Supplemental Document

applied optics

Channel transformer U-Net: an automatic and effective skeleton extraction network for electronic speckle pattern interferometry: supplement

Biyuan Li,[1,*] Zhuo Li,[1] Jun Zhang,[1] Gaowei Sun,[1] Jianqiang Mei,[1] and Jun Yan[2]

[1]*Tianjin University of Technology and Education, School of Electronic Engineering, Tianjin 300222, China*
[2]*Tianjin University, School of Mathematics, Tianjin 300072, China*
[*]*Corresponding author: lby@tute.edu.cn*

Supplement DOI: https://doi.org/10.6084/m9.figshare.21676565

Parent Article DOI: https://doi.org/10.1364/AO.477083

Supplementary Materials for

Channel transformer U-Net: an automatically and effective skeletons extraction network for electronic speckle pattern interferometry

Biyuan Li*[1] ,Zhuo Li[1],Jun Zhang[1], Gaowei Sun[1], Jianqiang Mei[1] and Jun Yan[2]

1Tianjin University of Technology and Education, School of Electronic Engineering, Tianjin 300222, China

2Tianjin University, School of Mathematics, Tianjin 300072, China

lby@tute.edu.cn

S1. The theory of the Channel-wise Cross fusion Transformer (CCT)

Channel-wise Cross fusion Transformer (CCT) is an encoder feature extractor that utilizes Transformer's long-term modeling advantages. The CCT module consists of three steps: multi-scale feature embedding, multi-head channel-wise cross attention (MCA) and Multi-Layer Perceptron (MLP). The multi-scale feature embedding performs tokenization by convolution process and reshaping the features into sequences of flattened 2D patches.

Example: the inputs of the multi-scale feature embedding step are the features obtained from the encoder noted as $E^{\frac{H}{i^2}\times\frac{W}{i^2}\times C_i}$. Where $H\times W$ represent the size of the feature map, C_i are the channel dimensions of the four skip connection layers. After the flat operation, the 3D features $E^{\frac{H}{i^2}\times\frac{W}{i^2}\times C_i}$ are reshaped into 2D patches $T_i \in R^{\frac{HW}{i^2}\times C_i}$.

The multi-head channel-wise cross attention (MCA) is to extract the detail features of feature images from 2D patches $T_i \in R^{\frac{HW}{i^2}\times C_i}$. Fig S1 shows the main structure of MCA. It maps the output of the embedded $T_i \in R^{\frac{HW}{i^2}\times C_i}$ to K,Q,V space, just like the original Transformer's multi-attentional mechanism. The main difference between the head-crossing attention mechanism and the traditional attention mechanism is that the instance regularization function is passed before inputting Softmax. The reconstruction process is symmetric with the embedding process. The 2D $O_i \in R^{\frac{HW}{i^2}\times C_i}$ obtained from MCA are transformed into 3D $B^{\frac{H}{i^2}\times\frac{W}{i^2}\times C_i}$ by convolution process and inverse flat operation.

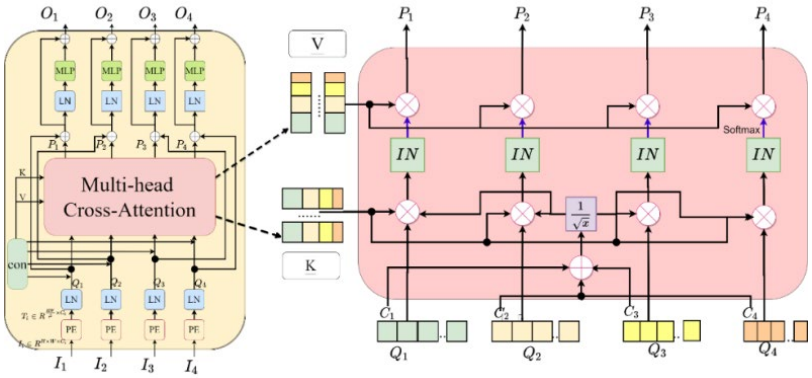


Fig. S1 The flowchart of Multi-head Cross-Attention

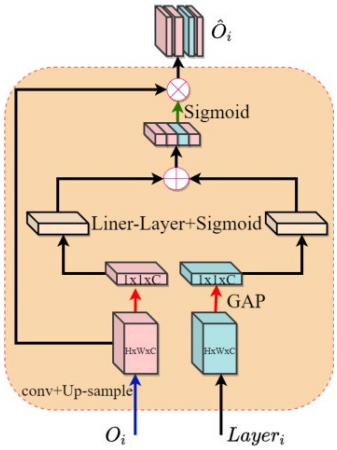


Fig. S2 The flowchart of Channel-wise Cross Attention

The Channel-wise Cross Attention (CCA) is shown in Fig. S2. The transformer outputs $B^{\frac{H}{i^2}\times\frac{W}{i^2}\times C_i}$ and the decoder feature map $D^{\frac{H}{i^2}\times\frac{W}{i^2}\times C_i}$ are fed into Channel-wise Cross Attention. Spatial squeeze is performed by a global average pooling (GAP) layer, we use a single Linear layer and sigmoid function to build the channel attention map $\hat{O}_i$. Finally, the $\hat{O}$ is concatenated with the up-sampled features of the i-th level decoder based on scale parameter η.

