Journal of

Applied and Computational Mechanics

Shahid Chamran
University of Ahvaz

Research Paper

A Bridge Vibration Measurement Method by UAVs based on CNNs and Bayesian Optimization

Gongfa Chen[1], Zhaocheng Yan[1], Shuai Teng[1], Fangsen Cui[2], David Bassir[3,4]

[1] School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou, 510006, China
[2] Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore 138632, Republic of Singapore
[3] UTBM, IRAMAT UMR 7065-CNRS, Rue de Leupe, 90010 Belfort Cedex, France
[4] Centre Borelli, ENS - Université Paris-Saclay, 4 avenue des Sciences, 91190 Gif-sur-Yvette, France

Abstract. A bridge vibration measurement method by Unmanned Aerial Vehicles (UAVs) based on a Convolutional Neural Network (CNN) and Bayesian Optimization (BO) is proposed. In the proposed method, the video of the bridge structure is collected by a UAV, then the reference points in the background of the bridge and the target points on the bridge in the video are tracked by the Kanade-Lucas-Tomasi (KLT) optical flow method, so that their coordinates can be obtained. The BO is used to find the optimal hyper-parameter combination of a CNN, and the CNN based on BO is used to correct the bridge displacement signal collected by the UAV. Finally, the natural frequency of the bridge is extracted by processing the corrected displacement signals with Operational Modal Analysis (OMA). Moreover, a steel truss is used as the experimental model. The number of reference points and the shooting time of the UAV with the optimal correction effect of the BO-based CNN are obtained by two groups of comparative experiments, and the influence of the distance between structure and reference points on the correction effect of the BO-based CNN is determined by another group of comparative experiment. The static reference points are not required for the proposed method, which evidently enhances the applicability of UAVs; the conclusion of this paper has great guiding significance for the actual bridge vibration measurement.

Keywords: Bridge vibration, Unmanned aerial vehicles, Correction effect, Convolutional neural networks, Bayesian optimization.

1. Introduction

Damage identification of bridge structures is an essential process in the long-term service of bridges, and measuring vibration is indispensable in the process of damage identification [1]. Strain gauges [2] and acceleration sensors [3] are the traditional sensors to measure the vibration of bridge structures, but due to the disadvantages of the contact measurement methods, they are gradually replaced by non-contact measurement ones such as Laser Doppler Vibrometer (LDV) [4] and Global Positioning System (GPS) [5]. With the development of computer vision, LDV and GPS are gradually replaced by some optical measurement methods with higher precision and lower cost. Kanade-Lucas-Tomasi (KLT) optical flow method (called KLT in this paper) [6] is a target tracking technique, which is the improvement of Lucas-Kanade (LK) optical flow method [7]. The inter-frame motion information of target points is obtained by tracking the target points on a set of images stored in the form of frames by KLT. KLT performs well in target tracking [8], which has been applied to vibration measurement of bridge models [9].

Fixed cameras are used in traditional optical measurement methods. However, it may be difficult to select appropriate installation locations for fixed cameras under some conditions (such as cross-valley bridges), hence, UAVs have become popular in bridge measurement [10,11]. Due to the movement of UAVs during the measurement, the false displacement is included in the measured displacement results [12]. Homography transformation is a commonly used method to correct the displacement measured by UAVs [13]. However, the premise of this method is that there exist at least four static reference points on the plane of measurement points, and that assumption is difficult to meet in field measurement. Though the modal parameters of a bridge can be obtained through the differential filtering technique [14] from the displacement signals measured by UAVs without using any reference points [15], however, the real displacement signals cannot be obtained by this method. The real displacement signal of the bridge can be extracted from the original displacement signal measured by UAVs by Empirical Mode Decomposition (EMD [16]) or a filter combined with Fourier Transform (FT [17]) and inverse Fourier transform (IFT [18]) without reference points [19], however, the premise of EMD is that the frequency of the bridge is very different from the motion frequency of the UAV and the premise of the FT filter is that the frequency of the bridge is roughly known, both are not always practical. On the other hand, the world coordinates of the target points can be recovered by the three-dimensional reconstruction method, and the target points displacement time-history curve can be obtained [20]. Similarly, this method requires that the plane of the reference points is

parallel to that of the target points, and the distance between two planes is also needed to be known in advance, and that is difficult to meet in the field measurement. It can be seen that the existing techniques for correcting the displacement measured by UAVs are extremely strict for the reference point selection limit the UAV applications, so it is very important to find a more advanced technique to make the measurement work more feasible.

With the advancement of neural networks in recent years, they have gradually been applied to solve the problem of unstable shooting of UAVs [21, 22]. The characteristic relationship between the target points on the measured bridge and the reference points on the background can be learned by the Feedforward Neural Network (FNN) when the bridge structure is stationary, and the coordinates of the target points can be predicted by the trained FNN when the structure is vibrating. Finally the real bridge displacement can be obtained through processing each frame in series [23]. Though the false displacement caused by UAVs can be eliminated by this method, however, the two-layer FNN is obviously difficult to achieve high precision. With the development of deep learning [24], CNNs [25] with higher feature extraction ability were applied to the field of UAV measurement. A CNN was used to detect vehicles based on UAV images, and its robustness was confirmed [26]. The cracks in bridge images taken by UAVs can be accurately detected by a CNN [27]. A CNN was used to predict the real displacement of a bridge to obtain better displacement correction effect of UAV measurement [28] than that of a FNN [23], however, the factors affecting the experimental results (such as the number of reference points and measurement time, etc.) and the hyper-parameters of the CNN affecting the network performance were not deeply investigated, which will be explored in this paper. Some CNN hyper-parameters, such as learning rate, number and size of convolution kernels, number of network layers and mini-batch size, are important to a network performance [29]. And it is difficult to manually select the best combination of hyper-parameters before network operation.

The optimization of network hyper-parameters is a key step to get a network with good performance. Grid search is a traditional hyper-parameters optimization method, which is time-consuming as it obtains optimal hyper-parameters by trying all of their possible combinations [30, 31]. Random search [32] is based on the assumption that the random sample size is large enough. The principle of Bayesian Optimization (BO) [33] is that the probability model of the objective function is established to select the optimal hyper-parameters to evaluate the real objective function, so as to optimize the hyper-parameters. Compared with grid search and random search, BO considers the previous hyper-parametric information, so as to achieve better robustness [34]. Hence, BO is used more widely to optimize the hyper-parameters of CNNs [35, 36].

The hyper-parameters of a CNN are optimized by BO in this paper. The correction effect of a BO-based CNN on UAV displacement measurement results is not only affected by the hyper-parameters, but also affected by some other variables in the measurement. In this paper, the bridge model video is collected by a UAV, the target points and the reference points are tracked with the KLT, and their coordinates on each frame can be obtained, which are used to train and optimize the CNN to eliminate the false displacement and the displacement time-history curves can be obtained. Finally, the natural frequency of the model can be extracted from displacement signals by OMA [37, 38]. Three comparative experiments are set up in this paper to explore the influence on the correction effect of the number of reference points, the shooting duration of the UAV and the distance between the model and reference points.

2. Methods

2.1 KLT optical flow method

The unchanged pixel value of each point and the small motion of each point between adjacent frames are presumed in the KLT. Moreover, spatial consistency is another premise of the KLT, that is, the adjacent pixels have similar motion between adjacent frames.

The target points selected in the first frame are used to calculate the inter-frame vector to realize target tracking by KLT. Suppose W is a constant size window, which is centered on the position of a target point on each frame image and d_x and d_y are the motion of the target point along the x-axis and y-axis between frames; the motion, $\boldsymbol{d} = [d_x \quad d_y]^T$, can be obtained by minimizing the value of the following formula [6]:

$$\varepsilon = \iint_W [I_2(\boldsymbol{X} + \boldsymbol{d}) - I_1(\boldsymbol{X})]^2 \omega(\boldsymbol{X}) d\boldsymbol{X} \tag{1}$$

where ε is the residual error, I_1 is the image of the previous frame, I_2 is the image of the next frame, $\omega(\boldsymbol{X})$ is the weight function, $\boldsymbol{X} = [x \quad y]^T$, (x, y) are the coordinates of the pixels in the window, W.
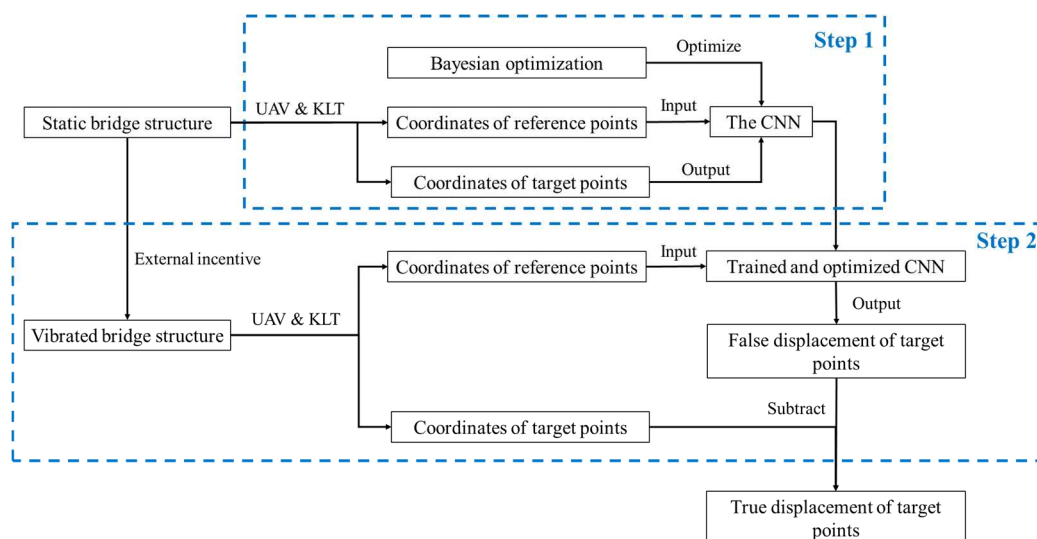

Fig. 1. Principle of displacement correction by BO-based CNN.

