processes

MDPI

*Article*

# Deformation Detection Method for Substation Noise Barrier Column Based on Deep Learning and Digital Image Technology

**Fayuan Wu [1,*], Mengting Mao [1], Sheng Hu [2], Xiaomin Dai [1], Qiang He [3], Jinhui Tang [1] and Xian Hong [4]**

[1] Electric Power Research Institute, Jiangxi Electric Power Co., Ltd., State Grid, Nanchang 330096, China; mengtingmao51@outlook.com (M.M.); xiaomindai605@outlook.com (X.D.); shu@163.com (J.T.)
[2] Electric Power Research Institute, State Grid Hunan Electric Power Co., Ltd., Changsha 410007, China; shenghu825@outlook.com
[3] State Grid Smart Grid Research Institute Co., Ltd., Beijing 102200, China; qianghe807@outlook.com
[4] State Key Laboratory for Performance Monitoring and Assurance of East China Jiao Tong University Rail Transport Infrastructure, Nanchang 330013, China; yanjunshu@ecjtu.edu.cn
* Correspondence: fayuanwu901@outlook.com

**Abstract:** The dynamic identification of the deformation of a noise barrier column is of great significance to the monitoring of its health. At the same time, the maximum stress of the column is an important indicator for the evaluation of its health status. Traditional contact displacement monitoring installs sensors on the structure, requires a lot of wiring and data acquisition equipment, and establishes a relatively independent and stable displacement reference system. Affected by the environment, wear, and material aging, the efficiency and reliability of data acquisition are reduced. A monitoring method based on digital image has the advantages of non-contact monitoring, high precision, and strong reliability. The existing DIC detection methods are limited by processor performance and image resolution, which are difficult to apply to engineering detection. In this paper, a structural displacement identification method based on convolutional neural networks (CNNs) and DIC technology is proposed. In this method, the data set is formed according to the column displacement cloud image obtained by DIC analysis, and the data set is enhanced by data normalization and region division. Through the analysis of the number of network layers and learning rate, the model design of the deep learning network is carried out. The high-speed camera image results of the test are introduced and identified by the static loading test of the equal-scale sound barrier. The results show that the structural displacement identification method based on CNN and DIC technology can accurately identify the displacement change in the structure, which greatly improves the efficiency of image displacement calculation using DIC technology.

**Keywords:** structural monitoring; substation noise barrier; deep learning; digital image technology; CNN

check for updates

Academic Editor: Mohd Azlan Hussain

Received: 28 November 2024
Revised: 31 December 2024
Accepted: 8 January 2025
Published: 14 January 2025

**Citation:** Wu, F.; Mao, M.; Hu, S.; Dai, X.; He, Q.; Tang, J.; Hong, X. Deformation Detection Method for Substation Noise Barrier Column Based on Deep Learning and Digital Image Technology. *Processes* **2025**, *13*, 215. https://doi.org/10.3390/pr13010215

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

With the rapid development of electric power technology and the continuous expansion of the scale of power grids [1], large-scale step-down substations in transmission and substation network systems are often built in urban areas closer to residential areas and public places [2]. As a special building structure, the substation has received extensive attention. Residents have complained more and more about the structural vibration noise generated by the transformer during operation; therefore, the main noise reduction properties of sound barriers are particularly important. The height of substation noise barriers is

usually more than 4 m due to the requirement of noise reduction, and the columns, as the main stress components of the whole structure, are subjected to greater deformation and stress [3]. Therefore, it is of great significance to study deformation detection for substation noise barrier columns to analyze the health status of substation noise barriers.

The working vibration of the transformer will drive the associated building structure to vibrate together with it [4]. In their analysis of substation sound barriers, numerous scholars have sought to elucidate the mechanisms by which these barriers reduce noise. However, less consideration is given to the mechanical structural analysis. When the health monitoring of a structure is conducted, sensors are typically employed to detect the structural responses [5], such as acceleration and displacement. This includes the use of the quadratic integration of acceleration. The strain and displacement are measured, and the displacement measurement is based on the inclination test, which can indirectly obtain displacement by combining the mechanical properties of the structure and the computational model. However, complex filtering techniques and large measurement errors make it difficult to popularize this approach [6].

At present, the research into monitoring technology for sound barriers is primarily concerned with the deformation and displacement of sound barrier columns and the breakage of sound barrier unit boards, among other things. The most commonly employed technical means include sensors, fiber optic technology, photogrammetry, and ground interferometric radar, among others [7]. However, the methods and means are not yet fully developed, mechanized and intelligent maintenance equipment has not yet been widely adopted, and sound barrier structure monitoring technology is still in the early stages of theoretical research, lacking the necessary informationized and networked monitoring system.

The rapid development of digital image technology has led to an increasing interest in structure monitoring and detection based on computer vision [8]. This interest has been driven by the development of monitoring images and digital image processing technology, which have led to the proposal of a method for characterizing the structural location of feature point displacement. This method employs high-precision displacement meters, with a relative error of less than 3%, to achieve mm level displacement monitoring and the accurate characterization of the structure corresponding to the feature point displacement change [9]. In addition, a dynamic measurement method based on computer vision has the advantages of high measurement accuracy, long distance, low cost, and so forth [10]. However, the existing DIC detection method is limited by processor performance, image resolution, and so forth, making it difficult for it to be applied to engineering inspection [11].

In recent years, the development of CNNs has brought new technological breakthroughs in image classification, target recognition, and other fields. In recent years, CNNs have been successfully applied to the detection of structures. Deng et al. [12] proposed a region-based CNN, which has a deformable module for visual classification of concrete cracks. Fan [13] and others proposed to apply a CNN to pavement crack detection, extract small patches from crack images, generate large training data sets, and then use the CNN to train the data to learn the structure of cracks from the image. Ref. [14] proposed a fully convolutional network for structural crack identification, and used the crack image of an indoor concrete beam test to verify its structural crack identification ability.

Based on the image recognition function of a CNN, this paper extracts the features of two structural experimental images with slight changes, identifies the displacement difference between the two images, obtains the displacement cloud map, and sets up the data set. The data set was enhanced by data normalization and region division. Through the analysis of the number of network layers and learning rate, the model design of the deep learning network was carried out. The designed model was used for training and testing, and the identified displacement was compared with the DIC detection results of the sound

barrier loading test. The results show that the structural displacement identification method based on CNN and DIC technology can accurately identify the displacement change in the structure and greatly improve the efficiency of image displacement calculation using DIC technology. Compared with other methods, the detection method proposed in this paper is more efficient, less error-prone, and more adaptable, and provides a new method for the calculation of structural surface displacement in testing.

## 2. Displacement Calculation Method Based on CNN and DIC

### *2.1. Displacement Field Calculation of Digital Image Method*

The DIC method employs the use of randomly distributed scattering spots [15], both natural and artificial, on the surface of an object. These spots are utilized to identify the most similar subset within the scattering image, which is then used to compute the displacement and strain at each point of interest (POI) within the region of interest (ROI) [16]. The fundamental concept is to ascertain the deformation of an object's surface by comparing the feature points in the scattering image, thereby enabling the measurement of displacement and strain.

The fundamental principle of the DIC method is illustrated in Figure 1. A reference subregion was established in the reference image prior to the deformation of the object, with the search point $p(x_0, y_0)$ serving as the center. The target subregion that most closely resembles the reference subregion was identified through a correlation function in the target image, with the coordinates of the center position of $p'(x'_0, y'_0)$ serving as the reference. In order to compute the requisite deformation variables, the gray values of the pixels in the target subregion were compared with those of the pixels in the reference subregion. The similarity between the reference and target subregions is typically described by the normalized least squares distance function (ZNSSD) [17], defined as follows:

$$C(p*) = \left\{ \sum_{(i,j)\in S} \frac{f(x_i, y_j) - \overline{f}}{\sqrt{\sum_{(i,j)\in S} \left[ f(x_i, y_j) - \overline{f} \right]^2}} - \frac{g(x'_i, y'_j) - \overline{g}}{\sqrt{\sum_{(i,j)\in S} \left[ g(x'_i, y'_j) - \overline{g} \right]^2}} \right\} \tag{1}$$

where $f(x_i, y_j)$ and $g(x'_i, y'_j)$ are the gray values of the point $Q(x_i, y_j)$ in the reference subdomain and the point $Q'(x'_i, y'_j)$ corresponding to point $Q(x_i, y_j)$ in the target subdomain, respectively; $p*$ is the deformation vector describing the relationship between $Q(x_i, y_j)$ and $Q'(x'_i, y'_j)$; $S$ is the set of all the points in the subdomain; $\overline{f} = \frac{\sum_{(i,j)\in S} f(x_i, y_j)}{n*}$ and $\overline{g} = \frac{\sum_{(i,j)\in S} g(x'_i, y'_j)}{n*}$ are the average values of the gray values in the reference subdomain and the target subdomain, respectively; and $n*$ is the number of points in the set $S$. The points in the reference subdomain are mapped to the points in the target subdomain by the corresponding equation. The point K in the reference subdomain is mapped to the point L in the target subdomain, and the corresponding equation is

$$\begin{cases} x'_i = x_i + u + \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y \\ y'_i = y_j + v + \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y \end{cases} \tag{2}$$