S2. The pseudo code fragments and the anti-noise performance analysis of the marking algorithm

We summarize the main steps of the marking algorithm as follows:

Algorithm 1 automatic skeleton marking algorithm

main program

Input: f that is a binary skeleton image.

λ that is the pixel minimum to determine if it is skeleton line.

Output: f' that is the labeled result

$\xi \leftarrow 0, \tau \leftarrow 0.5\pi, \hbar \leftarrow 1$

#Initialize f' to an all-zero matrix of the same size as the input image f

$V \leftarrow serach\_center(f)$

for v_0 in V

$f \leftarrow getdata(v_0)$ #getdata mainly obtains a set of coordinates centered on v

for v_0 in f

$Labeled\_skeleton\_line\left(v_1, \xi\right)$

if $\hbar == 0$

$\xi = \xi + \tau$

endif

endfor

endfor

end

Algorithm 1 is described by two parts of pseudo-code snippets, which we call the main algorithm segment and the sub-algorithm segment. In the main algorithm segment, the first step is to binarize the input skeleton line image; The second step is to filter the image, which aims to filter out the isolated pixels that are not skeleton lines (the number of isolated pixels is determined by the threshold λ). In the third step, we searched the center of the skeleton line. In the fourth step, we search the skeleton line outwards with each center v_0 as the center. If the current skeleton mark ends, update ξ, and then use the updated ξ to mark the new skeleton line.

sub-program

funcation $Labeled\_skeleton\_line\left(v,\xi\right)$

Input: v are the current pixel coordinates. ξ is the mark value of the skeleton line where the current pixel is located.

output:none .

if $f'_{ij}==0\ and\ f_{ij}==1$

$f'_{ij}\leftarrow\xi$

$\hbar\leftarrow 0$

endif

for v' in the eight domains of v

if v' do not go beyond the graph

$Labeled\_skeleton\_line\left(v',\xi\right)$ #Recursively marks skeleton lines

endif

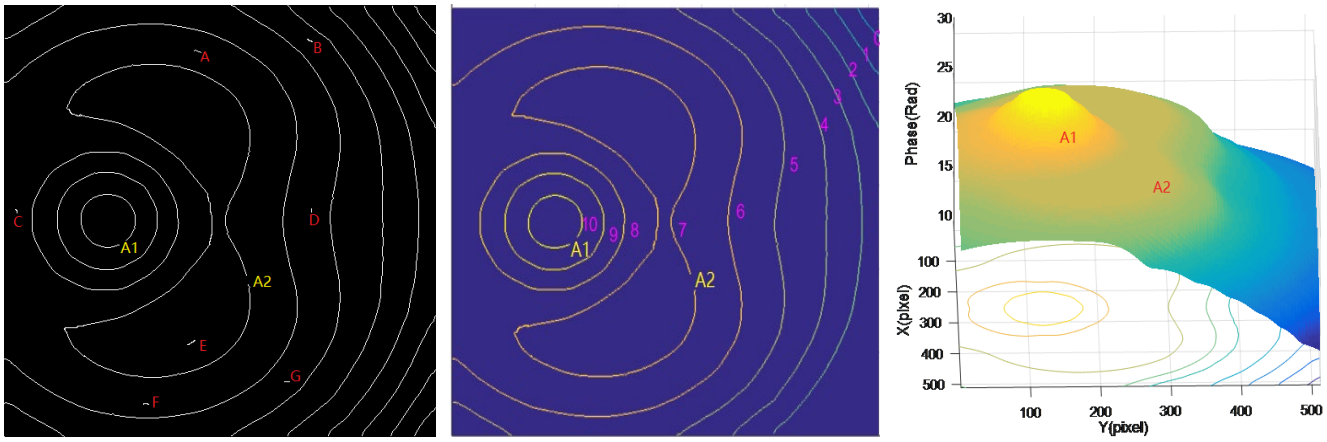
endfor

end.

In the sub-algorithm segment, the first step is to judge whether the current skeleton line pixel has been marked. If there is no mark, the current pixel will be marked. The second step is to recursively search eight directions in turn centering on the current pixel coordinates until the current skeleton line is marked and returned to the main algorithm.

Variable description in Algorithm 1: ξ in the main algorithm fragment is the current skeleton line identifier value; τ is the step length of the line marking. In this paper, we set the step length of the adjacent skeleton line as. The value of $\hbar$ indicates whether the identification of the current skeleton line is complete. If $\hbar$ is equal to 0, it means that the identification of the current skeleton line is complete. Then update ξ to mark the new skeleton line. λ means that the larger the image filtering coefficient is, the stronger the ability to filter out the noise of the isolated non-skeleton line is, but too large a number of skeleton line pixels will be lost. The experiment shows that the value of 30 is more appropriate.

