

Precise Displacement Measurement of Long-Span Spatial Structure of Buildings via Deep Learning and Machine Vision Technology

YINGHAO KONG

*College of Civil Engineering and Architecture
Shandong University of Science and Technology
Qingdao, 266590 P.R. China
E-mail: 201801041204@sdust.edu.cn*

At present, there are still problems of inaccurate labeling, inaccurate target matching, and excessively complex image processing of machine vision in spatial structure displacement measurement. Therefore, machine vision measurement technology is combined with deep learning algorithm to construct a precise displacement measurement model for large-span spatial structure of buildings based on machine vision technology and deep learning. First, the binocular stereo vision technology based on machine vision is expounded to analyze the deficiencies in measuring displacement of large-scale spatial structures. Second, related concepts of Convolutional Neural Networks (CNNs) are introduced, and You Only Look Once v2 (YOLOv2) is selected as the recognition algorithm of sphere joints in the long-span spatial structure of buildings. Finally, a displacement measurement system for long-span spatial structures is constructed based on CNN models and YOLOv2 algorithm. Images of various spatial structures collected through the Internet and shot by smartphones constitute the data set to test the performance of YOLOv2 recognition algorithm and three kinds of CNN models, namely the AlexNet model, Darknet19 model, and Resnet152 model. The experimental results demonstrate that the Recall of Darknet19 model is up to 93.9%, and the Intersection Over Union (IOU) is 82.02%, which are at least 3.6% and 4.4% higher than those of the other two models, respectively. Besides, the Darknet19 model performs better in the task of identifying and locating sphere joints in spatial structures. In addition, when the Darknet19 model is trained about 20,000 times, the Recall of the YOLOv2 recognition algorithm reaches 94.69%, and the IOU attains 84.98%. Moreover, the average recognition accuracy of YOLOv2 algorithm is 2.3% higher than other recognition algorithms, which has a high recognition level. Furthermore, the relative error of displacement measurement in the single direction of each stage of the measurement model designed here is controlled within 12%, which can meet the requirements of displacement measurement of spatial structure. The purpose of this work is to provide essential technical support for the upgrading of displacement measurement technology of large-span spatial structure of buildings and the improvement of measurement accuracy.

Keywords: binocular stereo vision, convolutional neural network, sphere joint, YOLOv2, long-span spatial structure, displacement measurement, image architecture

1. INTRODUCTION

The spatial structure developed rapidly in recent years, and a proliferation of various long-span spatial structures are put into practical application scenarios. However, the complicated long-span spatial structure has problems including material aging and foundation settlement under the influence of environmental loads. Such consequential problems cannot be ignored because they will render the long-span spatial structure broken-down or

even damage it [1, 2]. Hence, the displacement measurement of large-span structures plays an influential role in structure protection [3, 4]. Because of its excellent characteristics, machine vision has been widely accepted in engineering practice and other fields [5]. As a representative artificial intelligence technology, deep learning has achieved a breakthrough in the intelligent classification and recognition of images, which has gradually developed into an effective displacement measurement method for large-scale spatial structures [6, 7]. According to the implementation principle of machine vision, deep learning can effectively match and intelligently recognize the images combined with the actual characteristics of the large-scale spatial structure measurement process, thereby providing solid support for the accurate measurement of large-scale spatial structure displacement.

Regarding the structure displacement measurement, Ge *et al.* (2020) proposed a real-time monitoring method of box girder bridge overturning risk with predictive ability. This method combined machine vision with dynamic weighing monitoring information to recognize the distribution of real vehicles queuing in the designated area of bridge deck in real time. The authors derived the analytical expression of the support reaction of the main girder, and realized the real-time calculation of the support reaction [8]. Kromanis and Kripakaran (2021) proposed a multi-camera positioning method for computer vision accurate displacement measurement. They adopted coordinated transformation techniques, and utilized identified features to calculate the structural displacement from images. Then, they employed the proposed method to a laboratory where multiple cameras were used for recording in a series of locations. Their experimental results showed that the response computed from the recordings were accurate [9]. Wadhwa and Mukherjee (2021) held that, in the case of a given frame sequence and large inter-frame optical flow vector displacement, it was difficult to predict the next few frames of video motion. Moreover, conventional Convolutional Neural Networks (CNNs) often failed to learn the dynamics of the objects across frames in case of large displacements of objects in consecutive frames. Therefore, the authors proposed a CNN algorithm based on the concept of feature pyramid to extract spatial features from multiple consecutive images. Besides, they utilized the improved PWC-Net framework to extract spatial features from continuous frames and input them into bidirectional Long Short-Term Memory network to obtain temporal features. They found that the proposed pyramid could capture the sudden motion of moving objects in video, especially when the displacement of objects was large between consecutive frames. In addition, the proposed spatial-temporal pyramid feature could effectively predict the light flow in the next few frames, rather than only the next frame. Ultimately, through experiments, they proved that the proposed method was better than other technologies available in the application of challenging data sets [10].

In displacement measurement for large-scale spatial structures, machine vision has some disadvantages, such as inaccurate labeling, inaccurate target matching, and extremely complicated image processing. Based on the above problems, machine vision measurement technology is integrated with a deep learning method here to propose a precise displacement measurement model of large-span spatial structure of buildings based on machine vision technology and deep learning. Firstly, binocular stereo vision technology based on machine vision and CNN are described. Secondly, the displacement measurement model of large-span spatial structure based on deep learning is constructed. Finally, the model reported here is tested by constructing data sets through using various spatial structure images online and shot by smartphones. It is expected to offer technical support to the

upgrading of displacement measurement technology of large-span spatial structure of buildings and the improvement of measurement accuracy.