The displacement components along the horizontal and vertical directions from the center point $p(x_0, y_0)$ of the reference subdomain are designated as $u$ and $v$, respectively. The four displacement gradient components of the subdomain are designated as $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial x}$, and $\frac{\partial v}{\partial y}$. The distances from point $Q(x_i, y_j)$ to the center point $p(x_0, y_0)$ of the reference

subdomain along the horizontal and vertical directions are designated as $\Delta x$ and $\Delta y$, respectively. In order to determine the extremum of the ZNSSD, it is typical to employ the first-order shape function, which is capable of satisfying the requisite conditions. This function is designated as $p* = \left[uv\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\frac{\partial v}{\partial y}\right]^{\mathrm{T}}$. The parameters associated with $p*$ are typically determined through the use of the Newton iterative method (FANR) or the inverse combinatorial Gaussian Newton method (ICGN). By employing the point-by-point moving least squares fitting method [18], using the quintuple-spline function as the weight function, selecting a rectangular support area, and assuming that the node support radius is R, the following can be obtained:

$$J_{MLS} = \sum_{I}^{N} \omega(x - x_I, y - y_I) \times \left[u^h_{MLS}(x_I, y_I) - \hat{u}(x_I, y_I)\right]^2 \tag{3}$$

$$\omega(\mathrm{s}) = \begin{cases} 1 - 10s^3 + 15s^4 - 6s^5, s \le 1 \\ 0 \qquad\qquad\qquad\quad s > 1 \end{cases} \tag{4}$$

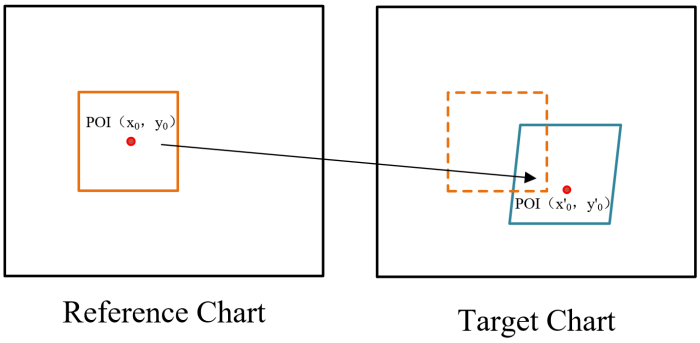



**Figure 1.** Schematic representation of the principle of the DIC method.

The displacement value, $u^h_{MLS}(x_I, y_I)$, is obtained through the point-by-point moving least squares fitting method. The weight function, $\omega(x - x_I, y - y_I)$, represents the point of interest in the support domain of the computation point, $(x, y)$. The approximate solution of the displacement field is derived from Equation (3).

$$u^h_{MLS}(x, y) = p^{\mathrm{T}}a = p^{\mathrm{T}}A^{-1}B\hat{u} \tag{5}$$

The location of the aforementioned phenomenon is as follows: $a$ is the coefficient of the basis function $p$; $\hat{u}$ is the displacement field calculated by digital image correlation (DIC).

$$p = \left[1\ xy\ x^2\ xy\ y^2\ \cdots\ xy^{p-1}\ y^p\right]^{\mathrm{T}} \tag{6}$$

$$A = p^{\mathrm{T}}Wp \tag{7}$$

$$B = p^{\mathrm{T}}W \tag{8}$$

$$\mathrm{W} = \begin{bmatrix} \omega(x - x_1, y - y_1) & 0 & \cdots & 0 \\ 0 & \omega(x - x_2, y - y_2) & \cdots & 0 \\ \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \omega(x - x_N, y - y_N) \end{bmatrix} \tag{9}$$

The strain at the detected point locations can be obtained by taking the derivative of Equation (5) and performing a moving least squares fit point by point for each data point, thereby obtaining the displacement field of the data point through Equation (5).

### *2.2. CNN-Based Structural Displacement Identification Method*

A CNN is usually composed of a convolution layer, an activation layer, a pooling layer, and a fully connected layer. These layers extract image features, introduce nonlinearity, and reduce the feature dimension to improve the generalization ability and recognition performance of the neural network. The parameters of the convolutional layer mainly include filter size, stride, and pad, which together determine the size of the feature map output by the convolutional layer. When the size of the receptive field is smaller than the size of the input image, the larger the size of the receptive field, the more complex are the features of the image that can be extracted. The step size determines the distance between the adjacent regions of the image scanned by the filter during operation. The function of boundary filling is to fill around the feature map, so as not to lose the edge information of the output feature.

As shown in Figure 2, when an image with a size of 5 × 5 is entered, pad = 1, and after boundary filling, the image size is 7 × 7. The convolution operation is performed using a filter with a size of 3 × 3. The entire convolution process can be seen as the filter sliding sequentially on the input image, and the sliding step size is 2. The specific performance is to start from the upper left corner of the image, slide from left to right, and perform convolution operations at the same time. When one row is processed, the process moves on to the next row for the same operation until the entire image area is covered. After the end, a total of nine convolution results are obtained, and a 3 × 3 matrix is output, which is called a feature map. The expression is as follows:

$$f(X) = \sum_{i,j}^{n} \theta_{i,j} \times X_{i,j} + b \tag{10}$$

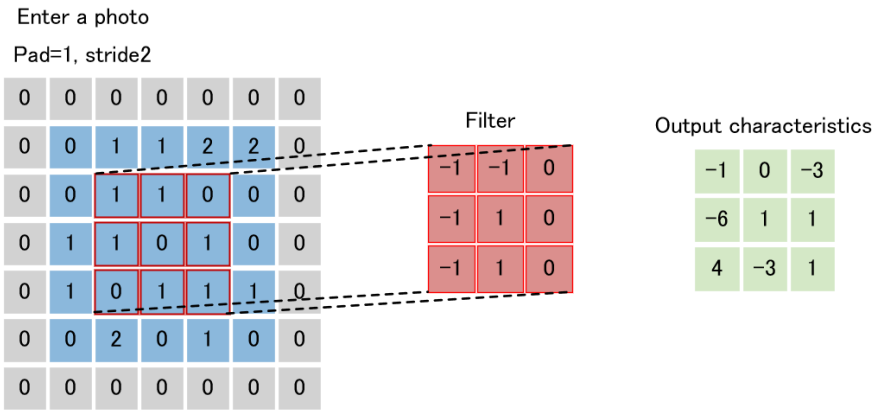



**Figure 2.** Convolution operation diagram.

Here, $f$ $(X)$ represents the output feature, $\theta_{i,j}$ represents the size of the convolution kernel element in column $j$ of row $i$, $X_{i,j}$, $j$ represents the size of column $j$ of row $i$, and b is the deviation.

The image features obtained by the over-convolution operation are linear, but the features of the actual samples are often nonlinear, so a nonlinear function is introduced to solve this problem. Generally, the introduced nonlinear function is called the activation function, which can use any value from 0 to 1 to represent all the pixels of the image, making the simulation process of the image more subtle. The activation function generally has the characteristics of monotonicity, nonlinearity, and continuous differentiability. Pooling the features of the upper layer through the pooling layer can effectively reduce the dimension of the features, and can also reduce the network parameters, thereby improving the problem of

network over-fitting. At the same time, the pooling layer can also maintain some invariance (such as rotation, translation, scaling, etc.). The role of the pooling layer is to compress the feature map, extract the main features, and simplify the complexity of network calculation.

The fully connected layer is located at the end of the CNN structure, which gives the final classification or regression results. In the operation of the fully connected layer, the output feature map will give up its spatial structure and expand into a feature vector. The final output is obtained by the nonlinear combination of the features identified in the previous level. The formula is expressed as follows:

$$f(X) = W \times X + b \tag{11}$$

Based on the image recognition function of the CNN, feature extraction is performed on two structural experimental images with slight changes, and the displacement difference between the two images is identified to obtain the displacement cloud map. Firstly, DIC technology is used to calculate the displacement value between the images during the structural experiment as a label. The CNN is used to train and test the images taken during the experiment. Then, the displacement identified by the CNN is compared with the DIC results to determine the accuracy of the method. The process is as follows: (1) The image is taken from the structural experiment process and recorded as the original image in the data set. The original image is processed by DIC technology to obtain the displacement of the structure, which is recorded as a label in the data set. (2) The data set is enhanced; that is, the pixel value and displacement label value of the image are adjusted to the interval of [0, 1], and the image is cut into a small patch area of 32 × 32 pixels. (3) Different neural network models are trained with the same data set. By comparing the RMSE of the validation set, the optimal CNN model is selected and the hyperparameters of the model are confirmed. (4) The data set is input, and the optimal CNN model is used to read the image for the displacement recognition test. The flow chart of the method is shown in Figures 2–5. The image before the structural deformation is selected as the initial image, and the deformed image is compared with the initial image in turn. It is represented by $X_i$ ($i$ = 1, 2. . .N), where $X_i$ represents the $i$-th deformed image compared with the initial image, and N represents the total number of deformed images. At the same time, the horizontal displacement obtained by the DIC algorithm is represented by $Y_i$ (i = 1, 2. . .N), and Y is the training label. Then, $X$ and $Y$ are used as input data, and the network parameters are trained according to the value of label $Y$ to continuously approach the label value.