It is well known that all of the marking algorithms can be susceptible to noise. Hence, even if small residual effect of noise is present after fringe skeletonizing, errors can still accumulate. Here we explore the anti-noise performance of the proposed marking algorithm. Fig. S3(a) is the fringe skeleton image with isolated pixels and broken lines. Fig. S3(b) is the labeling results of Fig. S3(a). Fig. S3(c) is the interpolation result of Fig. S3(b).



(a) (b) (c)

Fig. S3 Anti-noise performance analysis of the marking algorithm

As can be seen is Fig. S3(a), the skeleton image contains isolated pixels and broken areas. This situation is due to the noise interference in the process of binarization. As shown in Fig. S3(b), the proposed marking algorithm give the desired result. All the fringe levels are labeled correctly. In addition, the interpolation result shown in Fig. S3(c) also shows that the proposed marking algorithm has a good ability to resist noise.

S3. The detailed simulation formula of ESPI fringe patterns

As a supervised deep learning model, CTransU-Net requires to be trained by the noisy ESPI fringe patterns and the corresponding skeleton maps. In order to make the training effective and reliable, we construct an available dataset for

ESPI skeleton extraction by means of computer-simulated method. The original noisy ESPI fringe patterns are simulated with the following equation.

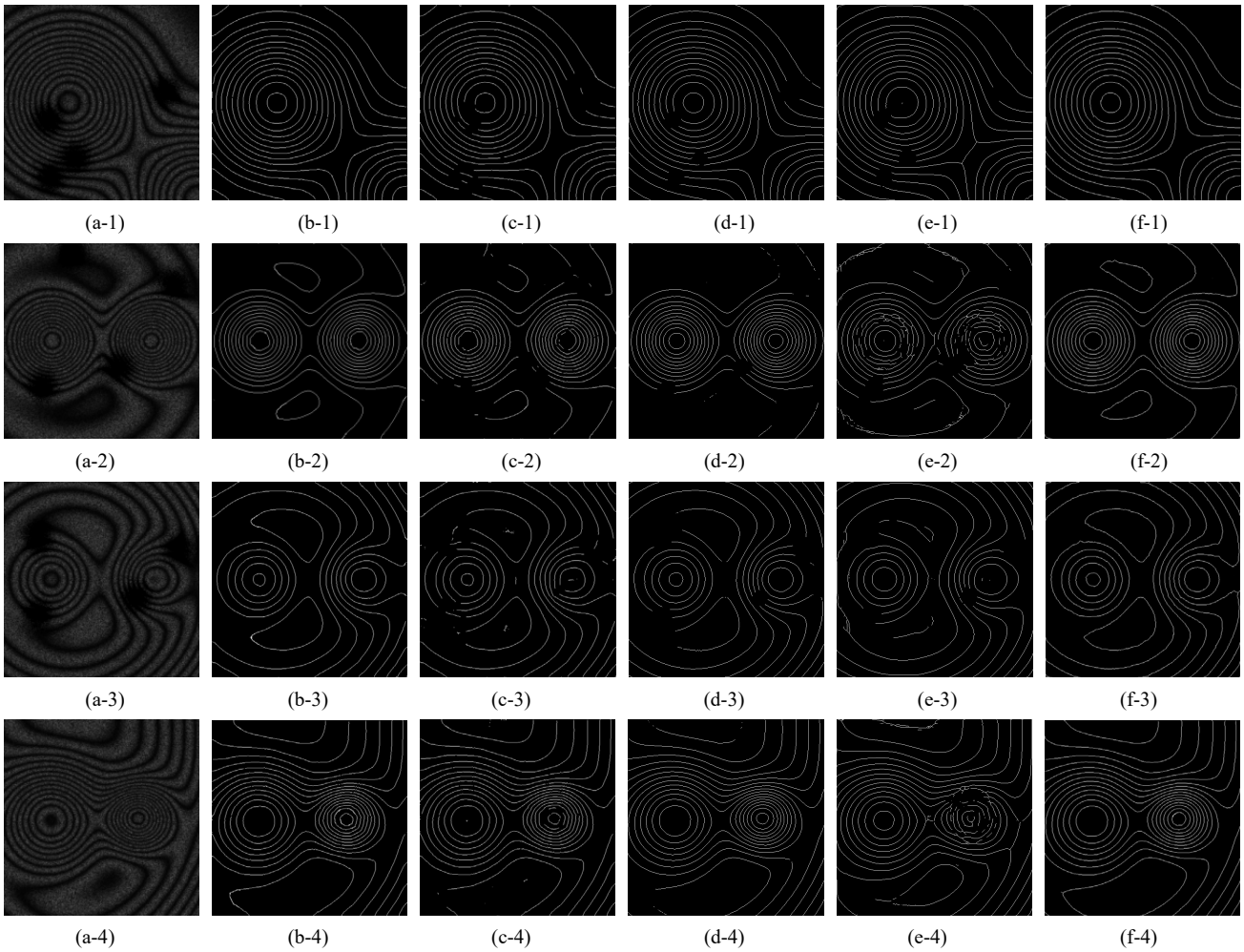
$$I_{sub} = \left| 4\sqrt{I_0 I_r}\sin(\varphi_r - \varphi_0 + \frac{\psi_{ij}}{2})\sin\left(\frac{\psi_{ij}}{2}\right)\right| + N * randoms(m,n) \tag{1}$$