2.2 Convolutional Neural Network

The principle of eliminating false displacement from the UAV displacement signal with a BO-based CNN is shown in Figure 1. The first set of coordinates of the target points and the reference points can be obtained by KLT from the video of the stationary bridge, then they are used respectively as the input and output of the network to train and optimize the network. Then the structure will be excited to vibrate, and the second set of coordinates of the target points and the reference points are obtained by KLT and the coordinates of the reference points are used as the network input to predict the target-point false displacement caused by the ego-motion of the UAV. Finally, the real displacement of the target points can be obtained by subtracting the predicted false displacement from the KLT-obtained displacement of the target points during vibration.

In the Step 1 of Figure 1, the input data size of the CNN is $2n \times p_1$, and the output data size is $2m \times p_1$, where n and m are, respectively, the numbers of the reference points and the target points, p_1 is the number of the training samples determined by the video duration of the static structure. In the Step 2 of Figure 1, the input data size of the CNN is $2n \times p_2$, and the output data size is $2m \times p_2$, where p_2 is the number of test set determined by the video duration of the vibrated structure. The coordinates of reference points and target points are the continuous coordinate sequences obtained by KLT, and the displacement time-history curve of a point of interest can be obtained from its continuous coordinate sequence.

The basic framework of the CNN in this paper is shown in Figure 2, where N is a positive integer, and the section depth of network is determined by $N + 1$.

The features of the input data are extracted by the convolution layers, which are the core of a CNN. The size and number of convolution kernels are closely related to the CNN ability to extract the characteristics of the input data [39]. The convolution process is shown in Figure 3, where, the stride is the movement distance of the convolution for scanning the input data.
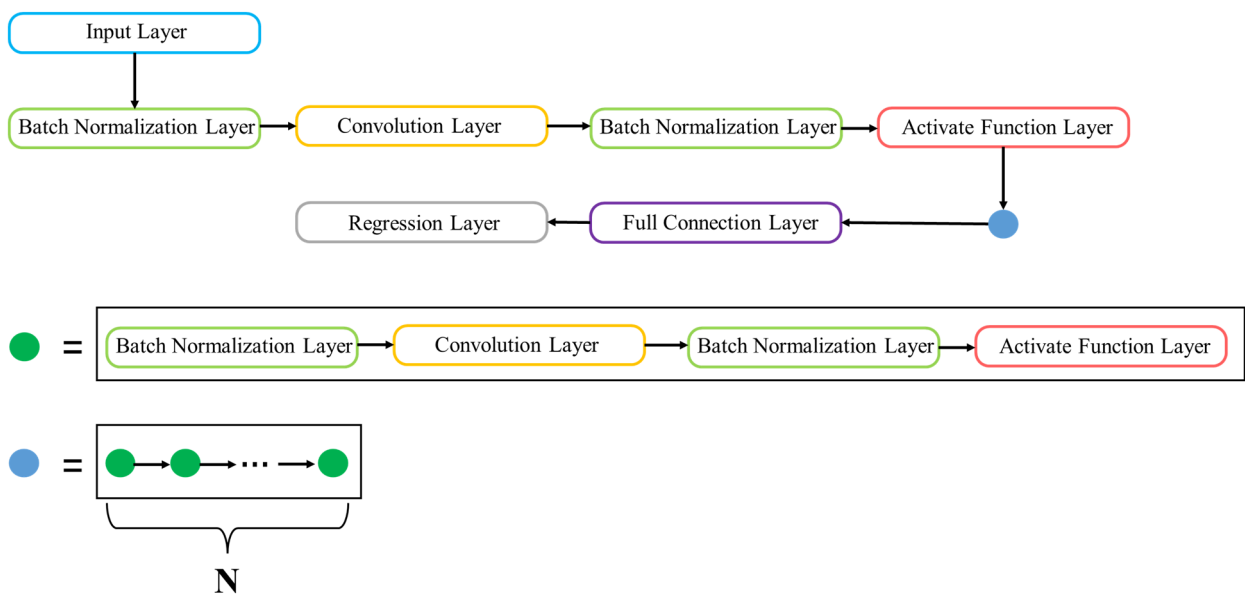
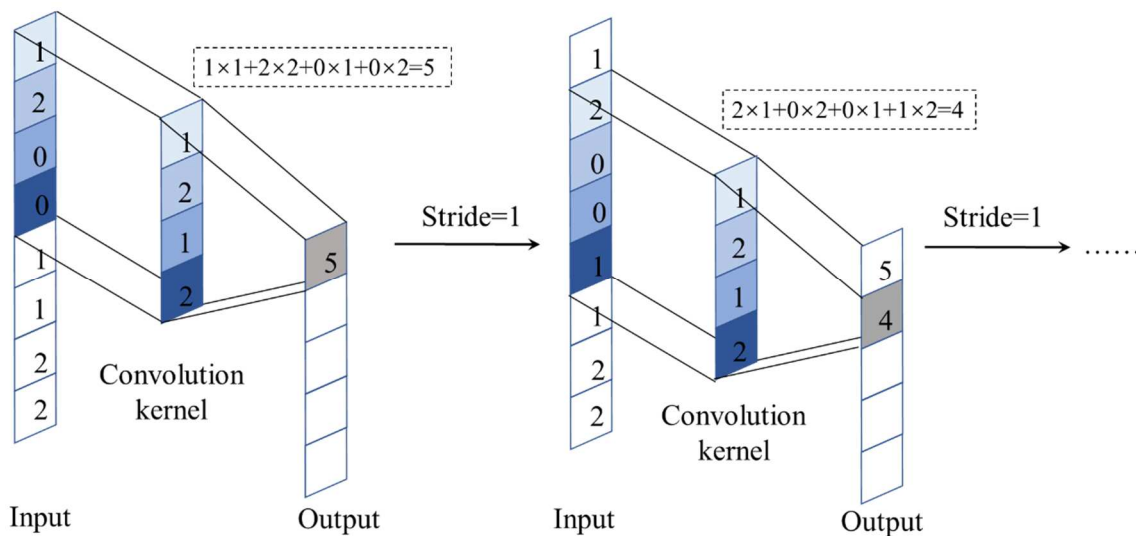

Fig. 2. Structure of the CNN.

Fig. 3. Convolution process.

The activation function layer is introduced to map the data nonlinearly, so that any nonlinear function can be approximated by a network. In this paper, the function used in the activation function layer is Leaky Relu, and its expression is [40]:

$$y_i = \begin{cases} x_i & , & x_i \geq 0 \\ scale \times x_i & , & x_i < 0 \end{cases} \tag{2}$$

The linear combination output of the previous layer can be normalized as shown in Equation (3) by the batch normalization layer to make the data distribution consistent, so as to speed up the training of the CNN and reduce its sensitivity to network initialization [41]:

$$y_i = \gamma \left(\frac{x_i - \mu(\boldsymbol{X})}{\sqrt{\sigma^2(\boldsymbol{X}) + \epsilon}} \right) + \beta \tag{3}$$

where, $\boldsymbol{X} = (x_1, \ x_2, \ \cdots, \ x_j)$, and j is the batch size, y_i is the output, $\mu(\boldsymbol{X})$ and $\sigma(\boldsymbol{X})$ are, respectively, the mean and standard deviation of all elements in $\boldsymbol{X}$, ϵ is a minimal quantity, γ and β are constants.

All neurons in the previous layer are connected with each neuron of the fully connected layer as shown in Figure 4 to integrate the local information transmitted from the previous layers. The input data is processed by the fully connected layer according to the following formula:

$$y_i = \omega x_i + b \tag{4}$$

where, y_i is the layer output and x_i is the layer input, b and ω are, respectively, the bias and weight of the fully connected layer.

