An End-to-end speckle matching network for 3D deformation measurement

Yahong Feng[1, a], Lianpo Wang[1, 2, b *]

[1]School of Software, Northwestern Polytechnical University, Xi' an, 710129, China

[2]Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen, 518063, China

[a]fengyh@mail.nwpu.edu.cn, [b]wlp@nwpu.edu.cn

Abstract. Digital image correlation (DIC) is a speckle image-based optical measurement technique for measuring deformation of object. In recent years, deep learning has been widely used in 2D-DIC, while research in the field of 3D-DIC is in its infancy, although 3D-DIC can measure 3D shape and deformation compared to 2D-DIC. 3D-DIC achieves 3D deformation measurement through temporal matching and stereo matching, and it is difficult to perform two matching tasks through an end-to-end network due to the different deformation types of them. To solve this problem, we propose an end-to-end speckle matching network for 3D deformation measurement, called 3D-DICNet. Considering that the difference in deformation types between the two matching tasks is mainly manifested in the deformation scales, we extract the features of different receptive fields, propose a new attention connect volume and multi-scale cost aggregation to achieve the deformation measurements at different scales. Experimental results show that the network can perform 3D deformation measurement with high accuracy and efficiency.

Keywords: 3D digital image correlation; deformation measurement; deep learning; optical measurement.

1. Introduction

Accurate measurement of 3D deformation of materials and structures under different loads is very important in the experimental characterization of mechanical behavior. As a non-contact optical measurement method, 3D-DIC has been widely used in experimental mechanics [1] and other scientific fields. The method combines DIC and stereo vision: The DIC obtains the in-plane displacement of the object by matching the speckle images before and after loading (temporal matching); Based on the principle of stereo vision [2], two calibrated cameras simultaneously record images from different angles, and calculate the disparity between the left and right images to obtain 3D position information of the object (stereo matching). As illustrated in Fig. 1.

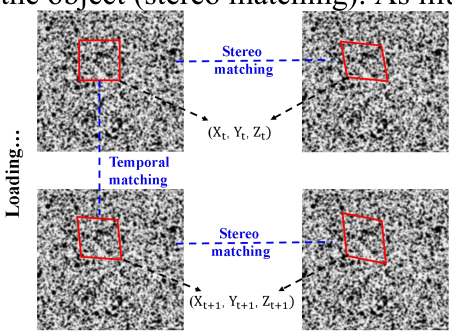


Fig. 1 The schematic of 3D-DIC for 3D deformation measurement

High-precision matching algorithms are essential for the measurement accuracy of DIC. At present, DIC widely uses the inverse combination Gauss-Newton (IC-GN) algorithm [3] with zero mean normalized difference sum (ZNSSD) correlation criterion for subpixel matching. IC-GN has high computational efficiency and good robustness, which can meet the quasi real-time measurement of 2D-DIC. However, compared to 2D-DIC, 3D-DIC requires not only temporal matching, but also stereo matching, which means that matches in 3D-DIC takes twice as long as 2D-DIC. In addition, stereo matching uses the IC-GN algorithm with second-order shape function, which is twice the deformation vector widely used in 2D-DIC, so the traditional 3D-DIC method is far from meeting the needs of real-time measurement.

In recent years, deep learning has been applied to 2D-DIC measurements [6] and speckle structured light (SLL) [7], which has greatly improved the computational efficiency and measurement accuracy. Essentially, both 2D-DIC and SSL are aim to find a matching relationship between the corresponding points, however, networks used in different fields have different architectures. At present, there are few researches on deep learning-based 3D-DIC. As far as we know, only StrainNet-3D [8] was proposed for 3D-DIC, this network achieves very good measurement accuracy for small in-plane displacement of about 2 pixels, but for disparity larger than 2 pixels, it is necessary to pre-process and correct the image pairs to be matched using traditional methods. The traditional methods not only affect the measurement accuracy, but also lead to a complex and time-consuming measurement process. Therefore, we believe that an end-to-end speckle matching network that can be used for two matching tasks in 3D-DIC has great significance for simplifying and accelerating 3D deformation measurements.

This paper aims to propose an end-to-end speckle matching network for temporal and stereo matching. In order to enable the network to measure deformation at different scales in temporal and stereo matching, the contribution of this paper is as follows: First, residual layers and spatial pyramid pooling layers are adopted to extract features with different receptive fields. Then, an attention connect volume that can represent the similarity information at different deformation scales is generated. Finally, a stacked hourglass-shaped network is adopted to ensure the accuracy of displacement prediction.

The rest of this article is organized as follows: Section 2 details the speckle matching network proposed in this paper. In section 3, simulation experiments and physical experiments are used to verify the measurement accuracy, efficiency and applicability of the network. The section 4 concludes this paper.