where, $\varphi_r - \varphi_0$, I_0 and I_r are taken as random variables with values uniformly distributed over the intervals $[-\pi,\pi]$、$[0,I_m]$ and $[0,\rho I_m]$, where I_m is a constant value, ρ is a normalized visibility parameter. Here, $I_m = 90$、$\rho = 0.2$. $N * randoms(m,n)$ noise intensity parameter. N=30, $m = 512$ $n = 512$。 ψ_{ij} represents the relative phase difference, and different ψ_{ij} generate different shapes of the fringe patterns. One of the ψ_{ij} is designed by

$$\psi_{ij} = \alpha \times \left(\exp(-\frac{(2i-m)^2 + \left(2j - \frac{11n}{8}\right)^2}{50000}) \right) + \exp\left(-\frac{(2i-m)^2 + \left(2j - \frac{n}{2}\right)^2}{35000} \right) + 10 \times \left(\left(\frac{6i-3m}{2m}\right)^2 + \left(\frac{3j-n}{n}\right)^2 \right) \tag{2}$$

S4. Qualitative evaluate from the Visual Effects

We apply the trained network to a series of ESPI fringe patterns. We compare the proposed method with VID-GVFs, and U-Net method and M-Net method. The tests are shown in Fig. S4.



(a-1) (b-1) (c-1) (d-1) (e-1) (f-1)

(a-2) (b-2) (c-2) (d-2) (e-2) (f-2)

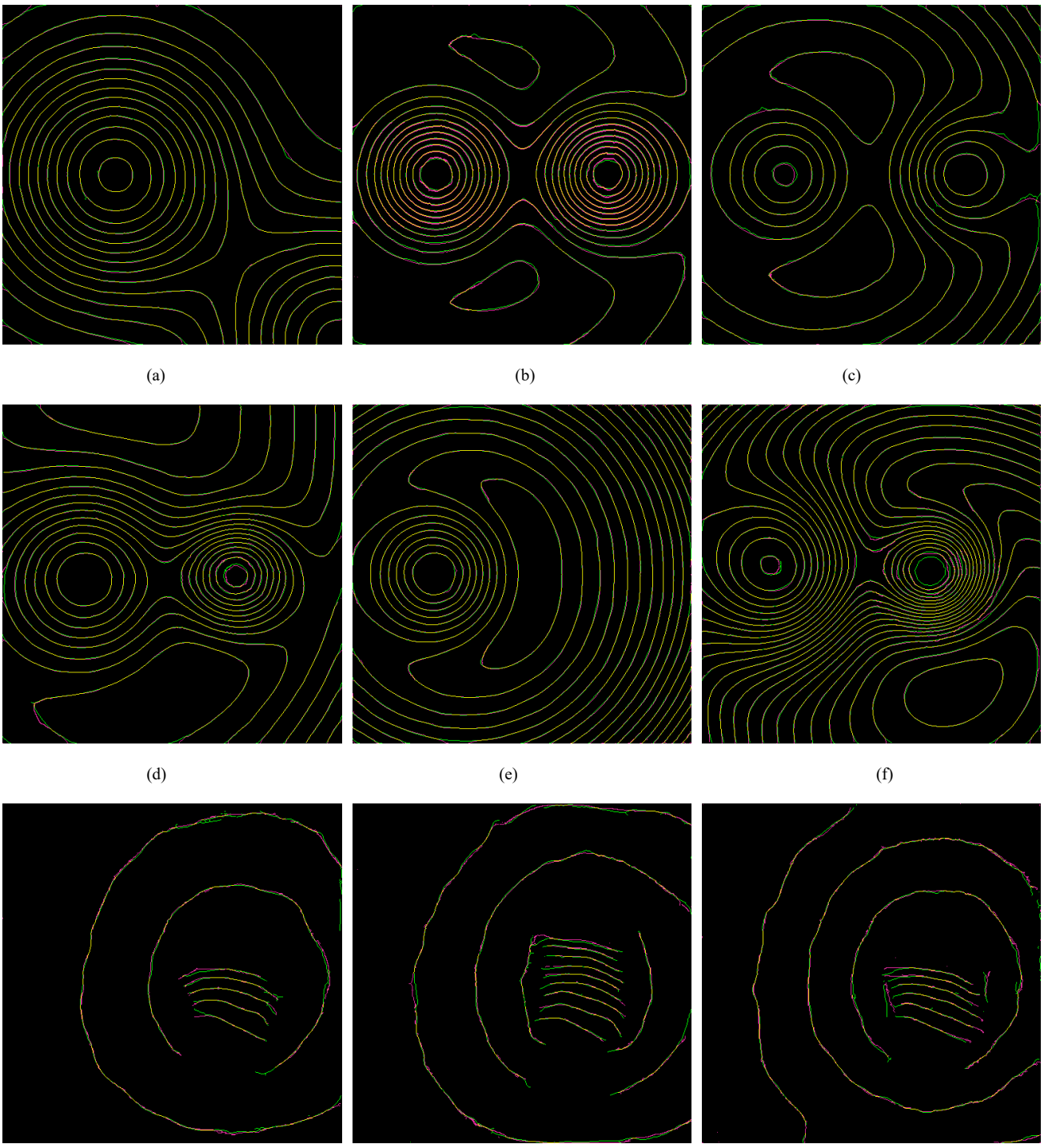
(a-3) (b-3) (c-3) (d-3) (e-3) (f-3)