The loss function of the regression layer is expressed as follows:

$$J = \frac{1}{2} \sum_{i=1}^{R} \frac{(t_i - y_i)^2}{R} \tag{5}$$

where J is the loss of the network, R is the size of data set used to calculate the loss, t_i and y_i are, respectively, the target output and the CNN prediction. With the network training, J continues to decrease.
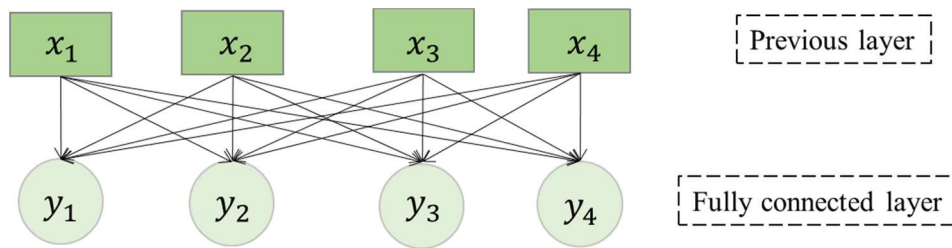

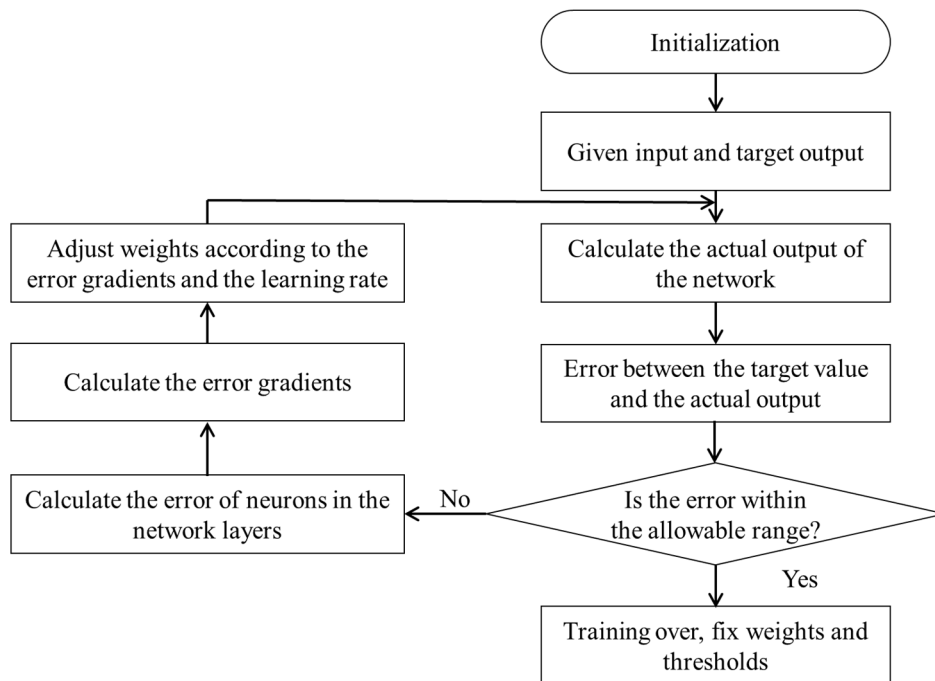**Fig. 4.** Connection of neurons of fully connected layer.

Fig. 5. Training process of CNN.

Table 1. The ranges of hyper-parameters.

Initial learning rate	Epoch	Mini-batch size	Convolution kernel size of the first convolution layer	Convolution kernel number of the first convolution layer	Convolution kernel stride	Section depth
[0.0001, 1]	[50, 300]	[10, 70]	[2, 6]	[64, 128]	[1, 4]	[1, 4]

2.3 Hyper-parameter optimization

2.3.1 Hyper-parameters

The hyper-parameters of the CNN optimized by BO in this paper include the initial learning rate, training epoch, mini-batch size, convolution kernel size, convolution kernel number, convolution kernel stride and section depth. The initial learning rate, training epoch, mini-batch size and section depth will be discussed in this section.

The performance of the network can be improved with appropriate hyper-parameter combination. The training process of a CNN is shown in Figure 5. The weights will be tremendously adjusted when the learning rate is too high, resulting in the training reaching a suboptimal result. While, if the learning rate is too small, the convergence speed of the model is slow. The mini-batch is the subset selected from the total training samples to save training time, while an over-small batch size cannot save the training time, while the network generalization ability will be compromised for an over-large batch size [42]. One training epoch is a forward and back propagation of the network training process shown in Figure 5. The prediction ability of the network is improved with the increase of the number of training epochs. But more epochs will increase the training time. With the deepening of the network hierarchy, the degree of feature abstraction of input data extracted by the network is improved, and the prediction effect of the model is better. However, the over fitting in the model training process can be caused for a very deep network.

2.3.2 Bayesian optimization

In this paper, the high-performance network is obtained by optimizing the CNN hyper-parameters. The BO process, shown in Figure 6, is to find the hyper-parameter combination that makes the network perform best in the validation metric, and its expression is as follow:

$$x^* = \underset{x \in X}{argmin} \ f(x) \tag{6}$$

where $argmin$ represents the function to find the independent variable corresponding to the minimum value of the objective function, x^* is the optimal combination of hyper-parameters and X is the range of hyper-parameters, which is shown in Table 1. And $f(x)$ is the objective function need to be minimized, that is, the loss of the validation set when the corresponding hyper-parameter combination is used by the model, and its expression is as follows:

$$f(x) = \frac{\sum_{i=1}^{t}(p_i^* - p_i)}{t} \tag{7}$$

where t is the size of the validation set. In the process of network validation, the validation set includes the coordinate set of the target points and the reference points, where p_i is the $i-th$ sample point of the target point coordinate set. The coordinate set of the reference point in the validation set are taken as the network input, and that of the corresponding target points (prediction set) can be predicted by the network, where p_i^* is the $i-th$ sample point in the prediction set. In this paper, the size of the validation set accounts for 20% of the total training samples of the network.

As shown in Figure 6, a probability model (Gaussian process) is established by BO based on the previous evaluation results of the objective function to find the hyper-parameter combination that minimizes the objective function value. The process of BO is as follows:

(1) A group of hyper-parameter combination x_0 is randomly selected within the value range of hyper-parameters shown in Table 1;

(2) The x_0 is taken as the hyper-parameter combination of the network to be trained, so as to obtain the corresponding objective function $f(x_0)$;

(3) The probability distribution of $f(x)$ value corresponding to x can be calculated and predicted through the Gaussian process by using all the input $(x, f(x))$;

(4) The optimal x can be determined by the acquisition function [33] in the probability distribution;

(5) The x obtained from Step (4) is taken as the hyper-parameter combination of the network to be trained, so as to obtain the corresponding objective function $f(x)$;

(6) Before the maximum iteration number is reached, the $(x, f(x))$ obtained in Step (5) will be used as the input of the network to continuously update the probability model to obtain a new $(x, f(x))$, and when the maximum iteration number is reached, according to Equation (5), the x corresponding to the minimum value of all objective functions $f(x)$ obtained in Step (5) is the optimal hyper-parameter combination x^*.

3. Experiment

3.1 Experimental process

In this paper, the steel frame shown in Figure 7a is used as the bridge model for experiment, which has 9.8 m span and 56 nodes on its front plane (Figure 7b). The model is composed by hollow round bars (Figure 7c) and bolt balls (Figure 7d). The model is hinged at both ends as shown in Figure 7e. As shown in Figure 8, the tool used to collect bridge model video is a quad-rotor UAV (Mavic Air 2, Da-Jiang Innovations, Shenzhen, China), with an image resolution of 3840 × 2160 and an acquisition frequency of 30 frames/s. Because the natural frequency of the model used in this experiment is low, the image acquisition frequency of the UAV can meet the experimental requirements. When the frequency of the measured object is high, a UAV with higher acquisition frequency can be selected. When the amplitude of the measured object is large, the distance between the UAV and the measured object can be increased to ensure that the target points can be tracked.
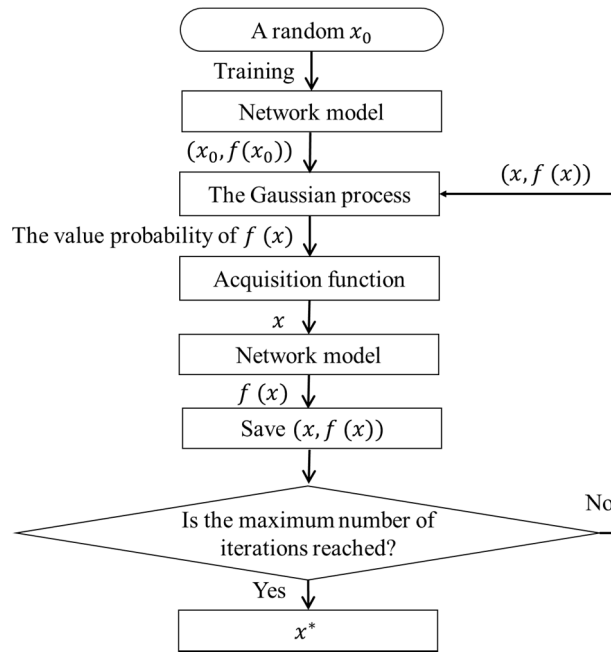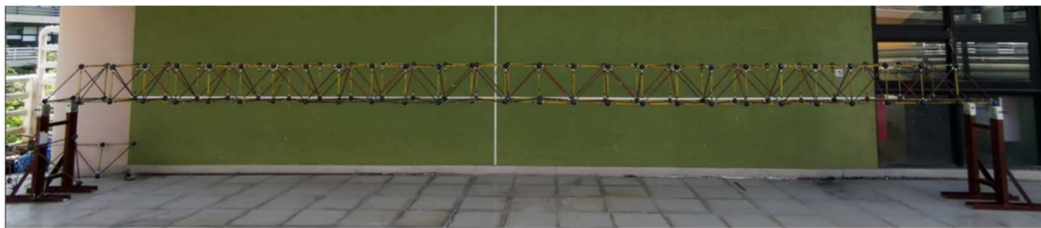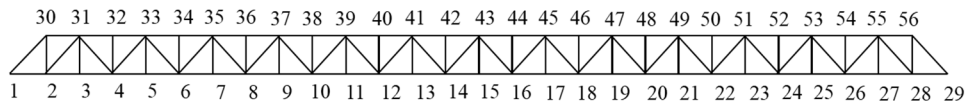
Fig. 6. The Bayesian Optimization process.