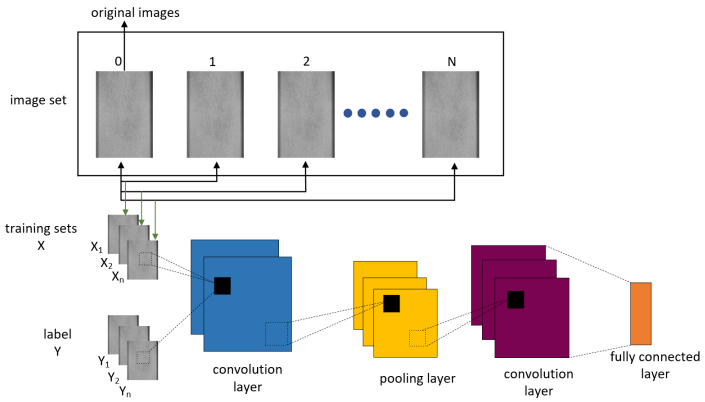



**Figure 3.** Flow chart of CNN displacement identification method.

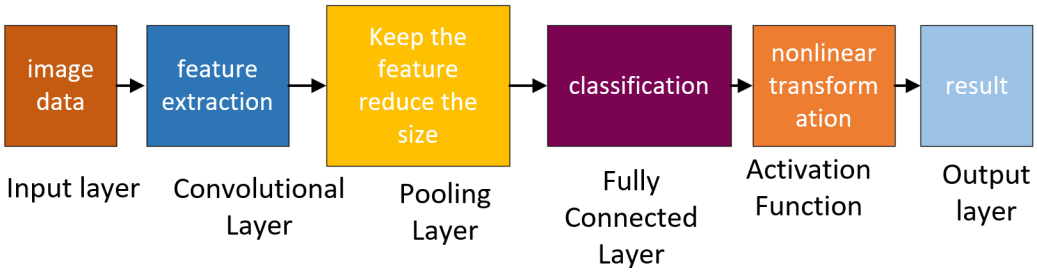



**Figure 4.** CNN architecture diagram.

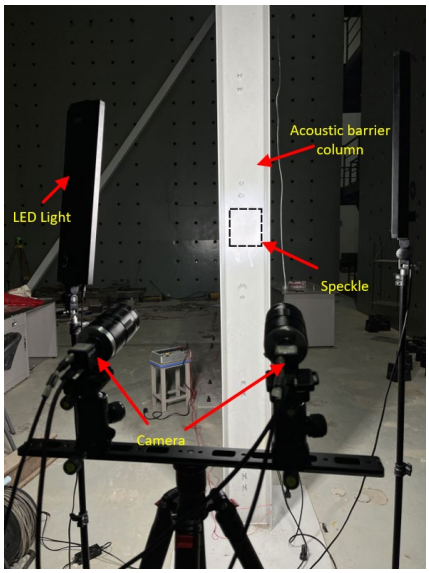



**Figure 5.** XTDIC 3D full-field strain measurement and analysis system.

Specifically, the image data are input into the convolutional layer to extract features through the filter. The next step is to enter the pooling layer to maintain important features and reduce the amount of calculation, and then enter the fully connected layer for classification or regression. The activation function is used for nonlinear transformation, and finally the prediction result is output.

## 3. Simulation Test Analysis

The top of the sound barrier column was loaded with a pulley group to simulate the stress of the sound barrier under different wind pressures. The camera was used to collect the image information of the specimen from the loading to the large deformation. The overall deformation trend and stress distribution of the sound barrier after wind pressure loading were analyzed by data.

The sound barrier is a vertical sound barrier, which is designed as a detachable metal sound barrier. The sound barrier is 6 m high and 4 m long. The column is H-shaped steel, and the material is Q235B. The H-shaped steel column is of the standard size, 250 mm × 125 mm × 6 mm × 9 mm, the column height is 6 m, and the two columns are separated by 2 m; a total of two spans. The column is welded onto the bottom plate with dimensions of 750 mm × 750 mm × 8 mm. The screen plate is made of 5A03 aluminum alloy, 2 m wide and 0.5 m high, and the upper end is fixed with nuts. The thickness of the sound barrier material is 100 mm, and the screen body is composed of a 1.2 mm galvanized plate + 100 mm of sound-absorbing cotton + a 0.8 mm louvered orifice plate. The model is simplified. The screen plate is established as a 100 mm aluminum alloy solid plate with a

perforation rate of 50%, and the bottom plate of the column is fixed with four bolts. The structural material parameters are shown in Table 1.

**Table 1.** Sound Barrier Material Parameters.

| Component | Material | Density (kg/m$^3$) | Young's Modulus (Gpa) | Poisson's Ratio |
|---|---|---|---|---|
| Column | Q235B | 7830 | 210 | 0.274 |
| Screen plate | Aluminum alloy | 2680 | 70 | 0.33 |

### 3.1. Main Equipment

The XTDIC three-dimensional full-field strain measurement and analysis system was used to detect the image of the noise barrier column at the speckle level. The frame rate of the camera is 75 fps, the strain measurement accuracy is less than or equal to 50 με, the measurement range is 400 mm × 300 mm, and the distance is 0 to 10 m. Two charge-coupled element sensor (CCD) cameras were used for shooting, and DIC technology was used to analyze the deformation of the structure. These two digital cameras are equipped with a manually controlled aperture, focusing, and zooming 50 mm macro lens, which can capture images with an area of about 200 mm × 200 mm. The digital camera has a resolution of 5 million pixels and provides 256 levels of grayscale output. The calculation resolution of the DIC program is 0.01 pixels. The two cameras are located 1.0 m in front of the loading specimen. At the same time, two high-power LED lights are set next to the cameras to provide a light source sufficient for the shooting process. During the experiment, two cameras were used to take 20 photos per second throughout the whole process. The collected images were stored in the system and then analyzed using DIC software (2020, v4.1.0).

### 3.2. Test Content

The pulley group was used to convert the gravity of the mass block into a lateral force, because the screen plate is relatively thin and the deformation of the concentrated force is large. In the actual process of the sound barrier being subjected to wind pressure loading, the screen plate was transferred to the column and the column was pulled. The deformation of the column was directly simulated in this experiment. At the same time, the DIC instrument was used to detect the deformation trend of the column during the loading process of the sound barrier. The receiving line of the strain gauge was summarized in the strain detection system, and the weight was increased step by step (Figure 6).



**Figure 6.** Sound Barrier Static Load Test Diagram.

The DIC detection system calculates the displacement and strain of each point of interest (POI) in the region of interest (ROI) based on the randomly distributed speckles of the structure, and sprays speckles on the region of interest of the column.

The equivalent pulley group load was applied to the three positions at the top of the noise barrier column, and the weights were added step by step to load the specimen, and the displacement and stress data of the noise barrier after loading were recorded.

In order to verify the calculation accuracy of the DIC technology, the column displacement measured by the displacement sensor at a height of 1.6 m on the sound barrier column was compared with the displacement value detected by the DIC. The displacement sensor uses an LVDT sensor; using the principle of electromagnetic induction, the displacement was measured according to the relative motion between the coil and the iron core (Figure 7).

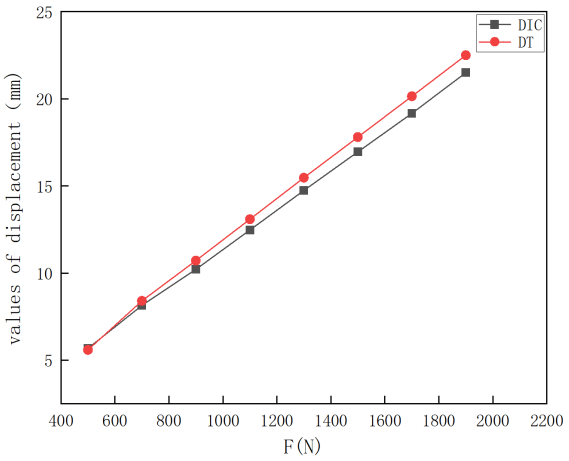



**Figure 7.** Comparison of column displacement values detected by DT and DIC.

It can be seen from the diagram that there is a small deviation in the measurement results of the two methods; it may be that the displacement sensor usually measures the displacement of a specific position, while DIC calculates the displacement on the whole surface. Because there may be errors in the measurement points of the two methods, this may lead to differences between the local deformation and overall deformation, especially for large deformations.

The original image resolution of the column loading test of the sound barrier was 1833 × 884 pixels, and the corresponding label was the displacement cloud map of the column surface calculated by DIC. Firstly, a region of interest (ROI) was set for the original image of the wood. The ROI area is the key area of image analysis. The size of the ROI area of the column is 200 × 186 pixels. Then, the ROI area is divided into a small patch area with a 32 × 32 pixel resolution. Then, the data set is normalized, that is, the image pixel value is adjusted from the [0, 255] interval to the [0, 1] interval; similarly, the displacement cloud map label (X-direction displacement) is also normalized to adjust the displacement value to the [0, 1] interval. In this way, the matrix dimension can be reduced, the network operation speed can be greatly improved, and the gradient information can be retained.

## 4. Displacement Identification Model

Five models are selected according to the number of network layers, and the best model is selected by comparing the training results of the model. The effectiveness of the model is also controlled by hyperparameters learning rate and activation function. The learning rate and activation function are determined according to the actual column displacement detection requirements.

### *4.1. Network Model*

The network model plays a key role in the recognition effect of structural displacement development, and the most influential factor for the network is the number of layers of the model. Therefore, this section selects five models according to the number of network layers—the LeNet, AlexNet, VGG 16, VGG 19, and GoogLeNet models—and selects the best model by comparing the training results of the five models. RMSE is a common index to measure the difference between the predicted results of a model and the actual results. It quantifies the accuracy of the prediction model by calculating the error between the predicted value and the real value, and the square and average and square of these errors. The smaller the RMSE, the more accurate the prediction of the model. The unit is the same as the original data:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

where $y_i$ is the true value, $\hat{y}_i$ is the predicted value, i is the data points corresponding to the serial number, n is the total number of data points. The data as shown in Table 2 are calculated.