(a-4) (b-4) (c-4) (d-4) (e-4) (f-4)

Fig. S4 Experimental fringe skeleton line extraction results: (a-1)-(a-9) experimentally obtained original ESPI fringe images. (b-1)-(b-9) corresponding ideal skeletons; (c-1)-(c-9) results of VID-GVFs; (d-1)-(d-9) results of U-net; (e-1)-(e-9) results of M-Net; (f-1)-(f-9) results of the proposed method

S5. The superposition results of Fig. S4

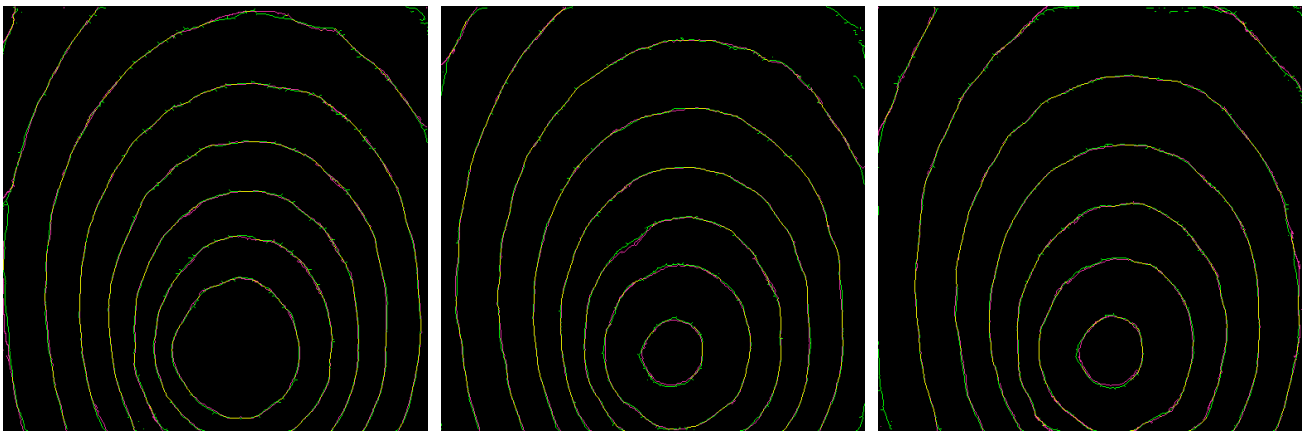
Figs. S5(a-L) are obtained by the superposition of the labeling images (Figs. S4(b-l)-(b-9)) and the predicted images (Figs. S4(f-l)-(f-9)), respectively. From the superposition results, it can be seen that the yellow lines (pixels that are completely classified correctly) are located in the main areas of skeleton line. Most of pink pixels and green pixels (pixels that are divided incorrectly) are concentrated at the edges of skeletons and these pixels have little influence on the phase extraction.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)



(j) (k) (L)

Fig. S5 Superposition results of Fig. 4 in pixel level

S6. The related statistics information of the four test methods

Table S1 shows the related statistics information about the comparison methods which used in Fig. S4. As shown in Table S1, the U-Net, M-Net and our method use the same training dataset. The loss function and number of iterations are set based on better performance shown in original literature. According to Table S1, one can find that our method achieves better results with fewer epochs. In conclusion, the highest network metrics confirm the strong capability of our method in detecting the skeleton pixels correctly.