2. METHOD

2.1 Image Processing and Selection of Deep Learning Model Under Machine Vision

Machine vision is an excessively effective optical measurement technique [11, 12], and the binocular stereo vision system can measure the spatial displacement of the target objects [13]. A binocular stereo vision system for measurement contains spatial three-dimensional (3D) coordinate reconstruction, camera calibration, and image processing [14, 15]. The overall model implementation process is shown in Fig. 1. Different from the monocular vision system model [16], the binocular stereo vision model adds a camera on the right in Fig. 1. With the cooperation of the two cameras, more effective image information can be collected, thereby obtaining the spatial position of the target to be measured.

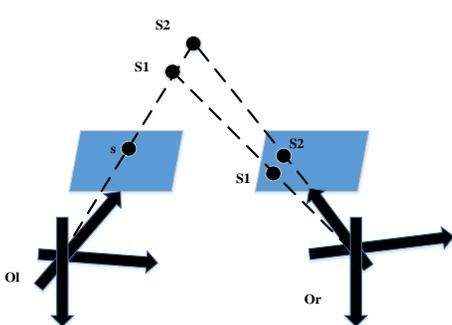


Fig. 1. Binocular stereo vision model.

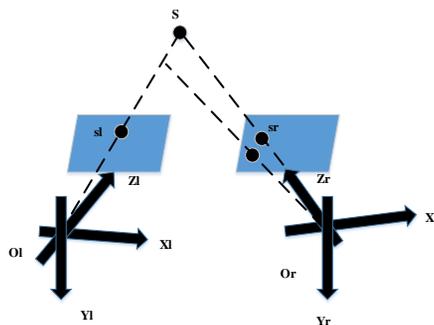


Fig. 2. 3D reconstruction model of the target point.

Camera calibration can properly set the internal and external parameters and distortion parameters [17, 18] to construct and correct the imaging model. Image matching is a crucial step in image processing is, that is, obtaining the required information from different digital images. Currently, gray-scale matching and feature matching are two common matching methods. The former relies on the gray-scale information of image pixels, and the latter is achieved by image preprocessing, feature extraction, and the establishment of mapping relationships. Specific applications require considering the actual situations [19, 20]. Fig. 2 reveals the model based on the 3D reconstruction of the target point.

Since the binocular measurement technology can quickly obtain the overall geometric spatial information of objectives to be measured at a distance, it can automatically identify and analyze the measurement information and achieve outstanding accuracy for measuring displacement of the large-span spatial structure of buildings. However, in actual scenarios, the requirements for camera calibration are extremely high, and there are problems such as difficulty in marking spatial positions and large errors. Therefore, it is necessary to seek an intelligent image recognition and processing method.

Compared with traditional machine learning approaches, deep learning pays more attention to the depth of the model; therefore, it can even extract effective information from

a large number of samples [21, 22]. CNN is a typical deep learning network, providing excellent performance for image processing and recognition [23]. Hence, CNN is introduced to process images of large-span spatial structure. The convolution part of CNN reflects its characteristics, which is composed of the convolution layer, pooling layer, and activation function [24].

The classification of images in computer vision aims to determine the type of image, and the purpose of target detection is to identify and locate related objects. The classification of images is the premise. The region-based CNN (RCNN) was proposed for image classification in 2014 [25]. The detection accuracy of this method under the PASCAL target database can reach 66%, but it is time-consuming and computationally intensive. The Spatial Pyramid Pooling Network (SPPNet) proposed later can achieve the classification and regression of different features; nevertheless, it has different mapping feature dimensions [26]. The Faster RCNN developed later solves the aforementioned problems and improves the detection efficiency significantly [27]. On this basis, the You Only Look Once (YOLO) algorithm transforms target detection into a regression problem, which significantly improves the detection speed. The YOLOv2 algorithm realizes higher accuracy and recall rate, so it is used for processing high-resolution images, which is even better than the Faster RCNN model [28, 29]. Hence, YOLOv2 is introduced into the sphere joint recognition for the large-span spatial structures; its corresponding size is 544*544.

2.2 Displacement Measurement of Large-Span Spatial Structure Based on Deep Learning

The identification of sphere joint structure is the foundation of displacement measurement of long-span spatial structures. The reason why the sphere joint is chosen is that it has obvious characteristics and fixed geometric characteristics in long-span buildings. In the process of image matching, accurate positioning of observation points is helpful to image processing. No matter what the direction is, the two-dimensional projection of the corresponding sphere joint is a circle. The sphere projected on the two-dimensional image is located in the center of the circle, which is conducive to image recognition and coordinate positioning. Fig. 3 illustrates the implementation process of the displacement measurement system of large-span spatial structures based on binocular vision technology and deep learning method.

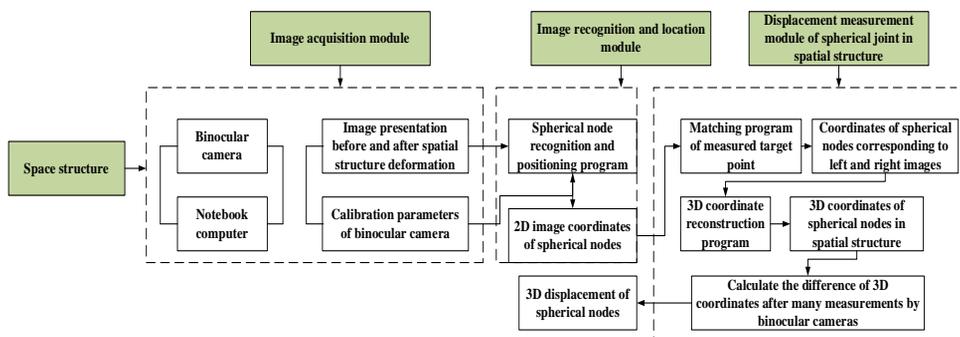


Fig. 3. Implementation process of long-span spatial structure displacement measurement system.