(a)

(b)

(c)   (d)   (e)

Fig. 7. Experimental model: (a) Bridge model; (b) Model node; (c) Hollow round bar; (d) Bolt connecting balls; (e) Supports.
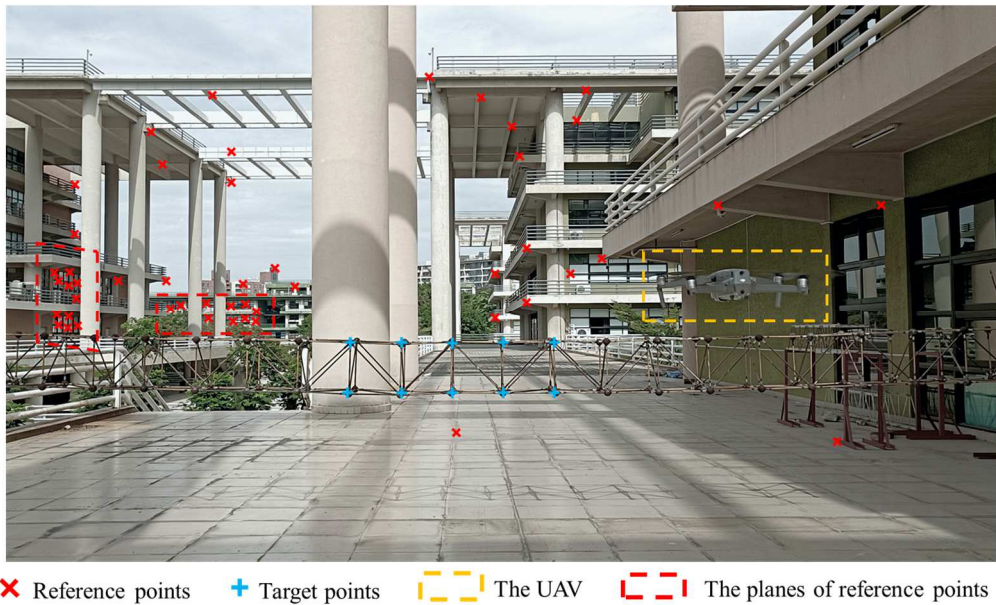

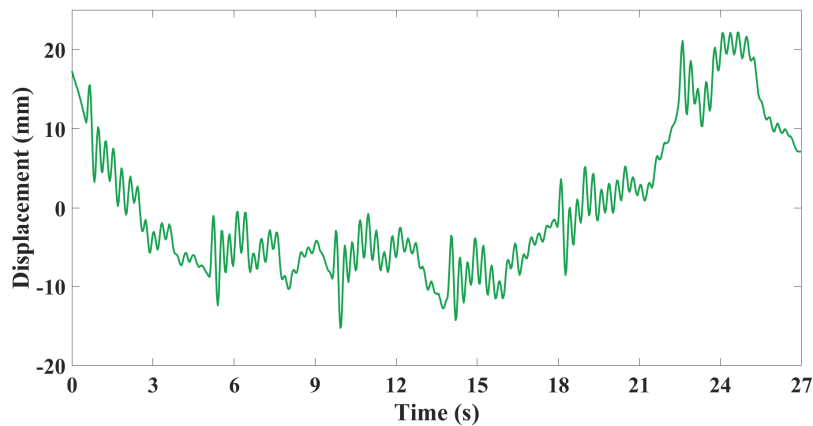Fig. 8. The UAV used in experiment.

Fig. 9. Experimental layout.

Fig. 10. The uncorrected bridge vibration displacement signal.

The experimental layout is shown in Figure 9. Some reference points are on the same lanes, the other reference points are from different planes of the model background. The distances between the model and two planes are respectively about 100 m and 300 m. The horizontal distance between the reference points and the model ranges from 2 m to 600 m, and the vertical distance between them ranges from 0 m to 25 m. The target points are the 13th to 17th nodes and the 41st to 45th nodes of the model. The KLT, CNN, BO and data processing are realized using MATLAB (MathWorks Inc., Na-tick, MA, USA) software.

3.2 Experimental scheme

Three comparative experiments are set up to study the influence of the number of reference points (N_R), the UAV shooting duration of the static model (T) and the distance between the model and reference points (D) on the correction effect of the BO-based CNN on displacement measurement results of the UAV. N_R, T and D are independent to each other. Firstly, T is kept unchanged and N_R is changed. By comparing the experimental results with different N_R, the optimal N_R is obtained. Here, all reference points are from different planes at different distance from the model. Then N_R is kept unchanged and T is changed, the correction results obtained with different T are compared to obtain the most appropriate T. Here, changing T leads to change of the number of the CNN samples. Finally, T and the same number of the reference points are kept unchanged, the correction effects of two planes at different distances from the model are compared to study the influence of D on correction effect.

4. Experimental Results and Analysis

For better comparison, the measurement results of Node 15 of the model are taken for display.

4.1 Influence of N_R on the correction effect

As shown in Figure 1, the uncorrected displacement signal can be obtained by KLT from the bridge video collected by a UAV. The uncorrected bridge vibration displacement signal is shown in Figure 10, it can be seen from the figure that the signal has serious drift caused by the UAV's ego-motion during the measurement. The signal correction in the experiment is realized by a set of algorithms combined with CNN framework and BO, that is, CNN the training process and the optimization process of BO are carried out together, as shown in Figure 6.
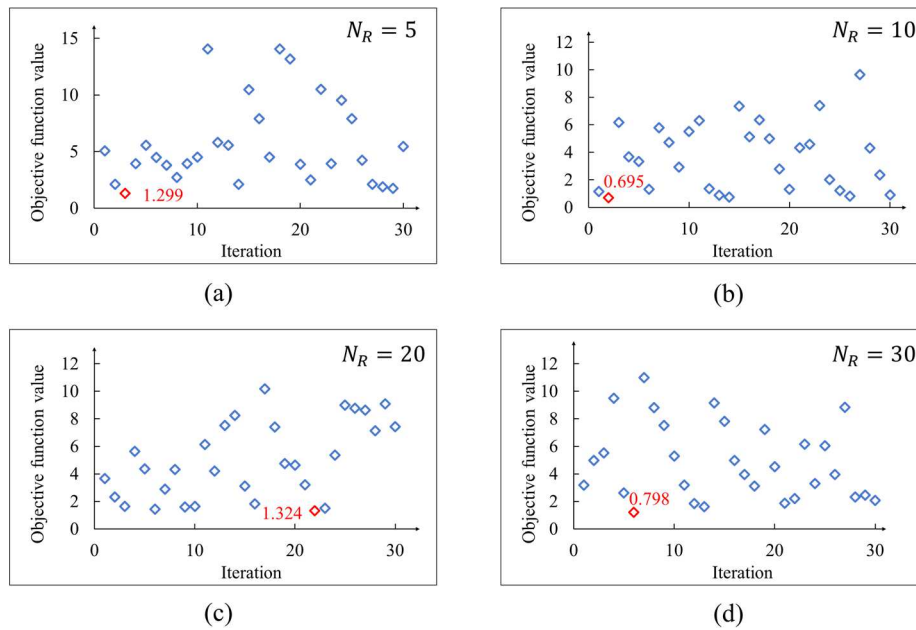
Fig. 11. The results of BO when T is 5 min: (a) Result at N_R of 5; (b) Result at N_R of 10; (c) Result at N_R of 20; (d) Result at N_R of 30.
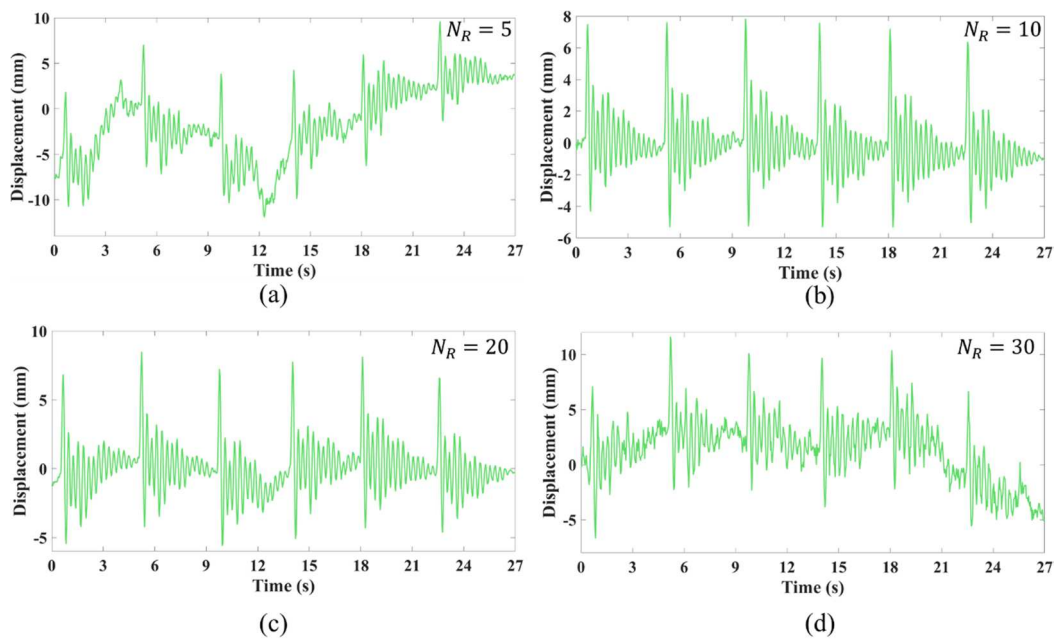