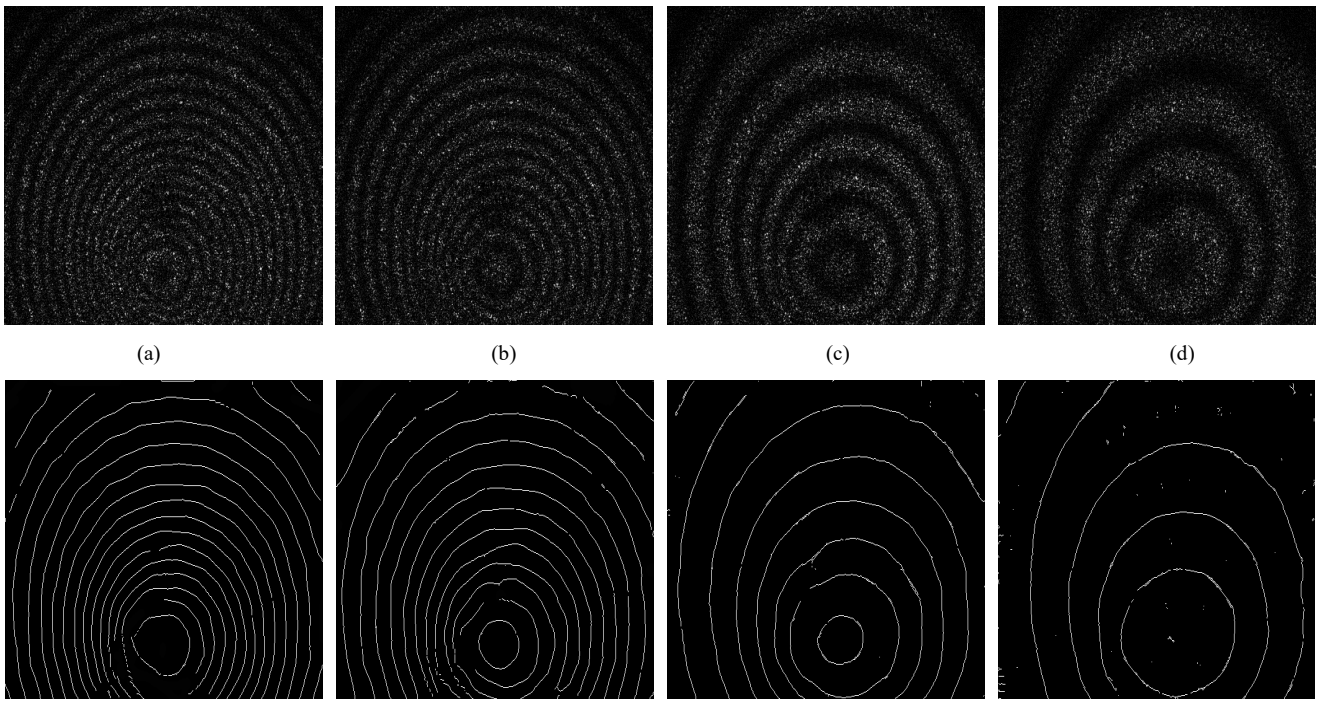
Table S1 The related statistics information about the comparison methods

Method	Pre-processing	Training	Number of training datasets	Number of test images	Loss function	epochs
VID-GVF	need	needless	none	100	none	none
U-Net	needless	need	30	100	cross entropy loss function	150
M-Net	need	need	30	100	focal loss+SSIM	150
CTransU-net	needless	need	30	100	PolyFocal+BCE	12

S7. Phase extraction of thermal deformation of alumina ceramics under laser irradiation

Fig. S6 shows part of the training dataset for the thermal deformation of Al_2O_3. Figs. S6(a)-(d) show four experimentally obtained Al_2O_3 ceramic ESPI fringe patterns with sizes of 512 × 512. As can be seen in Figs. S6(a)-(d), the quality of 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. Figs. S6(e)-(h) show the corresponding skeleton images of Figs. S6(a)-(d).

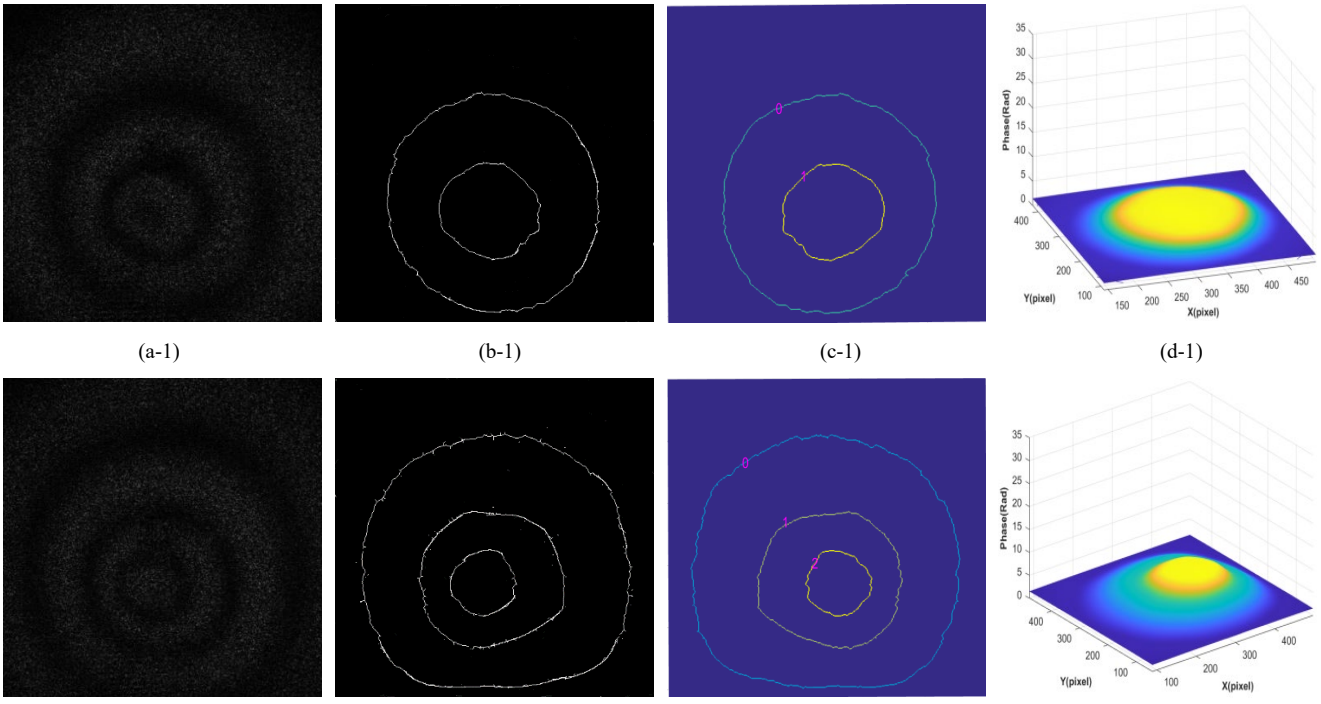
Fig. S7 shows part of the test results. Figs. S7(a-1)-(a-8) show parts of the test ESPI fringe patterns. Figs.7(b-1)-(b-8) shows the corresponding skeleton images of Figs. S7(a-1)-(a-8), respectively. Figs. S7(c-1)-(c-8) show the fringe level by the proposed the marking algorithm. In the previous works, the skeleton is marked manually and this is a time-consuming and laborious task. By using the proposed marking algorithm, all the skeleton can be marked automatically. Figs. S7(d-1)-(d-8) show the phase values of Figs. S7(a-1)-(a-8) by the interpolation method, respectively. In order to show the thermal deformation under heating more clearly, the phase values in column 256 in Figs. S7(d-1)-(d-8) are given in Fig. S8.