Upon analysis of Figure 8 after comparing the five different network models, LeNet, AlexNet, VGG16, VGG19, and GoogLeNet, we found that RMSE showed a certain decreasing trend with the increase in network layers, and the training time also increased accordingly. This shows that although a deeper network model may have stronger expression ability, there may be overfitting or computing resource bottlenecks during the actual training process, resulting in an increased error and a prolonged training time. Specifically:

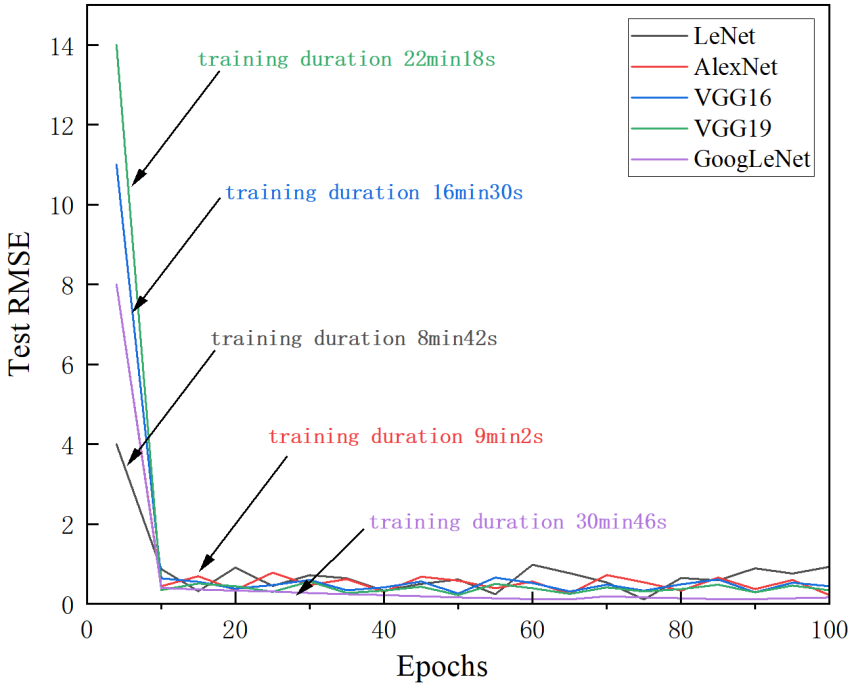



**Figure 8.** RMSE of test set for different model training.

Although LeNet has the least number of layers, it shows lower training error and faster training speed, which is suitable for simpler tasks. AlexNet introduces more convolutional layers, which can achieve better results in complex tasks. The deeper network structures of VGG16 and VGG19 give them a stronger feature extraction ability when dealing with high-dimensional data, but the training time is significantly increased, and, in some cases, RMSE exhibits a slight increase. By introducing the Inception module, GoogLeNet balances

network depth and computational complexity, and can maintain a high training efficiency while ensuring a low RMSE. In summary, GoogLeNet provides a lower RMSE and a longer training time, but it can maintain a stable lower RMSE. In the deformation detection of the column, the stability of the detection data and the lower error are the most important factors. Therefore, GoogLeNet was considered to be the most suitable model.

**Table 2.** RMSE of validation set and test set of different networks.

| Network Class | Network Layers | Validation RMSE | Test RMSE |
|---|---|---|---|
| LeNet | 7 | 0.33 | 0.66 |
| AlexNet | 8 | 0.21 | 0.38 |
| VGG16 | 16 | 0.24 | 0.45 |
| VGG19 | 19 | 0.28 | 0.56 |
| GoogLeNet | 22 | 0.15 | 0.29 |

### *4.2. Activation Function*

In order to select the optimal activation function, five functions, Maxout, Tanh, Sigmoid, ReLU, and Leaky ReLU, were introduced to retrain GoogLeNet. The results are shown in Table 3, and Figure 8 shows the change in the test set RMSE of the network under different activation functions during the training process.

It can be seen from Table 3 and Figure 9 that the test set error of the Leaky ReLU activation function is the smallest, and the error of ReLU is slightly higher than that of the Leaky ReLU function. Due to the saturation characteristics of the Sigmoid and Tanh functions, there are more loss characteristics, so that the error of these two functions is significantly higher than that of the Leaky ReLU function. In summary, the Leaky ReLU function with the smallest error was selected as the activation function of the network.

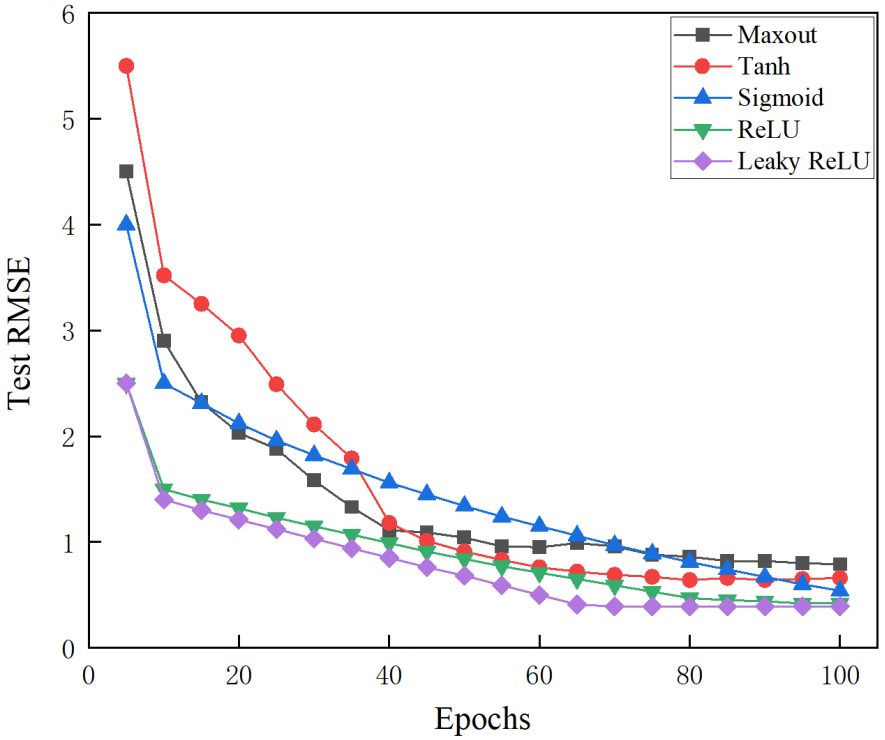



**Figure 9.** The training process of different activation functions.

**Table 3.** RMSE of validation set and test set of different activation functions.

| Activation Function | Validation RMSE | Test RMSE |
|---|---|---|
| Maxout | 0.36 | 0.79 |
| Tanh | 0.28 | 0.66 |
| Sigmoid | 0.42 | 0.54 |
| ReLU | 0.25 | 0.42 |
| Leaky ReLU | 0.19 | 0.39 |

### *4.3. Learning Rate*

The training process of the neural network is essentially a 'step-by-step' search process to find the optimal weight to minimize the loss function, where the training gradient depends on the learning rate. In the early stage of iterative optimization, when the learning rate is larger, the step length will be longer, and then the gradient can be decreased at a faster speed. In the later stage of iterative optimization, the value of the learning rate and the step length will gradually be reduced, which will help the convergence of the algorithm and make it easier to approach the optimal solution.

In order to determine the optimal learning rate, four different learning rates in the range of 0.0005–0.01 were selected to retrain GoogLeNet, and the results are shown in Table 4.

**Table 4.** RMSE of test sets with different learning rates.

| Learning Rate | (Test RMSE) | (Epochs) |
|---|---|---|
| 0.005 | 0.46 | 53 |
| 0.01 | 0.44 | 47 |
| 0.015 | 0.42 | 61 |
| 0.02 | 0.41 | 52 |
| 0.025 | 0.39 | 67 |
| 0.03 | 0.38 | 62 |
| 0.04 | 0.36 | 73 |
| 0.05 | 0.34 | 72 |
| 0.075 | 0.35 | 66 |
| 0.1 | 0.37 | 59 |

If the learning rate is too small, the parameter update rate of the model will be very small, which may lead to the training process becoming slow, and it may even not be able to fully learn the pattern of the data within a limited number of training rounds, resulting in underfitting. In this case, the model may not achieve optimal performance. If the learning rate is too large, the step size of each update of the model will be too large and may skip the best solution, resulting in oscillation or divergence, so that the ideal optimal solution cannot be found, and this may even lead to overfitting or training failure. When the learning rate is within a certain range, the parameter update is neither too small nor too large. The model can effectively find the law of the training data and maintain a good generalization ability on the test data. Therefore, when the learning rate is 0.05 in this training set, it performs best in terms of the test set RMSE.

## 5. Analysis of Recognition Results

The experimental environment is as follows: Intel (R) i9-13400KCPU @ 3.60GHz processor, RAM32GB. The programming language used in this paper is MATLABR2021 b, and Deep Network Designer is used to build a convolutional neural network. The network model is GoogLeNet. The Batch Size of the training set and the verification set is 128. The learning rate adopts the piecewise constant attenuation method. The initial learning rate

is 0.05, the learning rate drop factor is 0.8, and the learning rate drop cycle is 20 rounds. The evaluation index can calculate the similarity between the displacement cloud map identified by the CNN and the displacement cloud map calculated by DIC, so as to judge the accuracy of the CNN in identifying structural displacement.

The evaluation indexes selected in this paper are: Mean Square Error (MSE), Structural Similarity Index (SSIM) and Correlation (C), Peak Signal to Noise Ratio (PSNR), and Structural Content (SC). MSE calculates the amplitude of the average error of the two images, and then determines the degree of distortion of the two images by the size of the error. The calculation method of MSE is as follows:

$$\mathrm{MES} = \frac{1}{[R \times C]^2} \sum_{i=0}^{n} \sum_{i=0}^{n} \left(X_{ij} - Y_{ij}\right)^2 \tag{13}$$

$R$ and $C$ are the number of rows and columns of the image pixel matrix, $X_{ij}$ is the image pixel value from the CNN recognition result, and $Y_{ij}$ is the image pixel value of the DIC calculation result.