Fig. 12. The displacement time-history curves when T is 5 min: (a) Result at N_R of 5; (b) Result at N_R of 10; (c) Result at N_R of 20; (d) Result at N_R of 30.

To explore the influence of N_R on correction effect of BO-based CNN, the correction results with N_R of 5, 10, 20 and 30 are compared when T is 5 min, and the total number of the CNN samples is 9000 in each of the four cases. The results of BO for the CNN in the four cases are shown in Figure 11, in which the maximum iterations number is 30. Therefore, the minimum value of the objective function in the scatter diagram (Figure 11) is the optimal case of the hyper-parameter combination defined in Equation (5).

The target points displacement time-history curves (Figure 12) can be obtained from the optimized CNN which is the network in the case of the minimum value of the objective function (Figure 11); the time-history curve at N_R of 5 and 30 are poor, which have evident drift, while that at N_R of 10 and 20 are relatively better. As shown in the Figure 12, the time-history curves fluctuate around the horizontal line. The more drift the curves have, the farther the curves are away from the displacement of 0. The Root Mean Square (RMS [43]) can represent the drift degree of the displacement curves [28]. The RMS at N_R of 5, 10, 20 and 30 is, respectively, 4.436, 1.805, 1.922 and 3.120, which shows that the correction effect of BO-based CNN on displacement is the best at N_R of 10.

The Power Spectral Density (PSD) diagrams (Figure 13) can be obtained from the above time-history curves. As demonstrated in OMA [44], the peaks of the PSD curves correspond to the natural frequency of structure [38]. Figure 13 shows that the peaks of the PSD curves in the four cases are similar, which are at 3.365 Hz, 3.369 Hz, 3.370 Hz and 3.370 Hz respectively. However, the curves at N_R of 5 and 30 have another peak close to 0 Hz, which are 0.043 Hz and 0.044 Hz respectively; these peaks do not correspond to the real natural frequency of the structure but are induced by the phenomenon of zero drift [15]. Therefore, to summarize the above analysis of time-history curves, the correction effect of BO-based CNN is the best when $N_R = 10$.
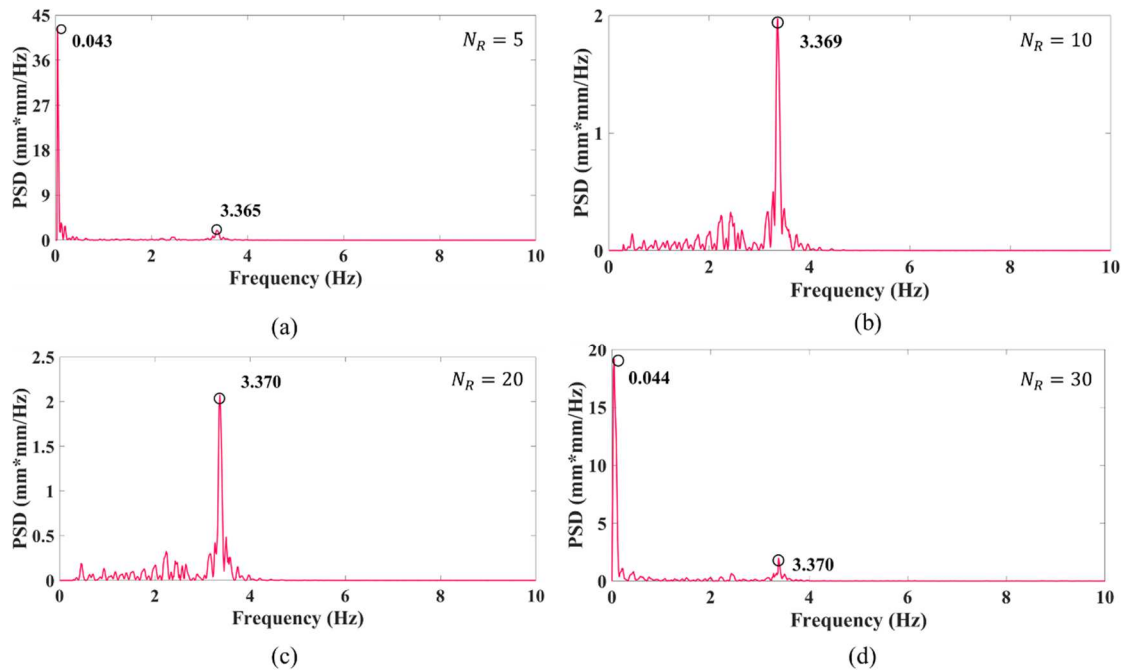
Fig. 13. The PSD curves when T is 5 min: (a) Result at N_R of 5; (b) Result at N_R of 10; (c) Result at N_R of 20; (d) Result at N_R of 30.
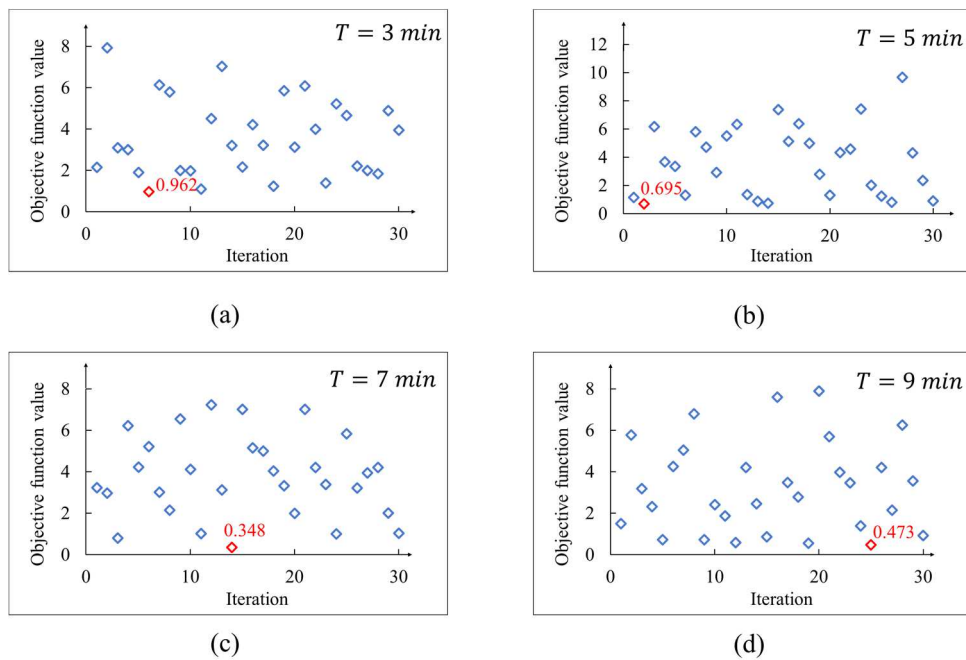

Fig. 14. The results of BO when N_R is 10: (a) Result at T of 3 min; (b) Result at T of 5 min; (c) Result at T of 7 min; (d) Result at T of 9 min.