(a) (b) (c) (d)

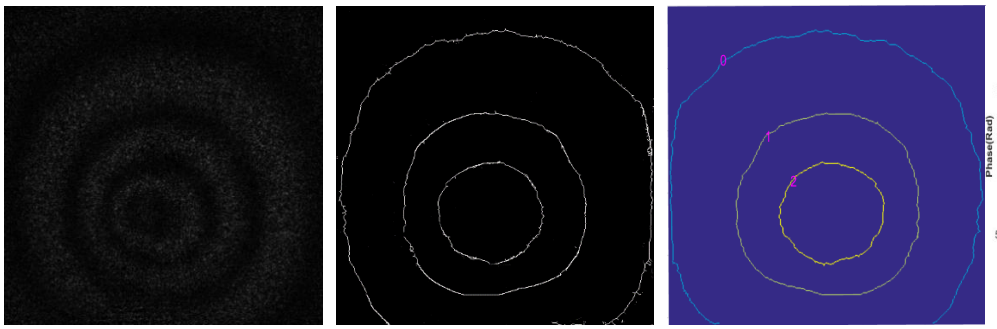
(e) (f) (g) (h)

Fig. S6 Parts of the training samples for thermal deformation of Al2O3 ceramic substrate: (a)-(d) real ESPI fringe images with low quality;(e)-(h) the corresponding skeleton images of (a)-(d), respectively

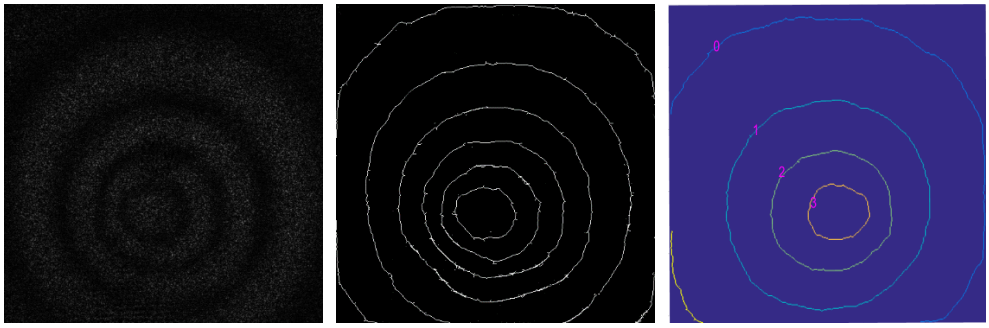


(a-1) (b-1) (c-1) (d-1)

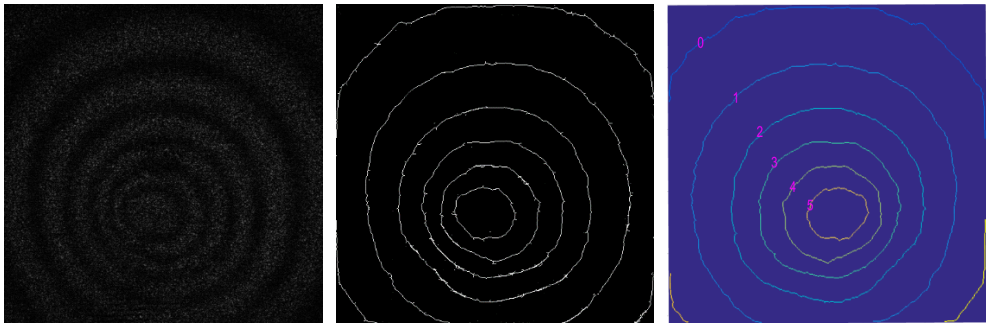
(a-2) (b-2) (c-2) (d-2)