The displacement measurement system consists of three modules, *i.e.*, image acquisition, recognition and positioning, and displacement measurement. The image acquisition module includes the binocular stereo vision system that aims to acquire the regional images before the spatial structure is deformed. The recognition and positioning module take the detected image information as input and employs CNN and YOLOv2 to recognize and detect the images to obtain the 2D image coordinates. The displacement measurement module is the core of the system. The coordinate information of the target point obtained by the above output is used as the input, and the 3D reconstruction of the coordinate can be formed after matching. On this basis, the 3D displacement of the target point before and after the deformation of the space structure can be completed.

All sphere joints in the overall structure are emphasized in the spatial structure displacement measurement herein. The 3D reconstruction and displacement calculation are the crux of the displacement measurement system. Because the camera calibration parameters and imaging position are all in a non-ideal state in the actual scenarios, there will be errors in the measurement process. Consequently, the solution to the coordinates is transformed into an optimization problem. If the optical center coordinates of the left camera are $O_l(X_{O_l}, Y_{O_l}, Z_{O_l})$ and the optical center coordinates of the right camera are $O_r(X_{O_r}, Y_{O_r}, Z_{O_r})$, the spatial point $S(X, Y, Z)$ is projected on the imaging plane of the two cameras as $S_l(X_l, Y_l, Z_l)$ and $S_r(X_r, Y_r, Z_r)$, then the projection vector of the imaging plane corresponding to the optical center coordinates of the left camera $O_l\vec{S}_l$ can be described as:

$$O_l\vec{S}_l = \begin{pmatrix} X_l - X_{O_l} \\ Y_l - Y_{O_l} \\ Z_l - Z_{O_l} \end{pmatrix}. \quad (1)$$

The projection vector of the left camera imaging plane corresponding to the space point O_lS can be described as:

$$O_l\vec{S} = \begin{pmatrix} X - X_{O_l} \\ Y - Y_{O_l} \\ Z - Z_{O_l} \end{pmatrix}. \quad (2)$$

The unit vector of $O_l\vec{S}_l$ can be defined as:

$$\vec{V}_l = (a_l, b_l, c_l). \quad (3)$$

In Eq. (3), a_l , b_l , and c_l are unit vectors on three directions of the X-axis, Y-axis, and Z-axis.

Then, the length of the vector $O_l\vec{L}$ can be presented as:

$$\|O_l\vec{L}\| = O_l\vec{S} \cdot \vec{V} = a_l(X - X_{O_l}) + b_l(Y - Y_{O_l}) + c_l(Z - Z_{O_l}). \quad (4)$$

Thus, the distances from point S to the left and right projection lines can be expressed as:

$$\|\vec{SL}\|^2 = \|\vec{O_rS}\|^2 - \|\vec{O_lS}\|^2 = (X - X_{O_l})^2 + (Y - Y_{O_l})^2 + (Z - Z_{O_l})^2 - [a_l(X - X_{O_l}) + b_l(Y - Y_{O_l}) + c_l(Z - Z_{O_l})]^2. \quad (5)$$

$$\|\vec{SR}\|^2 = (X - X_{O_r})^2 + (Y - Y_{O_r})^2 + (Z - Z_{O_r})^2 - [a_r(X - X_{O_r}) + b_r(Y - Y_{O_r}) + c_r(Z - Z_{O_r})]^2. \quad (6)$$

Therefore, the objective function T of the optimization problem can be expressed as:

$$T = \|\vec{SL}\|^2 + \|\vec{SR}\|^2. \quad (7)$$

Eq. (8) can be derived by sorting out the above findings:

$$\begin{bmatrix} (1-a_l^2)+(1-a_r^2) & -a_l b_l - a_l b_r & -a_l c_l - a_l c_r \\ -a_l b_l - a_l b_r & (1-b_l^2)+(1-b_r^2) & -b_l c_l - b_l c_r \\ -a_l c_l - a_l c_r & -b_l c_l - b_l c_r & (1-c_l^2)+(1-c_r^2) \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (1-a_l^2)X_{O_l} + (1-a_r^2)X_{O_r} - a_l b_l Y_{O_l} - a_l b_r Y_{O_r} - a_l c_l Z_{O_l} - a_l c_r Z_{O_r} \\ -a_l b_l X_{O_l} - a_l b_r X_{O_r} + (1-b_l^2)Y_{O_l} + (1-b_r^2)Y_{O_r} - b_l c_l Z_{O_l} - b_l c_r Z_{O_r} \\ -a_l c_l X_{O_l} - a_l c_r X_{O_r} - b_l c_l Y_{O_l} - b_l c_r Y_{O_r} + (1-c_l^2)Z_{O_l} + (1-c_r^2)Z_{O_r} \end{bmatrix} \quad (8)$$

The 3D coordinates of point S (X , Y , Z) can be obtained through linear solving, and all the nodes can be solved to obtain the spatial 3D coordinates of all sphere joints in the world coordinate system. The 3D spatial coordinates of all target points on the image can be calculated using the MATLAB for programming. The coordinate difference obtained by two measurements is the 3D displacement of the sphere joint. Fig. 4 displays the implementation process of the displacement measurement module.