4.2 Influence of T on the correction effect

The influence of T on BO-based CNN correction effect is explored in this section. The previous section shows that the correction effect is the best at N_R of 10. Therefore, N_R is kept at 10, the correction effect of BO-based CNN is compared in this section under the four T cases of 3 min, 5 min, 7 min and 9 min; the total numbers of the CNN samples in the four cases are 5400, 9000, 12600 and 16200, respectively. The BO results in the four cases are shown in Figure 14. The displacement signals in the four cases are shown in Figure 15; there is drift in the time-history curve at T of 3 min and its RMS is 2.183, while the RMS of the other three cases are respectively 1.805, 1.810 and 1.798, which are close to each other. The degree of consistency between time-history curves can be assessed by $TRAC$ (Time Response Assurance Criterion) value [45]. The $TRAC$ between the time-history curves at T of 5 min and 7 min is 0.978, and that between the time-history curves at T of 5 min and 9 min is 0.986, therefore, these three time-history curves are very close. According to the RMS and $TRAC$, the curve is stable at T of 5 min, and with the increase of T, the change of curve is not obvious, which shows that the correction effect of BO-based CNN reaches the best at T of 5 min.

The PSD curves of the above displacement signals are shown in Figure 16, their peaks are at very similar locations, which are 3.369 Hz, 3.369 Hz, 3.370 Hz and 3.368 Hz respectively. These frequencies obtained at N_R of 10 are also very close to those obtained in the previous section, which proves the reliability of extracting the natural frequency of the structure at N_R of 10.
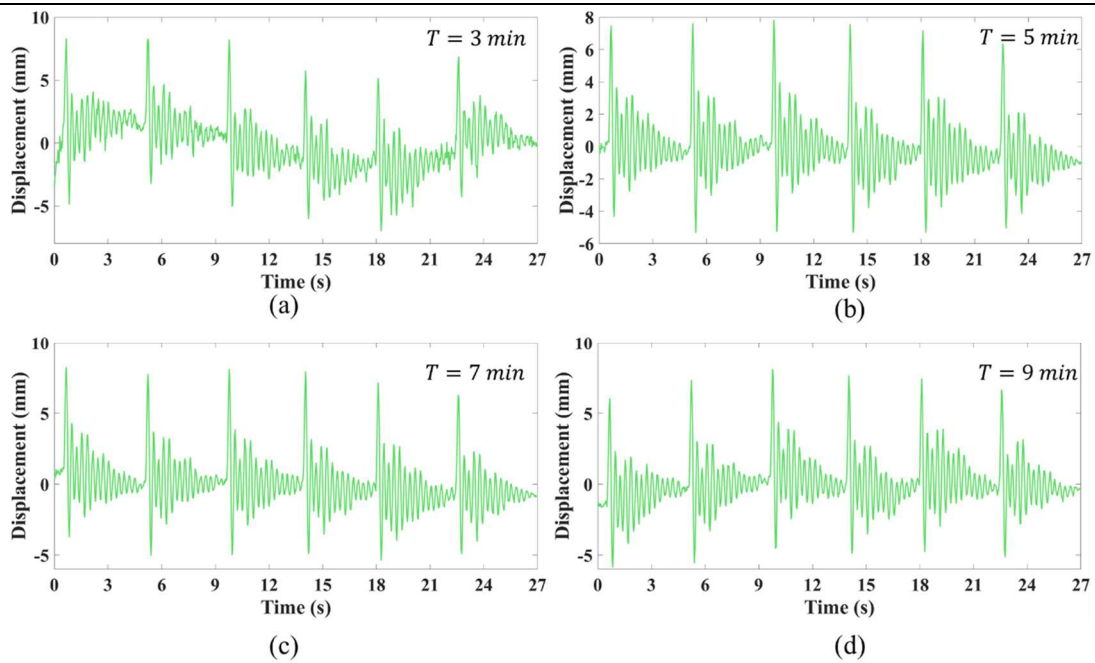
Fig. 15. The displacement time-history curves when N_R is 10: (a) Result at T of 3 min; (b) Result at T of 5 min; (c) Result at T of 7 min; (d) Result at T of 9 min.
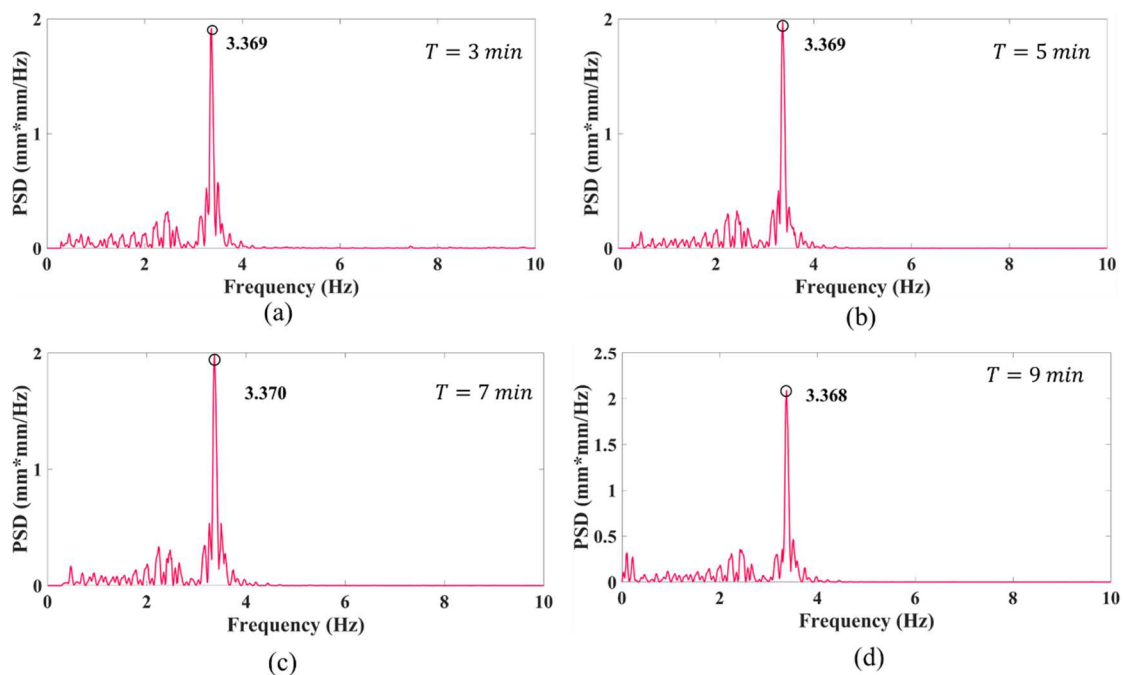

Fig. 16. The PSD curves when N_R is 10: (a) Result at T of 3 min; (b) Result at T of 5 min; (c) Result at T of 7 min; (d) Result at T of 9 min.

4.3 Influence of D on the correction effect

To explore the influence of the distance between the reference points and structure on the correction effect of the BO-based CNN, the correction results obtained for the reference points of two planes with different distances from the model are compared in this section at N_R of 10 and T of 5 min, and the total numbers of the CNN samples in both cases are 9000. The distances between the two planes and the model are about 100 m and 300 m, respectively. The BO results of two cases are shown in Figure 17. The displacement time-history curves are shown in Figure 18; the time-history curve has more drift for $D = 300\ m$, while the curve is relatively stable for $D = 100\ m$, and the *RMS* of two cases are 0.369 and 1.641, which show that the curve for D of $100\ m$ is better than that for D of $300\ m$. The PSD curves (Figure 19) are obtained from the two displacement signals (Figure 18). The relative error of natural frequencies obtained in the two cases is only 0.029%. To summarize the above analysis, the closer the reference points are to the structure, the better the correction effect of the BO-based CNN.
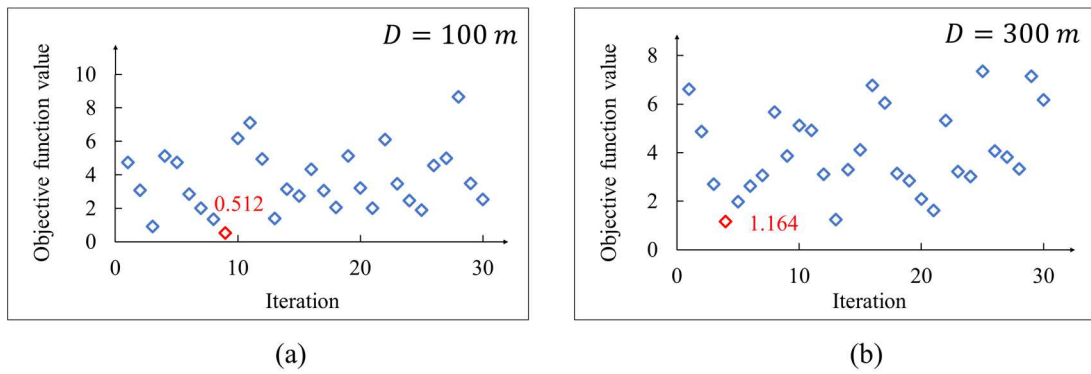
Fig. 17. The BO results of different D: (a) Result at D of 100 m; (b) Result at D of 300 m.
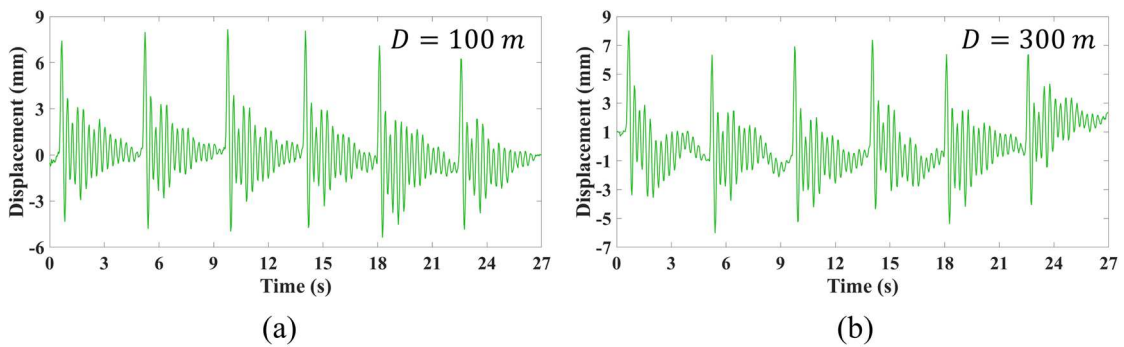