SSIM can measure the structural similarity between two images, ranging from $-1$ to 1. When the two images are almost the same, the SSIM is close to 1. The following formula is used to calculate the SSIM between the given two image pixels $\mathrm{seq}_1$ and $\mathrm{seq}_2$, where seq denotes an array of all the pixel values in an image arranged in the order

$$\mathrm{SSIM} = \frac{(2 \times \mathrm{mu}_1\,(\mathrm{p})\mathrm{mu}_2\,(\mathrm{p}) + \mathrm{c}_1)(2 \times \mathrm{cov}(\mathrm{p}) + \mathrm{c}_2)}{(\mathrm{mu}_1\,(\mathrm{p})^2 + \mathrm{mu}_2\,(\mathrm{p})^2 + \mathrm{c}_2)(\mathrm{s}_1\,(\mathrm{p})^2 + \mathrm{s}_2\,(\mathrm{p})^2 + \mathrm{c}_2)\Big)} \tag{14}$$

Among them, $\mathrm{mu}_1$ (p) and $\mathrm{mu}_2$ (p) are the pixel means of $\mathrm{seq}_1$ and $\mathrm{seq}_2$, respectively, $\mathrm{mu}_1$ $(\mathrm{p})^2$ and $\mathrm{mu}_2$ $(\mathrm{p})^2$ are the pixel standard deviations of $\mathrm{seq}_1$ and $\mathrm{seq}_2$, respectively, and cov (p) is the covariance of $\mathrm{seq}_1$ and $\mathrm{seq}_2$. $\mathrm{c}_1 = (\mathrm{K}_1 \times \mathrm{L})^2$ is a regularization constant, $\mathrm{c}_2 = (\mathrm{K}_2 \times \mathrm{L})^2$ is a regularization constant, $\mathrm{K}_1$ and $\mathrm{K}_2$ are regularization parameters (must be > 0), and L is the dynamic range of pixel values.

$C$ determines the similarity or different degree of two signals or vectors in phase and amplitude. The larger the correlation coefficient is, the higher the matching degree of the two images is, and the closer the matching result is. The calculation formula is as follows:

$$C = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \tag{15}$$

where $n$ is the number of data pairs, $x$ is the image pixel value of the CNN recognition results, and $y$ is the image pixel value of the DIC calculation results.

PSNR can calculate the imperceptibility of the image. The higher the PSNR value of the image, the higher the quality of the image, and the lower the image distortion. PSNR is calculated according to the following formula:

$$\mathrm{PSNR} = 10 \times \log_{10}\left[\frac{\mathrm{I}^2}{\mathrm{MSE}}\right] \tag{16}$$

where I represents the maximum possible value of the pixel in the image (for example, the maximum value of the gray image is 255), and MSE is the mean square error.

*SC* is a measurement based on image correlation, which can measure the similarity between two images. The closer the *SC* value is to 1, the higher the image similarity is. The calculation formula is as follows:

$$SC = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (y(i,j))^2}{\sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j))^2} \tag{17}$$

where $x$ $(i, j)$ represents the image pixel value of the recognition result, and $y$ $(i, j)$ represents the image pixel value of the DIC calculation result.

There are 500 images in the test set of this training process. Next, the overall displacement cloud map of the ROI and the displacement change in the column in the selected height range of 1.4 m–1.6 m in the ROI are analyzed, respectively. According to the comparison of the displacement cloud maps in Figures 10 and 11 it can be seen that the displacement cloud map identified by the CNN is roughly the same as the DIC calculation result, which can clearly show the displacement change process of the structure.

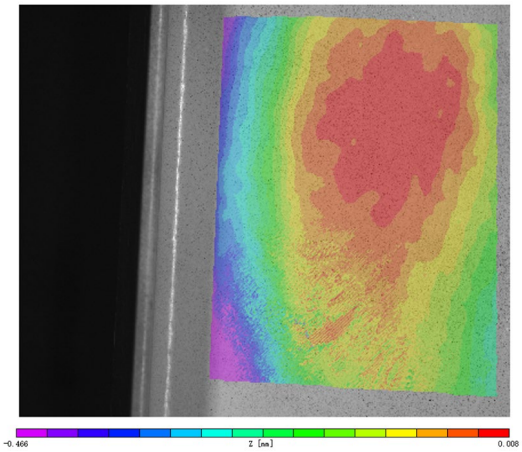


**Figure 10.** Displacement cloud map identified by CNN.

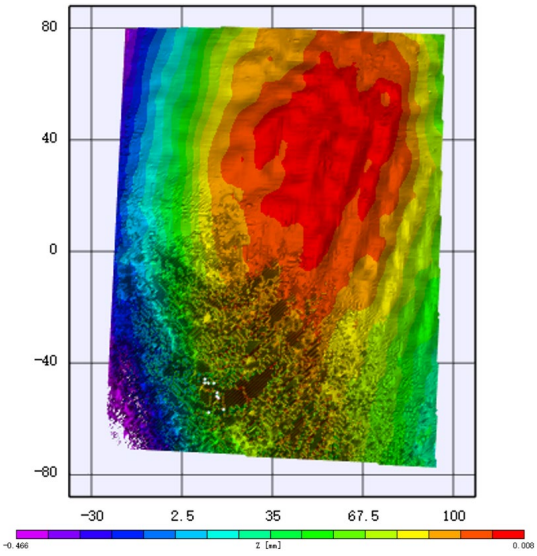



**Figure 11.** Displacement cloud image identified by DIC.

The error of the cloud images under each load calculated by the CNN and DIC is calculated to obtain Table 5.

**Table 5.** Evaluation index of image in loading process.

| F(N) | MSE | SSIM | C | PSNR | SC |
|---|---|---|---|---|---|
| 500 | 0.00141 | 1 | 0.9983 | 94.9887 | 1 |
| 700 | 0.00158 | 1 | 0.9946 | 96.5406 | 1 |
| 900 | 0.00112 | 1 | 0.9924 | 95.9382 | 1 |
| 1100 | 0.00176 | 1 | 0.9917 | 96.8369 | 1 |
| 1300 | 0.00105 | 1 | 0.9962 | 94.9265 | 1 |
| 1500 | 0.00198 | 1 | 0.9893 | 96.3471 | 1 |
| 1700 | 0.00124 | 1 | 0.998 | 94.3514 | 1 |
| 1900 | 0.00167 | 1 | 0.9901 | 96.0523 | 1 |

By comparing the similarity of the displacement cloud images identified by the two methods in Table 5, it was found that the six groups of displacement cloud images a–f have higher SSIM, C, PSNR, SC values and smaller MSE values, which indicates that the method based on the CNN and DIC proposed in this paper can effectively identify the expansion process of the displacement cloud image on the surface of the sound barrier column.

Five feature points were selected from the ROI area of the column, which were recorded as $P_1$~$P_5$, as shown in Figure 12. The abscissa of the five points were the same, and the ordinate interval was 10 mm.

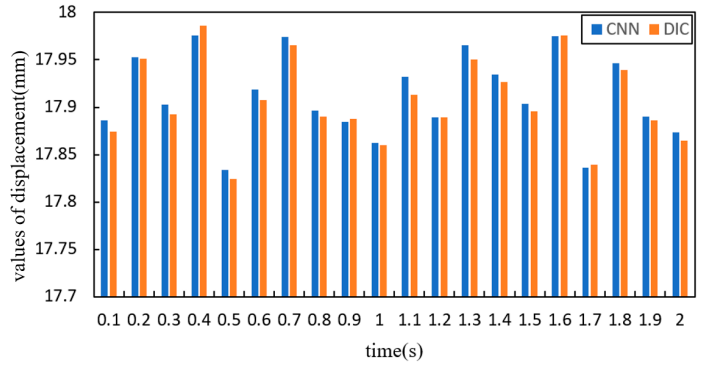



**Figure 12.** Feature point diagram.

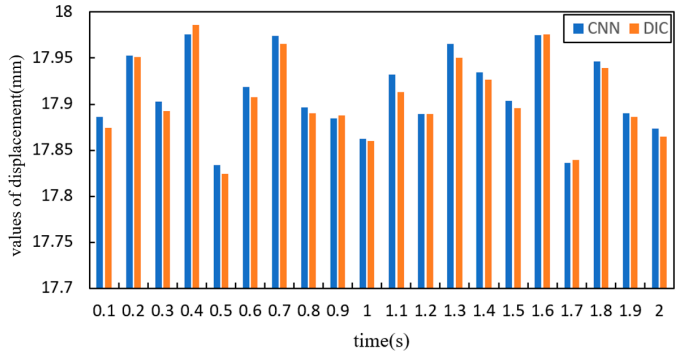
In the case of a full load of 1900 N, the displacement changes of these five points in the CNN recognition results are compared with the DIC calculation results within 2 s. The displacement changes measured by the two methods during the loading of the noise barrier column are shown in Figure 13.

It can be seen from the displacement changes of the five feature points in the figure that the point displacement identified by the convolutional neural network (CNN) is consistent with the change trend of the point displacement calculated by the digital image correlation (DIC) technology. The displacement curves of the two show a similar shape, indicating that the CNN-based recognition method can effectively capture the displacement mode matching the DIC calculation method. Specifically, by learning the displacement information of different feature points in the image, the CNN model can extract the corresponding spatial displacement features from the input image data and form a consistent trend with the DIC results. Further analysis shows that the error between the two is controlled within 5%, which indicates that the CNN method has reached a high level in terms of accuracy and can reliably identify the displacement change process of a certain point of the sound barrier column. Especially when dealing with large-scale or high-dimensional data, the

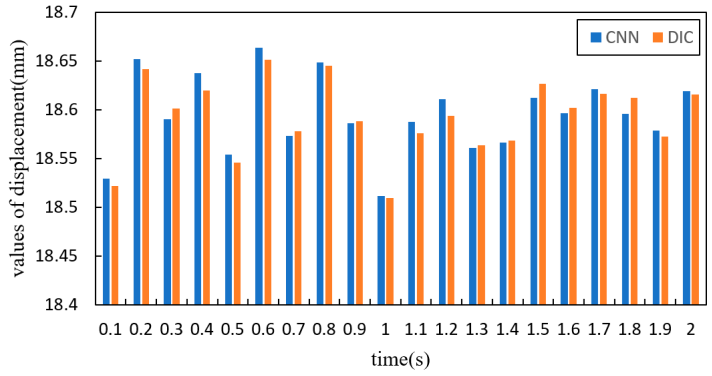
CNN has a higher computational efficiency and stronger generalization ability than the traditional calibration-based DIC method.