2. The proposed method

In this paper, we propose 3D-DICNet, aimed at end-to-end perform temporal and stereo matching in 3D-DIC. The 3D-DICNet consists of three modules: i. Feature extraction: using CNN with shared weights and spatial pyramid pools to extract image features with different receptive fields. ii. Cost volume generation: combine the advantages of connect volume and correlation volume, attention connect volume are proposed to represent the similarity information at different deformation scales. iii. Displacement prediction: hourglass-shaped sub-net are repeatedly processing cost volume in a top-down and bottom-up manner to predict displacement. Fig. 2 shows our network architecture, and below we will detail each of module.

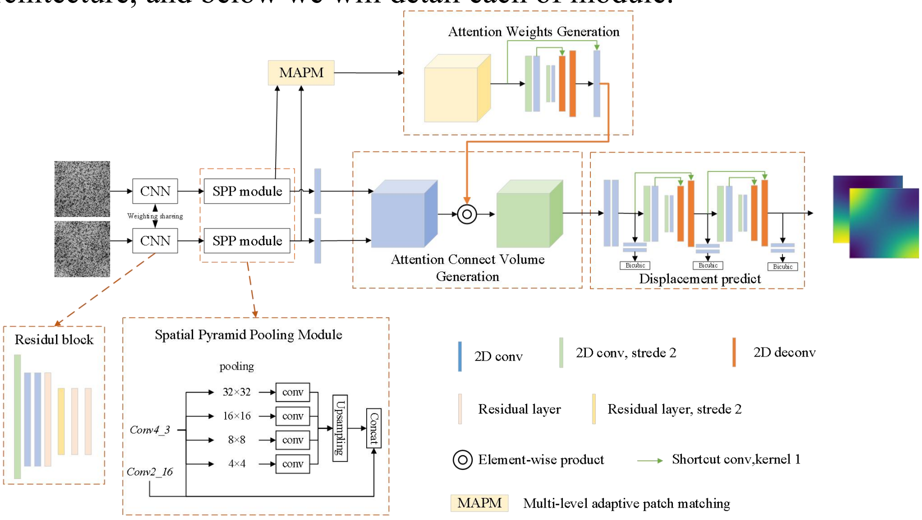


Fig. 2 The network structure of 3D-DICNet

2.1 Feature extraction

Features with different receptive fields are suitable for deformation measurement at different scales, features with large receptive fields make large deformation measurement more robust, and features with small receptive fields are suitable for accurate small deformation measurements. In order to enable the network to handle deformation at different scales, we use a multi-scale residual pyramid network to process the input speckle image pairs. The residual layers with different expansion ratios and the average pooling layers with different convolution sizes are used to obtain the feature of different receptive fields.

2.2 Cost volume generation

In a deep learning-based matching task, the cost volume provides pixel similarity information. Common cost volume includes connect volume and correlation volume. The connect volume directly connect the features in the channel dimension, which is generally used for small deformation measurement. The correlation volume can be used for large deformation measurement by calculating the correlation pixel-by-pixel. In order to enable the network to measure deformation at different scales, we hope to combine the advantages of connect volume and correlation volume, filter the connect volume by extracting attention weights from the correlation volume to generate an attention connect volume.

The process of generating an attention connect volume consists of three steps: building initial connection volume, attention weight generation, and attention filtering. Firstly, the features extracted from the images were connect in the channel dimension to obtain the connect volume. Then, in order to adapt to the deformation at different scales, we adaptively select patches of different sizes in the deformed features, the correlation volume was generated through the correlation operations of patch-by-patch, and use the hourglass-shaped network to learn the similarity information of correlation and obtain attention weight. And finally, use attention weight to filter connect volume to obtain attention connect volume.

2.3 Displacement prediction

The hourglass-shaped network aggregates the similarity information of cost volume at different scales, and improves the measurement accuracy of deformation at different scales. Therefore, a stacked hourglass-shaped network is used to predict the displacement. As shown in Fig. 2, the output of each hourglass-shaped sub-net participates in the training of the network.

2.4 Network training strategy

Supervised learning strategy requires real labels to continuously optimize network parameters. Since it is difficult to obtain a large number of accurate in-plane displacement and disparity from real experiments, the multi configuration stereo speckle image generation algorithm proposed by Wang et al. [8] was used to generate dataset in our work.

The network is trained on the device with NVIDIA GeForce 3090, using a batch size of 8, with a learning rate starting at 1^{e-4} and Adam is used to perform gradient descent. $Smooth_{L1}$ as a loss function for network training.

2.5 Calculate 3D deformation

As shown in Fig. 2, the output of the network is not a 3D deformation field, but in-plane displacement or disparity. When the camera calibration parameters and disparity are known, the depth information of the object can be calculated based on the triangulation principle, as shown in Eq. 1. Where Z_W is depth, b is the binocular camera baseline distance, f is the camera focal length, and D is the disparity.

$$Z_W = \frac{b\,f}{D}. \tag{1}$$

Combined with the in-plane displacement that reflects the two-dimensional deformation information, the deformation in three-dimensional space can be obtained.