Fig. 18. The displacement time-history curves of different D: (a) Result at D of 100 m; (b) Result at D of 300 m.
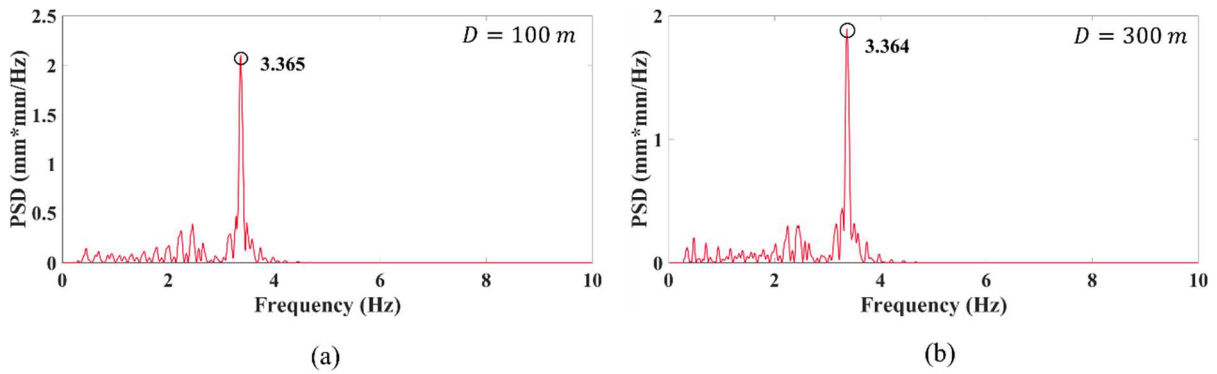

Fig. 19. The PSD curves for different D: (a) Result at D of 100 m; (b) Result at D of 300 m.
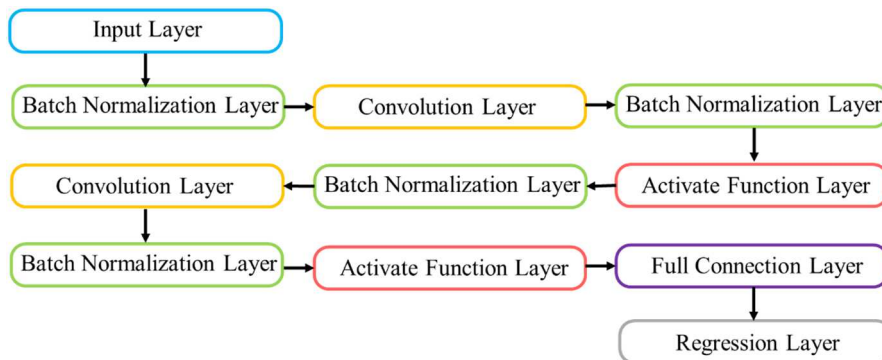

Fig. 20. The optimized CNN structure at T of 5 min and N_R of 10.

Table 2. The optimal hyper-parameter combination of the network after BO at T of 5 min and N_R of 10.

Initial learning rate	Epoch	Mini-batch size	Convolution kernel size of the first convolution layer	Convolution kernel number of the first convolution layer	Convolution kernel stride	Section depth
0.00016209	279	48	5	88	1	2

According to Section 4, the BO-based CNN has the best correction effect at T of 5 min and N_R of 10. Moreover, the closer the reference points are to the structure, the better the correction effect of the BO-based CNN. The optimal hyper-parameter combination of the BO-based CNN at T of 5 min and N_R of 10 (for the case of the objective function value of 0.695 in Figure 11b) is shown in Table 2, and the corresponding CNN structure is shown in the Figure 20.

5. Discussion and Conclusion

5.1 Discussion

(1) It can be seen from Section 4.1 that the correction effect of the BO-based CNN on the displacement signal is the best when N_R is 10 among the four cases of N_R, and when N_R is 10, there is no other peak in the PSD curve to affect the extraction of the natural frequency. Poor correction effect of the BO-based CNN can be found in the cases with too few or too many reference points. When N_R is too small, the feature relationship between the target points and the reference points may not be fully learned by the network to effectively predict the false displacement of the target points, while interference in network learning process can be caused by too many reference points to reduce the prediction ability of the network.

(2) Section 4.2 shows that T reaches the best correction effect at 5 min. As mentioned earlier, changing T leads to changing the number of the CNN samples (N_S). The network prediction ability can be improved with the increase of N_S, which is relatively stable when N_S reaches a certain amount, and the network prediction ability even can be degraded by noise caused by too many samples.

(3) Section 4.3 shows that, although the natural frequency can be easily extracted from the PSD curves of both D cases, the corrected displacement curve in the case of $D = 100\ m$ is better than that in the other case. It demonstrates that the correction effect of the BO-based CNN is better when the distance between reference points and structure is closer. The farther the reference point is from the UAV, the lower its resolution in the image, that is, the fewer pixels it occupies in the image. As the KLT tracks the pixels in an image, if the target points occupy only fewer pixels in the image, the tracking effect may be poor. This is why the farther the reference point is from the model, the worse the correction effect of the BO-based CNN is.

(4) Although the optimal values of N_R and T have been determined through the experiment, the universality of the optimal values and the influence of D on correction effect still need to be verified through experiments in more practical scenarios in follow-up work. In addition, the real bridge measurement environment is more complex, so the method proposed in this paper will be tested in actual bridge measurement in follow-up work.

5.2 Conclusion

In this paper, a bridge vibration measurement method by a UAV based on BO and a CNN was proposed. The conclusions based on the experiment are as follows:

(1) The correction effects of the BO-based CNN on the displacement measured by a UAV are the best when N_R is 10 and T is 5 min.

(2) The smaller the distance between the reference points and the structure is, the better the correction effect is.

In this paper, based on the use of a UAV as a measurement tool, a CNN was used as a correction technology, and BO was used to optimize the network and improve the performance of the CNN. In addition, the influencing factors of the proposed method in the process of bridge vibration measurement also explored in this paper, which provided guides on how to determine the location and number of the reference points and the shooting time of UAVs in actual bridge displacement measurement. Experiments in this paper have shown that the reference points do not have to be on the same plane, so the method of correcting the bridge displacement measured by UAVs with a BO-based CNN has a good prospect in the actual bridge displacement measurement.

Author Contributions

G. Chen planned the scheme and initiated the project; Z. Yan suggested the experiments and conducted the experiments; S. Teng analyzed the empirical results; F. Cui developed the mathematical modeling; D. Bassir examined the theory validation. The manuscript was written through the contribution of all authors. All authors discussed the results, reviewed, and approved the final version of the manuscript.

Acknowledgments

Conflict of Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and publication of this article.

Funding

Data Availability Statements

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Reference

[1] Doebling, S.W., Farrar, C.R., Prime, M.B., Digest, V., A Summary Review of Vibration-Based Damage Identification Methods, *The Shock and Vibration Digest*, 30, 1998, 91-105.

[2] Kovačič, B., Kamnik, R., Štrukelj, A., Vatin, N., Processing of Signals Produced by Strain Gauges in Testing Measurements of the Bridges, *Procedia Engineering*, 117, 2015, 795-801.

[3] Soyoz, S., Feng, M.Q., Long-Term Monitoring and Identification of Bridge Structural Parameters, *Computer-aided Civil & Infrastructure Engineering*, 24, 2010, 82–92.