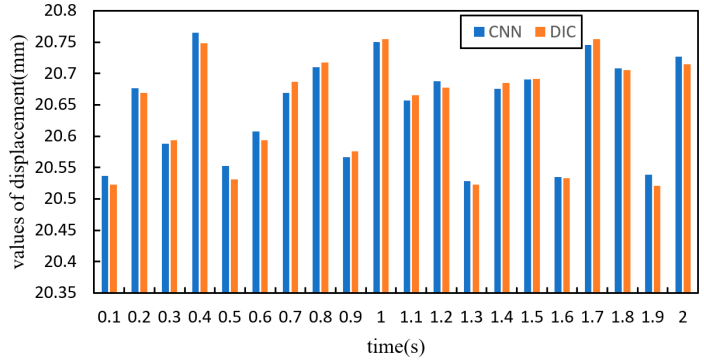

(**a**)




(**b**)

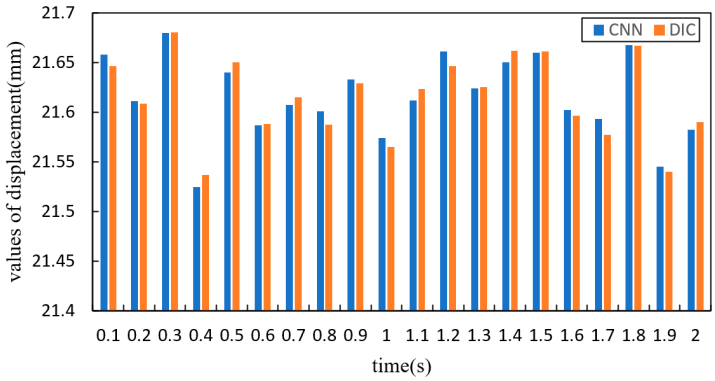



(**c**)




(**d**)

**Figure 13.** *Cont*.




(**e**)

**Figure 13.** A comparison of the CNN and DIC detection data. (**a**) A comparison of the two data sets at point $P_1$; (**b**) a comparison of the two data sets at point $P_2$; (**c**) a comparison of the two data sets at point $P_3$; (**d**) a comparison of the two data sets at point $P_4$; (**e**) a comparison of the two data sets at point $P_5$.

## 6. Simulation Model Verification

The deformation of the column obtained by the test may be accidental, and the test was simulated. The static strain testing system was employed to identify the position of the detection point in the process of gradually increasing the strain value. This was accomplished by integrating the modulus of elasticity of the material with the incremental increase in the strain value. The resulting stress value was then employed to determine the position of the column detection point, as illustrated in Table 6.

**Table 6.** Static strain detection system measurement data.

| Concentration Force (N) | Column Strain (με) | Column Stress (Pa) |
|---|---|---|
| 500 | 410 | $8.61 \times 10^7$ |
| 700 | 765 | $1.61 \times 10^8$ |
| 900 | 1030 | $2.16 \times 10^8$ |
| 1100 | 1290 | $2.71 \times 10^8$ |
| 1300 | 1556 | $3.27 \times 10^8$ |
| 1500 | 1805 | $3.79 \times 10^8$ |
| 1700 | 2045 | $4.29 \times 10^8$ |
| 1900 | 2283 | $4.79 \times 10^8$ |

In a similar manner, the noise barrier was subjected to incremental loading in the simulation in order to obtain the stress value at each corresponding point. A comparison between the measured and simulated data is presented in Figure 14, which indicates that the simulated and measured stress data are approximately the same, with a relative error of less than 1% at the detection points. This serves to validate the accuracy of the simulation model.

The output model of the analysis system was selected as Classical Partial Least Squares (PLS) regression through the DIC method. The Classical PLS method was employed to minimize the sum of the squares of deviations between the fitted values of the points of interest within the window and the values calculated by DIC.

$$J_{PLS} = \sum_{I}^{N} \left[ u_{PLS}^{h}(x_1, y_1) - \hat{u}(x_1, y_1) \right]^2 \tag{18}$$

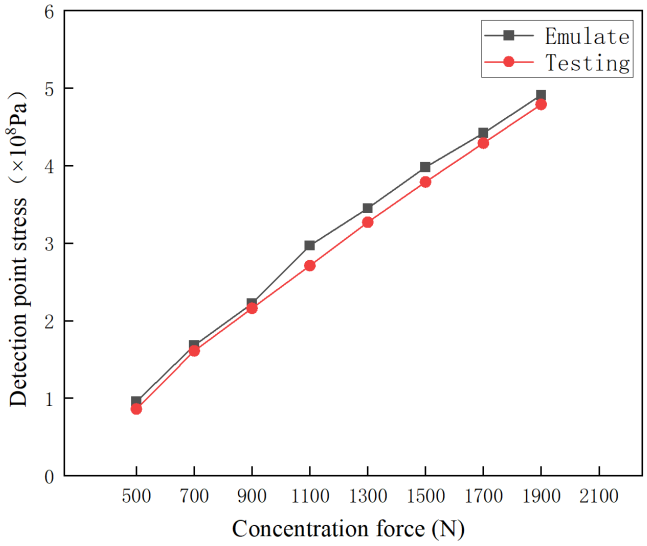



**Figure 14.** A comparison of the stresses at the measured and simulated test points.

The displacement value was calculated by the digital image correlation (DIC) method and subsequently fitted by the partial least squares (PLS) method. Additionally, the total number of points of interest within the window was taken into account. For the fitting of more uniform deformation, the classical PLS method was more accurate. Figure 11 depicts the deformation cloud map of the sound barrier column subjected to an 1800 N load in the detection area of 1. The column, 6 m in height and 20 cm by 20 cm in cross-section, was tested. The digital image correlation (DIC) detection system recorded a total of 409,290 data points, which were subsequently analyzed using the Vic-3D software (2020, v4.1.0). All images of the deformation process were subjected to automated object identification (AOI) and post-processing analysis. The projection error was 0.052. This is due to the fact that a considerable number of offsets were observed on the right side of the column outside the detection area. Consequently, the displacement of the *x*-axis on the left was selected. The deformation of the sound barrier column at a distance of 1.6 m was recorded to be 21.5 mm.

Figure 15 shows the measured displacement cloud diagram of DIC. The deformation of all nodes situated at a distance of 1.6 m from the column in the simulation model, in the direction of the loading force, was calculated in order to determine the maximum deformation of 21.68 mm. A comparison of the DIC inspection data with the modeling data revealed a high degree of agreement between the two.

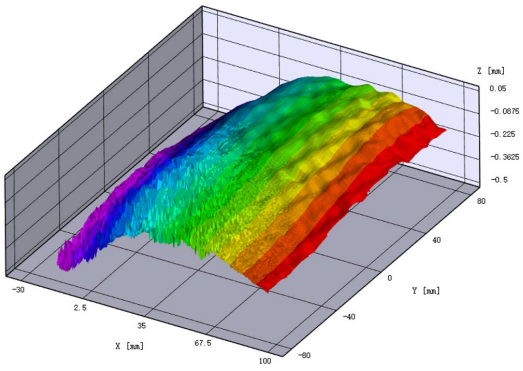



**Figure 15.** The deformation cloud of the column detection region.

*Calculate Working Conditions*

In order to simulate the deformation state of the noise barrier under the action of wind at all levels, the horizontal wind load can be designed and calculated according

to the formula for horizontal wind load, as specified in the relevant provisions of the Standard [19] GB50009-2012 (China; Beijing, 2012) Code for Structural Loading of Buildings. The calculation formula is as follows:

$$W_k = \beta_{gz}\mu_{sl}\mu_Z W_0 \tag{19}$$

Formula: $W_k$—default wind load (kN/m$^2$); $\beta_{gz}$—gust factor at altitude z; $\mu_{sl}$—wind load-localized body shape factor; $\mu_Z$—coefficient of variation of wind pressure height; $W_0$—basic wind pressure (kN/m$^2$).

Figure 16 shows the simulation model. The standard value of the wind load of 0.445 kN/m$^2$ in a class 7 wind with a wind speed of 14.5 m/s was calculated and applied to the sound barrier screen plate in the form of a uniformly distributed force. The maximum value of the stress was found to be 2.45 × 10$^9$ Pa at the bottom bolt, while the maximum value of the displacement was 0.08238 m at the top of the column.

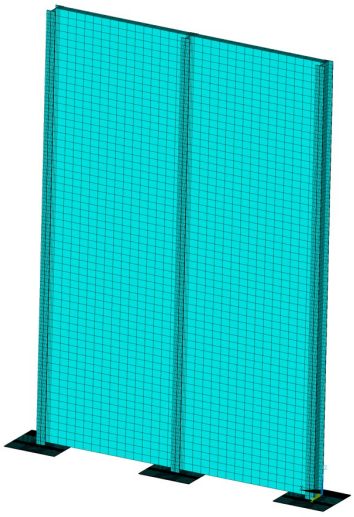


**Figure 16.** Finite element model.

In the test to verify the accuracy of the model, it is challenging to simulate the wind load applied to the sound barrier in the form of uniform load distribution. Consequently, the equivalent analogy method was employed, whereby a concentrated force was applied to the center of the top of the sound barrier column. The maximum stress and maximum displacement of the sound barrier were then compared with those of the sound barrier under the corresponding level 7 wind. It was observed that the two values were reduced to within the permissible error, indicating that the deformation of the wind load applied to the barrier was equal to the deformation of the concentrated force applied to the barrier. The deformation of the wind load on the sound barrier was equivalent to the deformation of the concentrated force on the sound barrier.