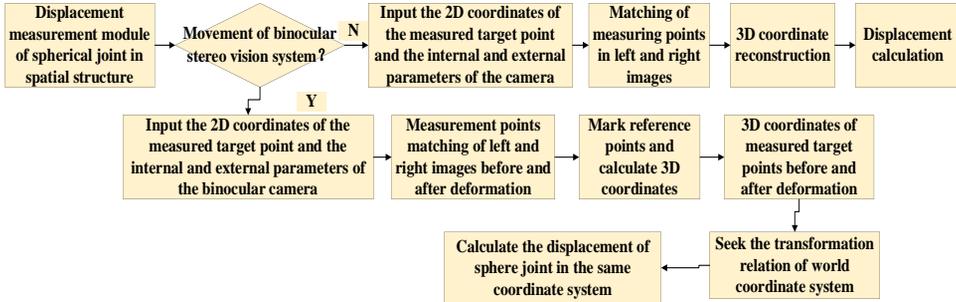


Fig. 4. Implementation process of the displacement measurement module.

3. RESEARCH MODEL AND RESEARCH PROCESS

3.1 Acquisition of Experimental Materials and Construction of Test Data Set

The data set used in this experiment is all kinds of spatial structure images containing sphere joints, of which 80% are collected from the network. The images with high resolution and clear sphere joints imaging are selected to form the data set. The remaining 20% are captured in the structural health monitoring by a smartphone. Most of the images in the data set have high resolution, but due to the large scale of spatial structure and the large number of nodes in the image, the nodes in the distance are difficult to identify in the image, which cannot directly meet the requirements of training data. Therefore, the collected

images are cut and segmented, and the part with clear nodes in the image is screened out to find out as many different types of structures and nodes as possible, and form a data sample set. In addition, some images are rotated by 90° or 180° to increase the number and diversity of samples.

After the collection of data set, it is necessary to record the position of the target node in the image and save it into the txt. document. The images and the txt. document are used as the input of the model training, as illustrated in Fig. 5.

The paramount implementation procedure is to obtain the center coordinates (A, B) of the image pixel by calculating the bounding box of the target node, the length A, and width B of the corresponding bounding box. The more accurate the position information labelled in the training image is, the stronger the recognition and positioning ability of weight files of the model will obtain. The former represents the coordinates of the center of the sphere in a specific image, and the latter denotes the size of the sphere joint on a specific image. The output data is normalized according to different image sizes. Finally, the editing process of txt. document is completed by Labeling software and programming software. Labeling software calculates the pixel coordinates and the pixel size of length and width of the boundary on the image by manually selecting the bounding box on the image. Then, matlab programming reads the image resolution information and processes the information output by Labeling to obtain the positioning txt. document required by the data set.

After the above processing, the final data sample set contains a total of 90 images. In addition, there are 90 existing documents and 800 sphere joints in total.

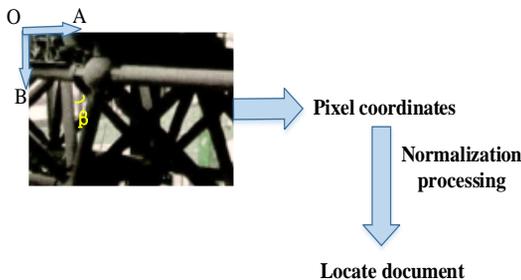


Fig. 5. Implementation of image target positioning.



Fig. 6. Spatial grid structure for performance verification.

3.2 Experimental Design

The spatial grid structure presented in Fig. 6 below is selected to test the effectiveness of the displacement measurement system reported here.

First, the vision measurement system is deployed. The two selected cameras are connected to the computer and fixed on the tripod crossbar. The angle of the camera is adjusted according to the camera imaging displayed on the laptop to present the measurement target at the center of the visual field. Based on the world coordinate system and the displacement measurement system, the X-axis direction is parallel to the line connecting the optical centers of the two cameras, the Y-axis direction is parallel to the lateral load direction, and the Z-axis is opposite to the vertical load direction. Before the loading starts, two dial gauges are placed next to the sphere joint to successfully measure the displacement of the sphere

joint in the Y-axis and Z-axis directions. After the displacement measurement system is deployed, the calibration ruler is adopted to measure the multiple fixed distances in the dual-camera image to calibrate the cameras.

The test will begin after the layout of the displacement measurement system and the calibration of the camera are completed. Specifically, the implementation process is as follows: (1) recording the reading of the dial indicator before the loading starts and using it as the basis to collect the structure state of the first stage; (2) unloading the corresponding structure after applying the load in the Y direction and recording the corresponding structural state; (3) unloading the corresponding structure after applying the load in the Z-axis direction. The whole process goes through three periods. The image information is collected by the binocular stereo vision system.

3.3 Configuration of Experimental Hardware and Parameter Setting

The measurement objects and hardware configuration of this experimental design are summarized in Table 1 below.

The detection target is a single class target of sphere joints of spatial structures, and the input image size is not certain. Therefore, the program code is modified in accordance with the computer hardware information, so that the program detection target and the training image scale meet the requirements. Then the training parameters are adjusted by pre-training, and the final training parameters are given by the parameter convergence process as follows: $batch = 64$; $subdivisions = 8$; $max\ batch = 30000$; $momentum = 0.9$; $decay = 0.0005$. Besides, to accelerate convergence in the early stage and refine the later training, the learning rate is set to 0.001 in 0-100 batches, and to 0.0001 in 100-10,000 batches, and 0.00001 later. Moreover, the default threshold of node detection is increased from 0.25 to 0.6 to reduce the false recognition and enhance the detection accuracy. The above parameters are used in the subsequent training of different CNN models and the sphere joint localization model.

Table 1. Measurement objects and hardware configuration of experimental design.