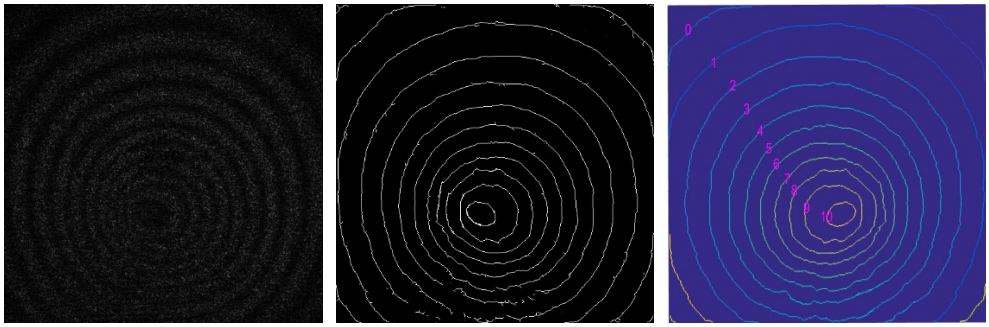
(a-3) (b-3) (c-3) (d-3)



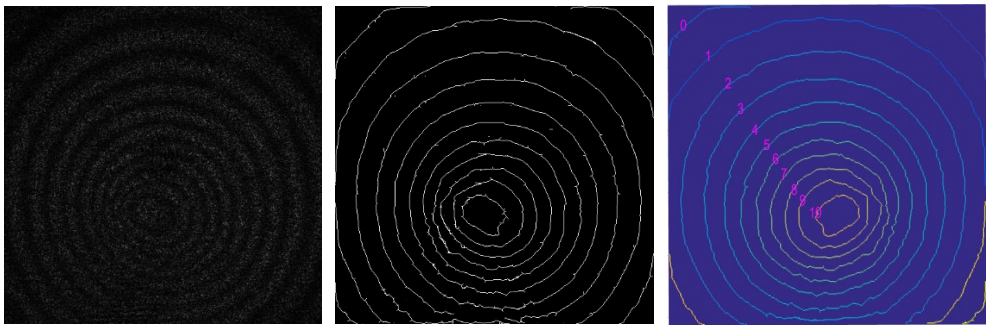
(a-4) (b-4) (c-4) (d-4)



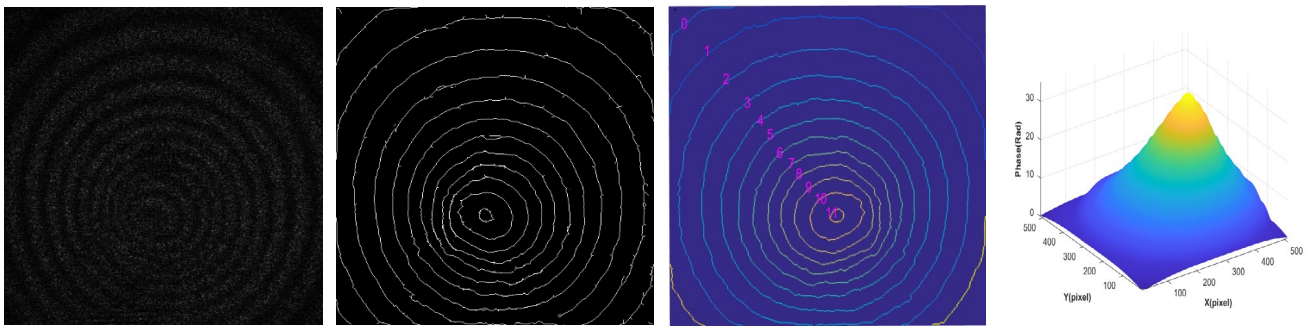
(a-5) (b-5) (c-5) (d-5)



(a-6) (b-6) (c-6) (d-6)



(a-7) (b-7) (c-7) (d-7)



(a-8) (b-8) (c-8) (d-8)

Fig. S7 Al2O3 ceramic ESPI experimental result

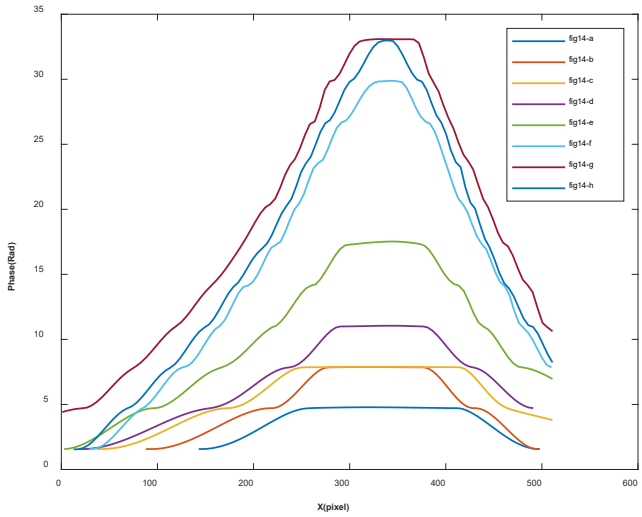


Fig. S8 the phase values in column 256