Following the completion of the simulation modeling, to obtain the cloud diagram in Figure 17, the calculated cloud diagrams of the sound barrier under the influence of uniform and concentrated forces could be obtained. Following the completion of the numerical analysis, it can be concluded that when the concentrated force was 1800 N, the maximum value of the stress of the sound barrier was 2.46 × 10$^9$ Pa and the maximum value of displacement was 0.083815 m, with an error of less than 1%. Furthermore, the overall cloud diagrams of the model are essentially identical.

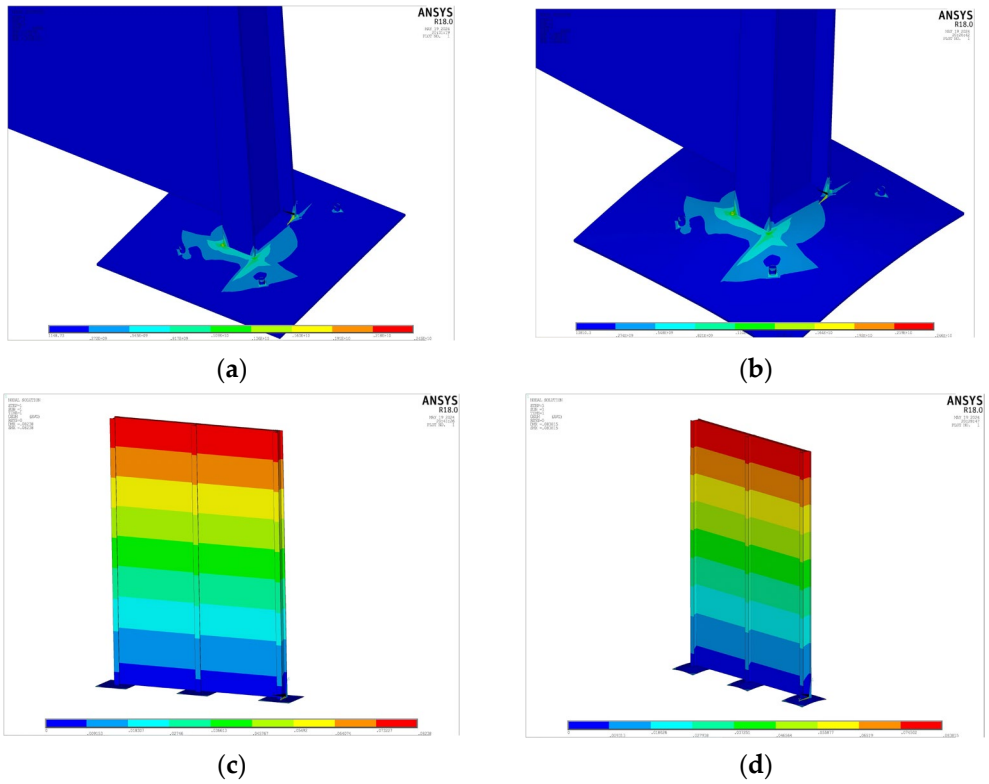



**Figure 17.** Cloud Diagrams Under the Effect of Uniform and Concentrated Forces. (**a**) Stress cloud with uniform force; (**b**) stress cloud under concentrated force; (**c**) displacement cloud under uniform force; (**d**) displacement cloud under concentrated force.

In the same case, the models in different studies were selected to calculate the index parameters of the images for comparative analysis. The data in Table 7 were obtained.

**Table 7.** Numerical comparison of the identification of each model.

| Model | MSE | SSIM | PSNR | C | SC |
|---|---|---|---|---|---|
| GoogLeNet | 0.00125 | 0.985 | 38.92 | 0.981 | 0.998 |
| U-Net | 0.0021 | 0.978 | 37.65 | 0.973 | 0.996 |
| tradition-CNN | 0.00345 | 0.96 | 36.12 | 0.95 | 0.99 |
| ResNet | 0.00185 | 0.982 | 39.1 | 0.975 | 0.997 |
| VGG-16 | 0.00275 | 0.975 | 37.98 | 0.965 | 0.993 |
| CNN + DIC | 0.0015 | 0.988 | 39.25 | 0.983 | 0.999 |

The combination of CNN + DIC has excellent performance for many indicators, especially SSIM, PSNR and SC, indicating that the method effectively preserves the structure and details of the image in the process of image denoising, and the denoising effect is significantly improved compared with the other models. Deep learning models such as GoogLeNet and ResNet also perform very well, but the CNN model combined with DIC is slightly better than the traditional CNN and other network models due to its multi-level image feature extraction and denoising capabilities.

## 7. Application Potential and Challenges

### 7.1. Application Development Prospects

The earliest non-contact monitoring methods mainly rely on physical sensors, such as temperature sensors, displacement sensors, etc., for direct measurement. For example, infrared sensors and laser range finders have been widely used in structural health monitoring and industrial inspection. However, these methods are often limited to a single physical measurement and cannot provide high-precision and high-resolution spatial data.

With the advancement of imaging technology, image-based non-contact monitoring technology has gradually become mainstream. In particular, DIC technology obtains information such as the displacement field and strain field on the surface of the object by photographing and comparing images under multiple different loads. This method has the advantages of being non-destructive and high precision, and can provide complete surface displacement field data [20].

In the application of DIC technology, how to efficiently process a large amount of image data has become a challenge. The traditional method relies on image feature matching and the manual selection of regions for displacement calculation. The operation is complex and susceptible to noise interference, and the efficiency is low. In order to overcome this problem, image processing methods based on machine learning and deep learning have begun to be applied, and the convolutional neural network (CNN) has gradually emerged in this context.

Although traditional DIC technology and CNN-based image processing methods have been applied in the field of structural monitoring, there are still some technical gaps and challenges in practical applications. In this study, a new structural displacement identification method is proposed by combining DIC technology and CNN, which effectively solves the following problems:

The displacement cloud generated by DIC technology provides detailed spatial information for the deformation analysis of the structure [21]. However, the image data obtained by DIC technology is usually affected by environmental noise, shooting angle, illumination changes, and other factors, which will lead to fluctuations in image quality. In order to solve this problem, this study proposes to enhance the data set by data normalization and region division. Through these methods, not only is the stability of the data improved, but also the robustness of the deep learning model to data under different environments and conditions is enhanced [22].

In terms of deep learning model design, this study analyzes the hyperparameters such as CNN network layer number and learning rate in detail, and optimizes the model architecture. By finely adjusting the network structure [23], this study can better extract the features in the displacement image and improve the accuracy of displacement recognition. Compared with traditional image processing methods, the CNN can automatically learn the high-dimensional features in the image and make adaptive adjustments, thereby improving the performance of the model under different test conditions [24].

The traditional displacement calculation based on DIC technology often requires a long calculation time, especially in the case of high-resolution images or complex structures. The CNN-based automated image analysis method proposed in this study significantly improves computational efficiency through an end-to-end learning framework. Compared with traditional manual feature extraction and matching methods, the CNN can quickly process large-scale image data [25], shorten the time of displacement recognition, and improve real-time performance.

Non-contact monitoring methods have gradually developed from traditional sensor measurement to intelligent monitoring systems based on image processing and deep learning. Although the existing technology has made progress in terms of application in

many fields, it still faces problems such as data processing complexity, low computational efficiency, and poor environmental adaptability. The structural displacement identification method based on CNN and DIC technology proposed in this paper overcomes these problems. Through data enhancement, network optimization, and automatic image analysis, the accuracy, efficiency, and real-time performance of displacement identification are significantly improved, which provides a new idea and solution for non-contact monitoring in the fields of structural health monitoring and engineering detection.

### 7.2. Challenges and Limitations

The displacement recognition method based on DIC and a CNN mentioned in this study mainly depends on the acquisition and analysis of image data. In extreme environments [26], image quality may be affected. For example, low temperatures may cause hardware problems for the equipment, and high humidity may increase image noise. Therefore, in practical applications, the input of the system (such as images collected by cameras or other sensors) can be detected and preprocessed in real time in combination with the online monitoring system to ensure that data can still be effectively obtained under extreme conditions. In addition, regarding the problem of equipment failure that may be caused by a high-temperature or high-humidity environment, a fault-tolerant mechanism can be designed to ensure the stable operation of the system. Intelligent algorithms for environmental changes can be further added to monitor and adjust the data processing flow in real time.

For the displacement identification of large structures or complex geometries, a multi-scale modeling method can be used to divide the structure into multiple subregions for local analysis. Each subregion is processed by a local CNN model, and finally the overall displacement result is obtained by the global fusion model. This can not only reduce the amount of calculation, but also effectively improve the scalability when dealing with complex structures [27].

For the case of poor image quality, image enhancement techniques (such as contrast enhancement, brightness adjustment, denoising, etc.) can be combined to improve image quality. In addition, the self-supervised learning method in deep learning can also extract effective information from low-quality images and improve the robustness of the model. In addition to traditional optical image data, consider combining other sensors (such as lidar, infrared imaging, acoustic sensors, etc.) [28] to provide supplementary information. Multi-modal data fusion can provide more comprehensive information by relying on the data of other sensors when the image data quality is poor, and can reduce the instability of a single image source. In order to adapt to image quality changes under different field conditions, an adaptive learning algorithm can be used. Through continuous learning and adaptive adjustment of the model, more accurate recognition results can still be obtained in the case of poor image quality. For example, ref. [29] low-quality images can be generated by generative adversarial networks (GANs) to simulate clear images to train the model, thereby improving the robustness of recognition.