Parameters		Detailed settings				
Measurement objects	Double layer spatial grid structure	Spatial scale		2.5*0.5*1.0m		
		Definition		2464*2056		
Camera	Manta G-507B industrial monochromatic camera	Focal length		16mm		
		Effective focal length	Pixel size	Scale factor	Position matrix	
Calibration parameters	Left camera	15.8262mm	0.00344mm	4587.26	$R_l = \begin{pmatrix} 0.975 & 0 & -0.207 \\ 0 & 1 & 0 \\ 0.207 & 0 & 0.975 \end{pmatrix} T_l = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	
	Right camera	15.7968mm	0.00344mm	4578.85	$R_r = \begin{pmatrix} 0.988 & 0.001 & 0.148 \\ 0.002 & 0.998 & -0.025 \\ -0.148 & 0.025 & 0.988 \end{pmatrix} T_r = \begin{pmatrix} -763.792 \\ -1.972 \\ 28.898 \end{pmatrix}$	

3.4 Performance Analysis of the Displacement Measurement System

CNNs of diversified structures has formed in the development process, and the depth of the corresponding neural network model has increased accordingly. Here, Recall and Intersection Over Union (IOU) are used as evaluation indicators to test the optimal model for node recognition in large-span spatial structures of buildings. Three CNN network models, namely AlexNet [30], Darknet19 [31], and Resnet152 [32] are included for comparative analysis.

The experiment tests the effect of the optical axis angle on the measurement error, and verifies the effectiveness of binocular stereo vision system in the displacement measurement system. Besides, the displacement measurement system is analyzed based on the actual displacement measurement results.

4. RESULTS AND DISCUSSION

4.1 Recognition Ability of CNN Model

Fig. 7 provides the recognition results of the three comparative models after 5k times of training and 10k times of training.

As shown in Fig. 7, after 5k cycles of training, the Recall of Darknet19 reaches 92.5%, and the IOU value attains 81.2%, which are better than those of AlexNet and Resnet152. After 10k cycles of training, Recall of Darknet19 is 93.9%, and IOU reaches 82.02%, which are also superior to those of the other two models. Therefore, the Darknet19 model is selected for the subsequent test of the YOLOv2 algorithm.

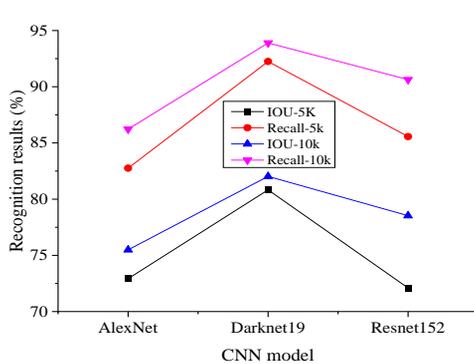


Fig. 7. Comparison of recognition results of three comparative models.

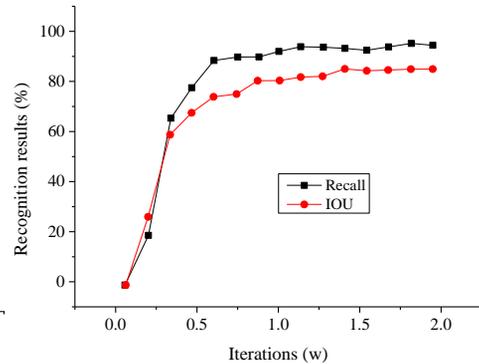


Fig. 8. Recognition result of the YOLOv2 algorithm.

Fig. 8 signifies recognition result of the YOLOv2 algorithm after training the Darknet19 model.

The curves in Fig. 8 suggest that when the Darknet19 model is trained about 20,000 times, Recall of the YOLOv2 recognition algorithm is 94.69%, and the IOU value is 84.98%. Hence, the YOLOv2 recognition algorithm achieves an excellent performance in recognizing and positioning the long-span spatial structure of buildings, which also reflects the outstanding applicability of the algorithm.

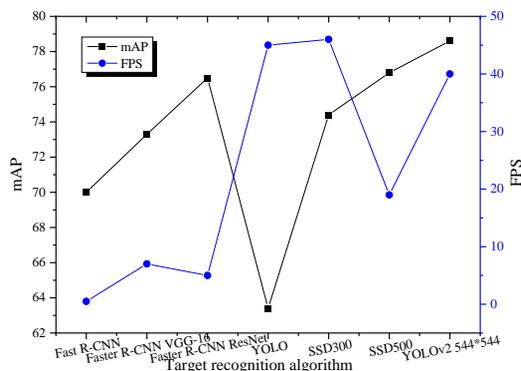


Fig. 9. Comparison of different target recognition and detection algorithms (SSD stands for Single Shot MultiBox Detector; mAP stands for mean Average Precision, which measures the detection effect of multiple classes; FPS stands for Frame Per Second, which denotes the detection rate.)

To validate the efficiency of the YOLOv2 algorithm, several common target recognition and detection frameworks are introduced for comparison, and the results are shown in Fig. 9.

Results obtained based on mAP and FPS are compared, suggesting that the YOLOv2 recognition algorithm reported here can provide the best detection accuracy and detection rate. This further confirms the results of the research. According to the evaluation results in Fig. 9, the Recall indicator is employed to evaluate the recognition performance of different algorithms. The final results demonstrate that the YOLOv2 algorithm realizes a strong recognition ability, and there are few omissions in the application scenario. The IOU indicator can evaluate the positioning performance of the algorithm, and the final result indicates that the YOLOv2 algorithm has good positioning ability. Srivastava and Srivastava (2020) used YOLOv2 as the bounding box to study the detection and localization of image targets. Based on the results obtained from the F measurement, they found that the model could provide good performance in image positioning detection, and the F measurement was above 0.7 [33]. Regarding the YOLOv2 applications in target detection and positioning, Shi *et al.* (2021) proposed a multi-target recognition and classification method based on YOLOv2 for vehicle classification and target detection. A comparison with existing models found that the accuracy of the model based on YOLOv2 could reach 98.6%, the misjudgment rate was low, and the robust performance was good [34]. These works further confirms the feasibility of the YOLOv2 algorithm and provides good support for the research in the present work.