3. Experiment

3.1 Simulation experiment

In order to verify the matching accuracy of the network, we compare the proposed 3D-DICNet with the traditional DIC method (SIFT+IC-GN2: SIFT for initial interpixel estimation, IC-GN1 for subpixel iteration, FFTCC+IC-LM2: FFTCC for initial interpixel estimation, and IC-LM2 for subpixel iteration). Since the measurement accuracy of the traditional DIC method is greatly affected by some prior parameters, after many trials, the subset radius is set to 16 and the step size is 5 in SIFT+IC-GN2, and FFTC+IC-LM has a subset radius of 13 and a step size of 5. Fig. 3 shows the temporal and stereo matching results of the three methods, as well as the corresponding error maps.

From Fig. 3, it can be seen that 3D-DICNet has better matching accuracy when dealing with stereo matching, but traditional DIC methods have poor accuracy; For temporal matching, 3D-DICNet has comparable matching accuracy to traditional methods.

In addition, we compared the computational efficiency with the traditional method on the same device, as shown in Table 1, it can be seen that our method is an order of magnitude more computationally efficient than the traditional method.

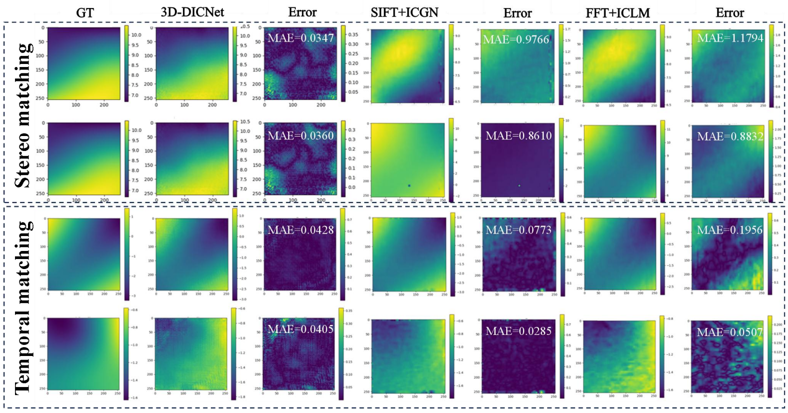


Fig. 3 Comparison of measurement results between 3D-DICNet and two traditional DIC methods in temporal and stereo matching. From top to bottom show the disparity at different moments, the in-plane displacement in the *x* and *y* directions

Table 1. Comparison of the efficiency of 3D-DICNet and traditional methods

Method	Computional efficiency (POI/S)
SIFT+IC-GN2	1.04×10^4
3D-DICNet	4.17×10^5

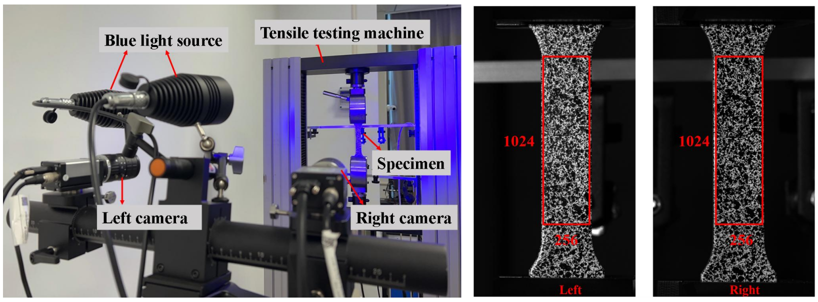
3.2 Physical experiment

In physical experiments, a binocular tensile experiment was designed to verify the performance of the 3D-DICNet in actual 3D mechanical deformation measurement. In this experiment, the tensile testing machine stretched the sample (Q235 mild steel specimen sprayed with black and white speckle patterns) at a speed of 10 mm/s, and a binocular camera (MV-A7A20MU201 with a 50m lens) with synchronous triggering captures images at 25 frames per second, as shown in Fig. 4 (a). Before measuring, the binocular camera needs to be calibrated, the epipolar rectified left and right camera images and ROI are shown in Fig. 4 (b).