Here are some suggestions for practitioners to use high-resolution industrial cameras, especially where small-size cracks or small deformations require high-precision monitoring. In addition to DIC and image processing, these data can be combined with other sensors (such as accelerometers, strain gauges, laser rangefinders, etc.) for multi-modal data fusion. Using edge computing hardware to process real-time collected data can speed up image processing and analysis, especially in large-scale monitoring tasks. For CNN training, sufficient labeled data sets are crucial. It is recommended to collect and generate a variety of training data sets. Image processing techniques (such as Gamma correction and contrast adjustment) can be used to perturb the input image to increase the diversity of the

training set. Through this method [30], the CNN model can learn more potential damage patterns during training and improve its robustness to different situations. For the task of image recognition, it is recommended to select the appropriate network according to the characteristics of the specific task. Lightweight networks such as Mobile Net and Efficient Net can be used in scenarios with limited computing resources

## 8. Conclusions

Based on computer vision technology, displacement field theory, and a simulation model, this paper proposes a structural displacement identification method based on convolutional neural networks (CNNs) and DIC technology. The data set was formed according to the column displacement cloud map obtained by DIC analysis, and the data set was enhanced by data normalization and region division. Through the analysis of the number of network layers and learning rate, the model design of the deep learning network was carried out. According to the number of network layers, the optimal network model was selected from five neural network models (LeNet, AlexNet, VGG 16, VGG 19, and GoogLeNet). At the same time, after a lot of training, the model was confirmed to be GoogLeNet, and the learning rate parameter was 0.05. The test set error of the Leaky ReLU activation function was the smallest, and the high-speed camera image results of the test were introduced and identified by the static loading test of the equal-scale sound barrier. The simulation results show that the structural displacement identification method based on the CNN and DIC technology can accurately identify the displacement change in the structure. The efficiency of the image displacement calculation using DIC technology is greatly improved, and the error can be controlled within 5%. Monitoring technology with high accuracy can identify potential structural problems and deformations in advance, thus effectively supporting preventive maintenance. Controlling the error within 5% can result in problems being found earlier, reduce sudden failures, optimize maintenance plans, reduce the risk of safety accidents, and save maintenance costs.

**Author Contributions:** The authors confirm their contributions to the paper as follows: study conception and design: M.M. and F.W.; data collection: S.H.; analysis and interpretation of results: X.D.; draft manuscript preparation: Q.H., J.T. and X.H., who also analyzed the simulation data. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Independent research project of the State Key Laboratory, HJGZ2021211; the Science and Technology Project of the Jiangxi Education Department, GJJ2200629; and the Science and Technology Project of the Jiangxi Education Department, Gjj190339.

**Data Availability Statement:** This paper relies on the data obtained by the previous experiments of the research group. Readers can obtain these data from the authors by email if they need to, so as to have a comprehensive understanding of the data and materials used.

**Acknowledgments:** Thank you very much for the support and help of each author.

**Conflicts of Interest:** Authors Fayuan Wu, Mengting Mao, Xiaomin Dai, and Jinhui Tang are employed by the company Jiangxi Electric Power Co., Ltd. Author Sheng Hu is employed by the company State Grid Hunan Electric Power Co., Ltd. Author Qiang He is employed by the company State Grid Smart Grid Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

## References


1. Yuchao, M.; Juan, M.; Meng, N.; Xiang, L. Research on Vibration Characteristics of Urban Substation Building and Transformer. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *371*, 022094. [CrossRef]
2. Wang, L.; Geng, M.; Bai, X.; Ma, J.; Zhao, Y.; Shen, C.; Yang, B. Urban 110 kV indoor substation noise analysis and control schemes: A real case study. *Appl. Acoust.* **2021**, *183*, 108290. [CrossRef] t
3. Wang, D.; Wang, B.; Chen, A. Vehicle-induced aerodynamic loads on highway sound barriers part1: Field experiment. *Wind. Struct.* **2013**, *17*, 435–449. [CrossRef]
4. Zhang, F.; Ji, S.; Shi, Y.; Zhan, C.; Zhu, L. Investigation on vibration source and transmission characteristics in power transformers. *Appl. Acoust.* **2019**, *151*, 99–112. [CrossRef]
5. Ma, Z.X.; Choi, J.; Sohn, H. Real-time structural displacement estimation by fusing asynchronous acceleration and computer vision measurements. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 688–703. [CrossRef]
6. Jastrzębska, M. Modern displacement measuring systems used in geotechnical laboratories: Advantages and disadvantages. *Sensors* **2021**, *21*, 4139. [CrossRef] [PubMed]
7. Wang, X.; Zhang, J.; Li, Z. Development of a real-time sensor system for deformation detection of noise barriers. *Sens. Actuators A Phys.* **2020**, *305*, 111827.
8. Bao, Y.; Tang, Z.; Li, H.; Zhang, Y. Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Struct. Health Monit.* **2019**, *18*, 401–421. [CrossRef]
9. Fukuda, Y.; Feng, M.Q.; Narita, Y.; Kaneko, S.I.; Tanaka, T. Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. *IEEE Sens. J.* **2013**, *13*, 4725–4732. [CrossRef]
10. Ye, X.W.; Dong, C.Z.; Liu, T. A review of machine vision-based structural health monitoring: Methodologies and applications. *J. Sens.* **2016**, *2016*, 7103039. [CrossRef]
11. Romanowicz, P.J.; Szybiński, B.; Wygoda, M. Application of DIC method in the analysis of stress concentration and plastic zone development problems. *Materials* **2020**, *13*, 3460. [CrossRef]
12. Deng, L.; Chu, H.-H.; Shi, P.; Wang, W.; Kong, X. Region-Based CNN Method with Deformable Modules for Visually Classifying Concrete Cracks. *Appl. Sci.* **2020**, *10*, 2528. [CrossRef]
13. Fan, Z.; Wu, Y.; Lu, J.; Li, W. Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network. *arXiv* **2018**, arXiv:1802.02208.
14. Ye, X.W.; Jin, T.; Chen, P.Y. Structural crack detection using deep learning–based fully convolutional networks. *Adv. Struct. Eng.* **2019**, *22*, 3412–3419. [CrossRef]
15. Yamaguchi, I. Speckle displacement and decorrelation in the diffraction and image fields for small object deformation. *Opt. Acta Int. J. Opt.* **1981**, *28*, 1359–1376. [CrossRef]
16. Peters, W.H.; Ranson, W.F. Digital imaging techniques in experimental stress analysis. *Opt. Eng.* **1982**, *21*, 427–431. [CrossRef]
17. Pan, B. Recent progress in digital image correlation. *Exp. Mech.* **2011**, *51*, 1223–1235. [CrossRef]
18. Dong, Y.L.; Pan, B. A Review of Speckle Pattern Fabrication and Assessment for Digital Image Correlation. *Exp. Mech.* **2017**, *57*, 1161–1181. [CrossRef]
19. *GB 50009-2012*; Load Code for the Design of Building Structures. China Architecture & Building Press: Beijing, China, 2012.
20. Baraniello, V.R.; Cicala, M.; Cicala, L. An algorit-hm for real time estimation of the flexible UAV structural motions using a video-based system. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–8.
21. Sun, Y.; Xu, Y.; Wang, X.; Zhu, H. Experimental studies on column foot connections of novel fully enclosed noise barriers. *J. Constr. Steel Res.* **2020**, *172*, 106179. [CrossRef]
22. Li, X.; Fang, G.; Zhao, J.; Zhang, Z.; Wu, X. Local Hermite (LH) Method: An accurate and robust smooth technique for high-gradient strain reconstruction in digital image correlation. *Opt. Lasers Eng.* **2019**, *112*, 26–38. [CrossRef]
23. Zhou, Z.; Shao, S.; Deng, G.; Gao, Y.; Wang, S.; Chu, X. Vision-based modal parameter identification for bridges using a novel holographic visual sensor. *Measurement* **2021**, *179*, 109551. [CrossRef]
24. Xu, Y.; Brownjohn, J.; Kong, D. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct. Control. Health Monit.* **2018**, *25*, e2155. [CrossRef]
25. Xu, Y.; Brownjohn, J.M.; Hester, D.; Koo, K.Y. Long-span bridges: Enhanced data fusion of GPS displacement and deck accelerations. *Eng. Struct.* **2017**, *147*, 639–651. [CrossRef]
26. Dang, H.V.; Raza, M.; Nguyen, T.V.; Bui-Tien, T.; Nguyen, H.X. Deep learning-based detection of structural damage using time-series data. *Struct. Infrastruct. Eng.* **2021**, *17*, 1474–1493. [CrossRef]
27. Nguyen-Ngoc, L.; Tran-Ngoc, H.; Le-Xuan, T.; Nguyen, C.-T.; De Roeck, G.; Bui-Tien, T.; Wahab, M.A. A two-step approach for damage identification in bridge structure using convolutional Long Short-Term Memory with augmented time-series data. *Adv. Eng. Softw.* **2024**, *198*, 103795. [CrossRef]

28. Sony, S.; Gamage, S.; Sadhu, A.; Samarabandu, J. Vibration-based multiclass damage detection and localization using long short-term memory networks. In *Structures*; Elsevier: Amsterdam, The Netherlands, 2022; Volume 35.
29. Fu, L.; Tang, Q.; Gao, P.; Xin, J.; Zhou, J. Damage identification of long-span bridges using the hybrid of convolutional neural network and long short-term memory network. *Algorithms* **2021**, *14*, 180. [CrossRef]
30. Ahmed, B.; Mangalathu, S.; Jeon, J.-S. Seismic damage state predictions of reinforced concrete structures using stacked long short-term memory neural networks. *J. Build. Eng.* **2022**, *46*, 103737. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.