4.2 Measurement Error of Binocular Stereo Vision System

Based on the above system layout, the results of the influence of the optical axis angle on the measurement error are shown in Fig. 10.

Fig. 10 demonstrates that the influence of the optical axis angle on the system measurement is affected by the changes in the θ value. When the optical axis angle is about 40° , and the θ value is in the range of 20° to 60° , the system's measurement error is at a comparably low level. The optical axis angle between the optical axes represents the distribution of the measurement area in the camera position arrangement of the binocular stereo

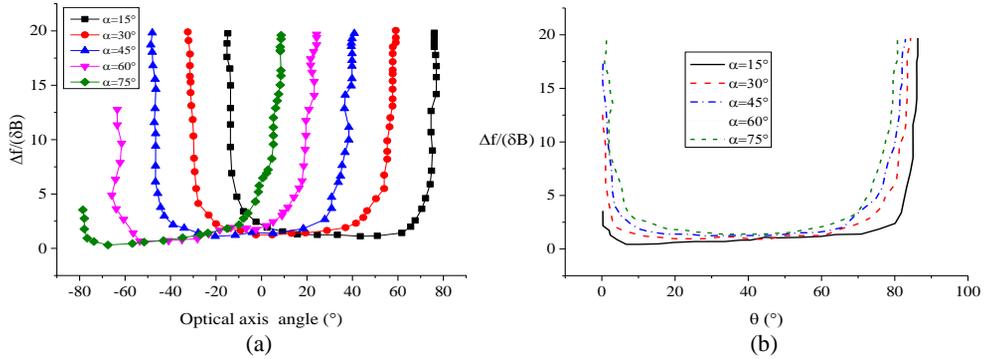


Fig. 10. The measurement error of the binocular stereo vision system; (a) The influence of the optical axis angle; (b) The influence of the θ value.

vision system. The above analysis of measurement errors reveals that when arranging the binocular stereo vision system, the measurement area should be kept within $\pm 20^\circ$ of the optical axis angle between the left and right cameras' optical axes. In this way, the measurement error of the system can be minimized to provide excellent prerequisite support for the subsequent displacement measurement. The above analysis suggests that based on the binocular stereo vision system, the measurement error can be reduced to lay a good foundation for the subsequent accurate displacement measurement.

4.3 Test Verification Results of the Displacement Measurement System

Table 2 displays the 2D pixel coordinate distribution corresponding to each time point in the test.

Table 2. The 2D pixel coordinates of the spatial structure sphere joint at different time points.

Numbering of sphere joints of the spatial structure	Time points		2D pixel coordinates			
			Left camera		Right camera	
			U	V	U	V
1	Before loading		260.32	493.76	270.56	648.97
	Lateral load	After loading	259.61	492.66	272.02	647.92
		After of-floading	258.02	493.36	271.52	649.04
	Vertical load	After loading	257.32	487.66	272.26	643.66
		After of-floading	257.65	493.69	272.19	648.77
2	Before loading		1693.11	541.92	1617.44	649.59
	Lateral load	After loading	1692.03	538.55	1614.22	647.37
		After of-floading	1692.93	538.52	1614.48	649.04
	Vertical load	After loading	1692.57	535.26	1614.17	645.47
		After of-floading	1693.25	539.08	1613.88	648.98

Table 2. (Cont'd) The 2D pixel coordinates of the spatial structure sphere joint at different time points.

3	Before loading		901.48	1335.86	1069.65	1445.48
	Lateral load	After loading	901.28	1335.57	1069.95	1444.46
		After of-flooding	904.74	1334.65	1069.47	1443.98
	Vertical load	After loading	903.98	1333.23	1069.44	1442.22
		After of-flooding	904.95	1335.97	1069.79	1443.96
4	Before loading		2091.85	1344.03	2299.12	1497.56
	Lateral load	After loading	2090.48	1344.58	2303.36	1495.88
		After of-flooding	2089.98	1341.87	2303.37	1496.19
	Vertical load	After loading	2089.37	1341.24	2303.32	1495.26
		After of-flooding	2090,05	1344.98	2304.57	1495.88

Table 3. Reconstruction results of space coordinates of sphere joints at different time points.

Numbering of sphere joints of the spatial structure	Time points		Spatial reconstruction coordinates of sphere joints /mm		
			X	Y	Z
1	Before loading		9.57	2062.29	-179.82
	Lateral load	After loading	9.36	2063.95	-179.55
		After of-flooding	339.82	2365.82	222.97
	Vertical load	After loading	8.33	2067.68	-181.86
		After of-flooding	8.44	2067.42	-179.32
3	Before loading		338.89	2365.78	223.92
	Lateral load	After loading	338.95	2368.73	223.18
		After of-flooding	339.82	2365.82	222.98
	Vertical load	After loading	339.57	2366.68	222.98
		After of-flooding	339.57	2366.38	222.48

Table 3 shows the reconstruction results of the space coordinates of the sphere joints corresponding to each time point in the experiment.

According to the results in Table 3, the 3D displacement at different time points can be obtained. Then, the error based on the machine vision measurement method can be finally obtained through the measurement results of the displacements at each time point obtained by the dial indicator, as presented in Fig. 11.

Through Fig. 11, at each time point, the absolute error of the measurement results of corresponding unidirectional displacement during the test process is 0.12mm on average, and the maximum value of the overall measurement error is 0.31mm. Besides, the average value of the relative error during the measurement is 8.2%, and the maximum value of the relative error is 11.8%.