[4] Siringoringo, D.M., Fujino, Y., Noncontact Operational Modal Analysis of Structural Members by Laser Doppler Vibrometer, *Computer-aided Civil and Infrastructure Engineering*, 24, 2010, 249-265.

[5] Psimoulis, P., Pytharouli, S., Karambalis, D., Stiros, S., Potential of Global Positioning System (GPS) to measure frequencies of oscillations of engineering structures, *Journal of Sound & Vibration*, 318, 2008, 606-623.

[6] Shi, J., Tomasi, C., Good Features to Track, *In Proceedings of Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 6000, 2002, 593-600.

[7] Lucas, B.D., An Iterative Image Registration Technique with an Application to Stereo Vision (DARPA), *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 81, 1981, 674-679.

[8] Liu, F., Yundong, W.U., Cai, G., Chen, S., Science, S.O., Research of Feature Point Tracking Algorithm for UAV Video Image Based on KLT, *Journal of Jimei University*, 22, 2017, 73-80.

[9] Zhu, J., Lu, Z., Zhang, C., A marker-free method for structural dynamic displacement measurement based on optical flow, *Structure and Infrastructure Engineering*, 18, 2020, 1-13.

[10] Ellenberg, A., Kontsos, A., Moon, F., Bartoli, I., Bridge related damage quantification using unmanned aerial vehicle imagery, *Structural Control & Health Monitoring*, 23, 2016, 1168-1179.

[11] Hoskere, V., Park, J.W., Yoon, H., Spencer, B.F., Vision-Based Modal Survey of Civil Infrastructure Using Unmanned Aerial Vehicles, *Journal of Structural Engineering*, 145, 2019, 04019062.

[12] Rodríguez-Canosa, G., Stephen, T., Jaime, D.C., Antonio, B., Bruce, M., A Real-Time Method to Detect and Track Moving Objects (DATMO) from Unmanned Aerial Vehicles (UAVs) Using a Single Camera, *Remote Sensing*, 4, 2012, 1090-1111.

[13] Zhai, Y., Shah, M., Visual attention detection in video sequences using spatiotemporal cues, *In Proceedings of Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara, CA, USA, October 23-27, 2006.

[14] Zhu, X., Chen, Z., Tang, C., Mi, Q., Yan, X., Application of two oriented partial differential equation filtering models on speckle fringes with poor quality and their numerically fast algorithms, *Applied Optics*, 52, 2013, 1814-1823.

[15] Zhang, J., Wu, Z., Chen, G., Liang, Q., Comparisons of Differential Filtering and Homography Transformation in Modal Parameter Identification from UAV Measurement, *Sensors*, 21, 2021, 5664.

[16] Wang, G., Chen, X., Qiao, F., Wu, Z., Huang, N.E., On Intrinsic Mode Function, *Advances in Adaptive Data Analysis*, 2, 2010, 277-293.

[17] Bochner, S., Fourier transforms, *Annals of Mathematics Studies*, 7, 1949, 145–151.

[18] Lazaro, J.A., Wessel, R., Koppenborg, J., Dudziak, G., Blewett, I.J., Inverse Fourier transform method for characterizing arrayed-waveguide gratings, *IEEE Photonics Technology Letters*, 15, 2003, 93-95.

[19] Yan, Z., Teng, S., Luo, W., Bassir, D., Chen, G., Bridge Modal Parameter Identification from UAV Measurement Based on Empirical Mode Decomposition and Fourier Transform, *Applied Sciences*, 12, 2022, 8689.

[20] Wu, Z., Chen, G., Ding, Q., Yuan, B., Yang, X., Three-Dimensional Reconstruction-Based Vibration Measurement of Bridge Model Using UAVs, *Applied Sciences*, 11, 2021, 5111.

[21] Lirong, D., Riuan, L., Tingting, C., Geometric Correction Algorithm for UAV Remote Sensing Image Based on Neural Network, *In Proceedings of 2016 International Conference on Artificial Intelligence and Engineering Applications (AIEA 2016)*, 63, 2016, 59-63.

[22] Ruian, L., Nan, L., Beibei, Z., Tingting, C., Ninghao, Y., Geometry correction Algorithm for UAV Remote Sensing Image Based on Improved Neural Network, *IOP Conference*, 322, 2018, 072002-.

[23] Weng, Y., Shan, J., Lu, Z., Lu, X., Spencer, B.F., Homography-based structural displacement measurement for large structures using unmanned aerial vehicles, *Computer-Aided Civil and Infrastructure Engineering*, 36, 2021, 1114–1128.

[24] Hao, X., Zhang, G., Ma, S., Deep Learning, *International Journal of Semantic Computing*, 10, 2016, 417-439.

[25] Hubel, D.H., Wiesel, T.N., Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *Journal of Physiology*, 160, 1962, 106-154.

[26] Luo, X., Tian, X., Zhang, H., Hou, W., Zhang, J., Fast Automatic Vehicle Detection in UAV Images Using Convolutional Neural Networks, *Remote Sensing*, 12, 2020, 1994.

[27] Aliyari, M., Droguett, E.L., Ayele, Y.Z., UAV-Based Bridge Inspection via Transfer Learning, *Sustainability*, 13, 2021, 11359.

v28. Yan, Z., Jin, Z., Teng, S., Chen, G., Bassir, D., Measurement of Bridge Vibration by UAVs Combined with CNN and KLT Optical-Flow Method, *Applied Sciences*, 12, 2022, 5181.

[29] Seunghye, L., Jingwan, H., Mehriniso, Z., Hyeonjoon, M., Jaehong, L.J., Background Information of Deep Learning for Structural Engineering, *Archives of Computational Methods in Engineering*, 25, 2017, 1-9.

[30] Boughorbel, S., Tarel, J.P., Boujemaa, N., The LCCP for Optimizing Kernel Parameters for SVM, *In Proceedings of International Conference on Artificial Neural Networks*, 3697, 2005, 589-594.

[31] Sun, Y., Bing, X., Zhang, M., Yen, G.G., An Experimental Study on Hyper-parameter Optimization for Stacked Auto-Encoders, *In Proceedings of 2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018, 1-8.

[32] Bergstra, J., Bengio, Y., Random Search for Hyper-Parameter Optimization, *Journal of Machine Learning Research*, 13, 2012, 281-305.

[33] Snoek, J., Larochelle, H., Adams, R.P., Practical Bayesian Optimization of Machine Learning Algorithms, *Advances in Neural Information Processing Systems*, 2, 2012, 2951–2959.

[34] Putatunda, S., Rama, K., A Comparative Analysis of Hyperopt as Against Other Approaches for Hyper-Parameter Optimization of XGBoost, *In Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 2018, 6-10.

[35] Hahn, L., Roese-Koerner, L., Friedrichs, K., Kummert, A., Fast and Reliable Architecture Selection for Convolutional Neural Networks, *In Proceedings of The 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2019, 179-184.

[36] Ekici, B.B., Detecting damaged buildings from satellite imagery, *Journal of Applied Remote Sensing*, 15, 2021, 032004.

[37] Brownjohn, J., Magalhaes, F., Caetano, E., Cunha, A., Ambient vibration re-testing and operational modal analysis of the Humber Bridge, *Engineering Structures*, 32, 2010, 2003-2018.

[38] Yan, W.J., Ren, W.X., An Enhanced Power Spectral Density Transmissibility (EPSDT) approach for operational modal analysis: Theoretical and experimental investigation, *Engineering Structures*, 102, 2015, 108-119.

[39] Lecun, Y., Bottou, L., Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86, 1998, 2278-2324.

[40] Maas, A.L., Hannun, A.Y., Ng, A.Y., Rectifier Nonlinearities Improve Neural Network Acoustic Models, *In Proceedings of International Conference on Machine Learning*, 2013.

[41] Ioffe, S., Szegedy, C., Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *Journal of Machine Learning Research*, 37, 2015, 448–456.

[42] Keskar, N.S., Mudigere, D., Nocedal, J., Smelyanskiy, M., Tang, P., On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima, *In Proceedings of the 5th International Conference on Learning Representations*, 2016.

[43] Margaliot, M., Hespanha, J.P., Root-mean-square gains of switched linear systems: A variational approach, *Automatica*, 44, 2008, 2398-2402.

[44] Chen, G., Wu, Z., Gong, C., Zhang, J., Sun, X., DIC-Based Operational Modal Analysis of Bridges, *Advances in Civil Engineering*, 2021, 2021, 6694790.

[45] Chen, G., Liang, Q., Zhong, W., Gao, X., Cui, F., Homography-based measurement of bridge vibration using UAV and DIC method, *Measurement*, 170, 2020, 108683.

ORCID iD

Gongfa Chen https://orcid.org/0000-0001-7232-9583
Zhaocheng Yan https://orcid.org/0000-0002-7803-3553
Shuai Teng https://orcid.org/0000-0003-1703-3362
Fangsen Cui https://orcid.org/0000-0001-6682-9768
David Bassir https://orcid.org/0000-0002-5364-9992

Publisher's Note Shahid Chamran University of Ahvaz remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.