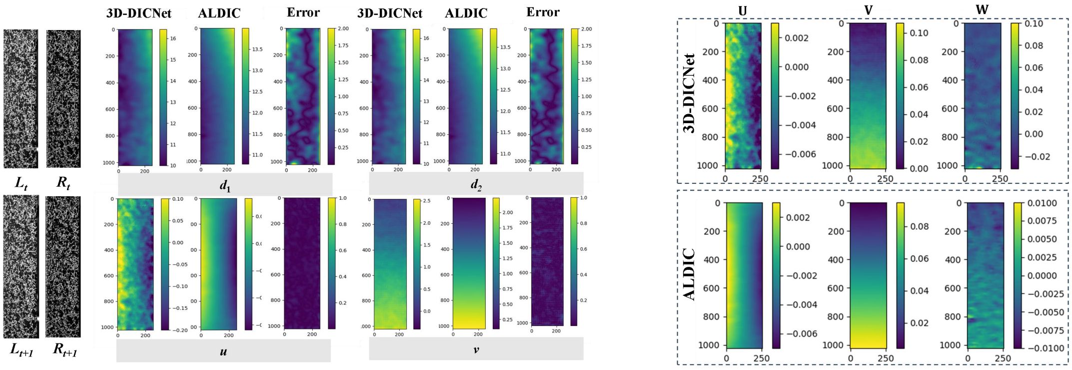
In order to ensure the generalization ability of the model, we used transfer learning to fine-tune the trained model. Fine-tuning models and ALDIC are used for temporal and stereo matching, respectively. We performed multiple tests, and Fig. 5 (a) gives one of the set of measurements , d_l

and d_2 are disparity before and after deformation, respectively. u and v are in-plane displacement in the x and y directions, respectively. Combining the calibration parameters of the binocular camera, the 3D deformation is calculated as shown in Fig. 5 (b).

Fig. 4 (a) Experimental device for binocular material stretching experiments (b) The epipolar



corrected left and right camera images and regions of interest (Red box)



(a) (b)

Fig. 5 (a) disparity and in-plane displacement predicted by 3D-DICNet and ALDIC methods (b) 3D deformation field calculated by 3D-DICNet and ALDIC methods

Compared with the ALDIC measurement results, the bias in the U, V, W three directions are 0.0011 pixels, 0.0021 pixels, and 0.0092 pixels, respectively. Real tensile experiments have shown that the proposed network has good 3D deformation measurement accuracy, at the same time, the network has good generalization ability and applicability.

4. Summary

In this paper, we propose an end-to-end speckle matching network 3D-DICNet for 3D deformation measurement. The results of simulation and physical experiments indicate that the network can be used for deformation measurement at different scales in temporal and stereo matching. Binocular tensile experiments showed, the MAE of 3D deformation measurement in the U, V, and W directions are 0.0011 pixels, 0.0021 pixels, 0.0092 pixels respectively. The 3D-DICNet proposed in this article has the following advantages:

(1) Compared with traditional 3D-DIC methods, the proposed 3D-DICNet method not only has higher efficiency, but also has higher measurement accuracy for stereo matching.

(2) Compared with other deep learning-based 3D-DIC method, the proposed network can be used for deformation measurement at different scales, and can be used for end-to-end temporal and stereo matching.

(3) Although the network is trained on synthetic data, after transfer learning, it can achieve good performance in physical experiments and show strong generalization ability.

Acknowledgements

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (2022A1515110036), National Natural Science Foundation of China (12302245), Natural Science

Basic Research Program of Shaanxi Province (2023-JC-QN-0026), and Jiangsu Shuangchuang Talent Plan Project (JSSCBS20220943).

References

[1] Zhang C, Li D, Ma J, et al. Dynamic shear fracture behavior of rocks: insights from three-dimensional digital image correlation technique[J]. Engineering Fracture Mechanics, 2023, 277: 109010.

[2] Hamzah R A, Ibrahim H, Hassan A H A. Stereo matching algorithm for 3D surface reconstruction based on triangulation principle[C]//2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2016: 119-124.

[3] Pan B, Li K, Tong W. Fast, robust and accurate digital image correlation calculation without redundant computations[J]. Experimental Mechanics, 2013, 53: 1277-1289.

[4] Yin W, Ji Y, Chen J, et al. Initializing and accelerating Stereo-DIC computation using semi-global matching with geometric constraints[J]. Optics and Lasers in Engineering, 2024, 172: 107879.

[5] Xue Y, Cheng T, Xu X, et al. High-accuracy and real-time 3D positioning, tracking system for medical imaging applications based on 3D digital image correlation[J]. Optics and Lasers in Engineering, 2017, 88: 82-90.

[6] Yang J, Qian K, Wang L. R^3-DICnet: an end-to-end recursive residual refinement DIC network for larger deformation measurement[J]. Optics Express, 2024, 32(1): 907-921.

[7] Dong Y, Yang X, Wu H, et al. Lightweight and edge-preserving speckle matching network for precise single-shot 3D shape measurement[J]. Measurement, 2023, 210: 112549.

[8] Wang G, Zhang L, Yao X. StrainNet-3D: Real-time and robust 3-dimensional speckle image correlation using deep learning[J]. Optics and Lasers in Engineering, 2022, 158: 107184.