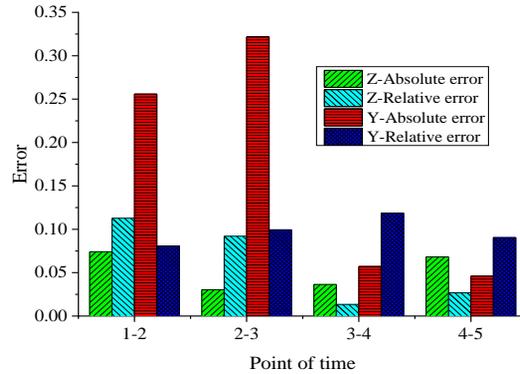


Fig. 11. Displacement error measurement results.

4.4 Experiment Results and Discussion

In the test results of recognition effects of three CNN models, namely AlexNet, Darknet19, and Resnet152, after 5k times of training and 10k times of training, Recall of the Darknet19 model reaches 93.9%, which is at least 3.6% higher than the other two models. Besides, the maximal IOU is 82.02%, which is at least 4.4% higher than of other models. Moreover, the Darknet19 model is superior to the other two CNN models in identifying and locating the sphere joints of spatial structures. This is because the Darknet19 model is portable and easy to install, and it can display images, which can be used as the backbone network for feature extraction. Based on the actual characteristics of long-span building structures, the Darknet19 model is better for the identification and location of sphere joints. In the identification test results of YOLOv2 algorithm, when the Darknet19 model is trained about 20,000 times, the corresponding Recall of YOLOv2 identification algorithm attains 94.69 %, and the IOU value reaches 84.98 %. Meanwhile, its average recognition accuracy is at least 2.3% higher than that of other recognition algorithms, showing excellent recognition precision and adaptability.

The precise displacement measurement model of large-span spatial structure of buildings designed here primarily depends on the binocular stereo vision system. When the optical axis angle of the binocular stereo vision system is within the range of $\pm 20^\circ$ and the θ value is within the range of $20^\circ - 60^\circ$, the measurement error is relatively low. The relative error of measurement in the whole experiment is controlled within 12%, which can meet the requirements of displacement measurement of spatial structures, but the accuracy needs to be improved. One of the reasons for the displacement measurement error of large-span spatial structure of buildings is the IOU value between the actual bounding box corresponding to the positioning sphere joint and the predicted bounding box. In addition, the angle between the left and right cameras is relatively small, making the displacement error outside the plane slightly larger than that in the plane.

5. CONCLUSION

By means of the binocular stereo vision measurement technology of machine vision, a CNN-based sphere joint recognition algorithm is proposed by applying the deep learning-

based YOLOv2 recognition algorithm. Finally, a framework of long-span spatial structure displacement system is constructed. According to the recognition results of different models and the measurement results of the displacement measurement system, recall of the Darknet19 model reaches up to 93.9%, the IOU value is 82.02%. The two indicators are at least 3.6% and 4.4% higher than those of the other two models. Moreover, the Darknet19 model performs better in the task of identifying and locating sphere joints in spatial structures. When the Darknet19 model is trained about 20,000 times, Recall of the YOLOv2 recognition algorithm attains 94.69%, and the IOU value is 84.98%. Besides, the average recognition accuracy of YOLOv2 algorithm is 2.3% higher than other recognition algorithms, showing a brilliant recognition effect. The relative error of the measurement model designed here in the single direction displacement measurement of each stage is controlled within 12%, which can meet the requirements of displacement measurement of spatial structures. The deficiency of this work is that the measurement object in the experiment is relatively single, and it is essential to enhance the measurement accuracy and further study the practical application in the field. The research outcome can provide critical reference for the upgrading of displacement measurement technology of large-span spatial structure of buildings and the improvement of measurement accuracy.

REFERENCES

1. M. Balázsová, M. Feistauer, J. Horáček, M. Hadrava, and A. Kosík, "Space-time discontinuous Galerkin method for the solution of fluid-structure interaction," *Applications of Mathematics*, Vol. 63, 2018, pp. 739-764.
2. T. Shen, B. Shirinzadeh, Y. Zhong, J. Smith, J. Pinskiar, and M. Ghafarian, "Sensing and modelling mechanical response in large deformation indentation of adherent cell using atomic force microscopy," *Sensors*, Vol. 20, 2020, p. 1764.
3. P. Robalinho and F. Orlando, "Nano-displacement measurement using an optical drop-shaped structure," *IEEE Photonics Technology Letters*, Vol. 33, 2020, pp. 65-68.
4. B. Shang, Y. Miao, H. Zhang, C. Fei, and L. Zu, "Ultralong-period microfiber grating for simultaneous measurement of displacement and temperature," *IEEE Photonics Technology Letters*, Vol. 31, 2019, pp. 1763-1766.
5. H. Wang, J. Zhao, J. Zhao, J. Song, Z. Pan, and X. Jiang, "A new rapid-precision position measurement method for a linear motor mover based on a 1-D EPCA," *IEEE Transactions on Industrial Electronics*, Vol. 65, 2017, pp. 7485-7494.
6. P. S. Tamizharasan and N. Ramasubramanian, "Analysis of large deviations behavior of multi-GPU memory access in deep learning," *The Journal of Supercomputing*, Vol. 74, 2018, pp. 2199-2212.
7. J. Jaworek-Korjakowska, "A deep learning approach to vascular structure segmentation in dermoscopy colour images," *BioMed Research International*, Vol. 2018, 2018, pp. 1-8.
8. L. Ge, D. Dan, X. Yan, and K. Zhang, "Real time monitoring and evaluation of overturning risk of single-column-pier box-girder bridges based on identification of spatial distribution of moving loads," *Engineering Structures*, Vol. 210, 2020, pp. 110383.
9. R. Kromanis, and P. Kripakaran, "A multiple camera position approach for accurate displacement measurement using computer vision," *Journal of Civil Structural Health Monitoring*, Vol. 11, 2021, pp. 661-678.

