeleton images directly.

4 Conclusion

We propose an intelligent skeleton extraction method based on deep learning. Our method has three main advantages: (1) our trained network can obtain the skeleton images of the multiframe noisy ESPI fringe patterns simultaneously, which is very efficient. With the development of computer hardware, we would expect that we can obtain more numbers of skeleton images at one time; (2) our trained network can give desired skeleton results for the noisy ESPI fringe patterns, and the shapes can be slightly different from those of the training samples. (3) In the dynamic ESPI measurement, if the numbers of the training samples are few, one can use some simulated ESPI images as the training samples.

Future areas of further development include optimizing the numbers of the training samples and improving the network structure. This work is currently underway. With the development of AI, we would expect that an intelligent ESPI system can be used to predict the skeleton images for all kinds of fringe patterns, which is quite a step forward in optical measurement.

Acknowledgments

The authors would like to thank the editor and the anonymous reviewers for their constructive and helpful comments on this manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 11772081).

References

1. F. Languy et al., "Vibration mode shapes visualization in industrial environment by real-time time averaged phase-stepped electronic speckle pattern interferometry at 10.6 μm and shearography at 532 nm," *Opt. Eng.* **55**(12), 121704 (2016).
2. C. Xiong et al., "Real-time one-point out-of-plane displacement measurement system using electronic speckle pattern interferometry," *Opt. Eng.* **55**(12), 121721 (2016).
3. R. Zemmamouche et al., "Use of specklegrams background terms for speckle photography combined with phase-shifting electronic speckle pattern interferometry," *Opt. Eng.* **54**(8), 084110 (2015).
4. C. Quan et al., "Phase extraction from a single fringe pattern based on guidance of an extreme map," *Appl. Opt.* **44**(23), 4814–4821 (2005).
5. C. Tang et al., "Performance evaluation of partial differential equation models in electronic speckle pattern interferometry and δ -mollification method of phase map," *Appl. Opt.* **45**(28), 7392–7400 (2006).
6. C. Tang et al., "Denoising by coupled partial differential equations and extracting phase by backpropagation neural networks for electronic speckle pattern interferometry," *Appl. Opt.* **46**(30), 7475–7484 (2007).
7. G. Wang, Y. Li, and H. Zhou, "Application of the radial basis function interpolation to phase extraction from a single electronic speckle pattern interferometric fringe," *Appl. Opt.* **50**(19), 3110–3117 (2011).

8. L. Lam, S. Lee, and C. Suen, "Thinning methodologies—a comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(9), 869–885 (1992).
9. C. Xu and J. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.* **7**(3), 359–369 (1998).
10. C. Tang et al., "Nearly preprocessing-free method for skeletonization of gray-scale electronic speckle pattern interferometry fringe patterns via partial differential equations," *Opt. Lett.* **33**(2), 183–185 (2008).
11. C. Tang et al., "Oriented couple gradient vector fields for skeletonization of gray-scale optical fringe patterns with high density," *Appl. Opt.* **49**(16), 2979–2984 (2010).
12. F. Zhang et al., "Skeleton extraction and phase interpolation for single ESPI fringe pattern based on the partial differential equations," *Opt. Express* **23**(23), 29625–29638 (2015).
13. X. Chen et al., "Gradient vector fields based on variational image decomposition for skeletonization of electronic speckle pattern interferometry fringe patterns with variable density and their applications," *Appl. Opt.* **55**(25), 6893–6902 (2016).
14. P. H. Winston, *Artificial Intelligence an MIT Perspective*, MIT Press, Cambridge, Massachusetts (1986).
15. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
16. C. Lee, A. Tyring, and N. Deruyter, "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," *Biomed. Opt. Express* **8**(7), 3440–3448 (2017).
17. H. Dong et al., "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand and Anal.*, pp. 506–517 (2017).
18. Ö. Çiçek, "3D U-Net: learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.* **9901**, 424–432 (2016).
19. Q. Dou et al., "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.* **41**, 40–54 (2017).
20. W. Li et al., "On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task," *Lect. Notes Comput. Sci.* **10265**, 348–360 (2017).
21. K. Konstantinos et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.* **36**, 61–78 (2017).
22. R. Holger et al., "DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation," *Lect. Notes Comput. Sci.* **9349**, 556–564 (2015).
23. H. Chen et al., "VoxResNet: deep voxelwise residual networks for volumetric brain segmentation," arXiv 1608.05895 (2016).
24. Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.* **15**(5), 749–753 (2018).
25. R. Li, W. Liu, and L. Yang, "DeepUNet: a deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**, 3954–3962 (2018).
26. F. Sawaf and R. Groves, "Statistically guided improvements in speckle phase discontinuity predictions by machine learning systems," *Opt. Eng.* **52**(10), 101907 (2013).
27. Y. Cai, J. Liang, and X. Yu, "A novel automated approach for noise detection in interference fringes pattern images using feature learning," *Proc. SPIE* **10835**, 108350I (2018).
28. N. Anantrasirichai, J. Biggs, and F. Albino, "Application of machine learning to classification of volcanic deformation in routinely generated InSAR data," *J. Geophys. Res.* **123**(8), 6592–6606 (2018).
29. D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent.* (2015).
30. F. Huazhu et al., "DeepVessel: retinal vessel segmentation via deep learning and conditional random field," *Lect. Notes Comput. Sci.* **9901**, 132–139 (2016).

Biyan Li received his BS and MS degrees from Tianjin University in 2013 and 2015, respectively. He is currently working toward his PhD at Tianjin University, Tianjin, China. His current research interests include image processing and applications in optical interferometric fringes.

Chen Tang received her PhD from Dalian University of Technology, Dalian, China, in 1996. She is currently a professor at Tianjin University, Tianjin, China. Her current research interests include image/video processing and research of deep learning algorithm.

Tingyue Zheng received her BS and MS degrees from Tianjin University in 2016 and 2019, respectively. Her current research interests include image processing and deep learning algorithms.

Zhenkun Lei received his PhD from Dalian University of Technology, Dalian, China, in 2002. He is currently a professor at Dalian University of Technology, Dalian, China. His current research interests include optical measurement mechanics and fringe analysis, image processing, and compound material mechanics.