10. L. Wadhwa and S. Mukherjee, "Learnable spatiotemporal feature pyramid for prediction of future optical flow in videos," *Machine Vision and Applications*, Vol. 32, 2021, pp. 1-10.
11. Y. Chen, C. Hou, Y. Tang, J. Zhuang, J. Lin, Y. He, and S. Luo, "Citrus tree segmentation from UAV images based on monocular machine vision in a natural orchard environment," *Sensors*, Vol. 19, 2019, p. 5558.
12. K. Wurtz, I. Camerlink, R. B. D'Eath, A. P. Fernández, T. Norton, J. Steibel, and J. Siegford, "Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review," *PLoS One*, Vol. 14, 2019, p. e0226669.
13. L. Yang, B. Wang, R. Zhang, H. Zhou, and R. Wang, "Analysis on location accuracy for the binocular stereo vision system," *IEEE Photonics Journal*, Vol. 10, 2017, pp. 1-16.
14. T. Wang, and A. K. Somani, "Aerial-DEM geolocalization for GPS-denied UAS navigation," *Machine Vision and Applications*, Vol. 31, 2020, pp. 1-12.
15. W. Wang, H. Song, Z. Zhang, and H. Cui, "Landslide multi-attitude data measurement of bedding rock slope model," *International Journal of Parallel Programming*, Vol. 48, 2020, pp. 928-939.
16. D. C. Andrade, F. Bueno, F. R. Franco, R. A. Silva, J. H. Z. Neme, E. Margraf, and R. dos Santos Amaral, "A novel strategy for road lane detection and tracking based on a vehicle's forward monocular camera," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, 2018, pp. 1497-1507.
17. J. Jiang, C. Mi, M. Wu, Z. Zhang, Y. Feng, and W. Mi, "Real-time container truck speed measurement at container port gates based on the binocular vision technology," *Journal of Coastal Research*, Vol. 93, 2019, pp. 998-1005.
18. Q. Hu, Z. Feng, L. He, Z. Shou, J. Zeng, J. Tan, and Y. Gu, "Accuracy improvement of binocular vision measurement system for slope deformation monitoring," *Sensors*, Vol. 20, 2020, pp. 1994.
19. J. Liu, Z. Zhou, W. Xu, and J. Hu, "Adaptive support-weight stereo-matching approach with two disparity refinement steps," *IETE Journal of Research*, Vol. 65, 2019, pp. 310-319.
20. B. J. M. Hess, "On the role of ocular torsion in binocular visual matching," *Scientific Reports*, Vol. 8, 2018, pp. 1-12.
21. A. Pla, X. Zhong, and S. Rayner, "miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts," *PLoS Computational Biology*, Vol. 14, 2018, p. e1006185.
22. A. Moshovos, J. Albericio, P. Judd, A. D. Lascorz, S. Sharify, T. Hetherington, and N. E. Jerger, "Value-based deep-learning acceleration," *IEEE Micro*, Vol. 38, 2018, pp. 41-55.
23. H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, and I. Zalaudek, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, Vol. 29, 2018, pp. 1836-1842.
24. F. C. Chen and M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Transactions on Industrial Electronics*, Vol. 65, 2017, pp. 4392-4400.

25. X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, Vol. 299, 2018, pp. 42-50.
26. N. Wang, B. Li, Q. Xu, and Y. Wang, "Automatic ship detection in optical remote sensing images based on anomaly detection and SPP-PCANet," *Remote Sensing*, Vol. 11, 2019, p. 47.
27. N. Mo and L. Yan, "Improved faster RCNN based on feature amplification and over-sampling data augmentation for oriented vehicle detection in aerial images," *Remote Sensing*, Vol. 12, 2020, p. 2558.
28. J. Wang, N. Wang, L. Li, and Z. Ren, "Real-time behavior detection and judgment of egg breeders based on YOLO v3," *Neural Computing and Applications*, Vol. 32, 2020, pp. 5471-5481.
29. Z. Wu, J. Sang, Q. Zhang, H. Xiang, B. Cai, and X. Xia, "Multi-scale vehicle detection for foreground-background class imbalance with improved YOLOv2," *Sensors*, Vol. 19, 2019, p. 3336.
30. T. Shanthi and R. S. Sabeenian, "Modified Alexnet architecture for classification of diabetic retinopathy images," *Computers & Electrical Engineering*, Vol. 76, 2019, pp. 56-64.
31. R. Broadhurst, M. Ball, and C. J. Jiang, "Availability of COVID-19 related products on Tor darknet markets," *Australasian Policing*, Vol. 12, 2020, pp. 8-13.
32. X. Du, J. Yuan, L. Hu, and Y. Dai, "Description generation of open-domain videos incorporating multimodal features and bidirectional encoder," *The Visual Computer*, Vol. 35, 2019, pp. 1703-1712.
33. G. Srivastava and R. Srivastava, "User-interactive salient object detection using YOLOv2, lazy snapping, and gabor filters," *Machine Vision and Applications*, Vol. 31, 2020, pp. 1-7.
34. B. Shi, X. Li, T. Nie, K. Zhang, and W. Wang, "Multi-object recognition method based on improved YOLOv2 model," *Information Technology and Control*, Vol. 50, 2021, pp. 13-27.



Yinghao Kong was born in Jining, Shandong Province in 2000. He received his bachelor's degree from Shandong University of Science and Technology. His research interests include spatial structure and parametric design.