

Article

Guided Wave-Convolutional Neural Network Based Fatigue Crack Diagnosis of Aircraft Structures [†]

Liang Xu, Shenfang Yuan *, Jian Chen and Yuanqiang Ren

Research Center of Structural Health Monitoring and Prognosis, State Key Lab of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

* Correspondence: ysf@nuaa.edu.cn

[†] This paper is an extension version of the conference paper: Xu, L.; Yuan, S.F.; Chen, J.; Bao, Q. Deep learning based fatigue crack diagnosis of aircraft structures. In Proceedings of the 7th Asia-Pacific Workshop on Structural Health Monitoring 2018, Hong Kong, China, 12–15 November 2018.

Received: 26 June 2019; Accepted: 10 August 2019; Published: 15 August 2019



Abstract: Fatigue crack diagnosis (FCD) is of great significance for ensuring safe operation, prolonging service time and reducing maintenance cost in aircrafts and many other safety-critical systems. As a promising method, the guided wave (GW)-based structural health monitoring method has been widely investigated for FCD. However, reliable FCD still meets challenges, because uncertainties in real engineering applications usually cause serious change both to the crack propagation itself and GW monitoring signals. As one of deep learning methods, convolutional neural network (CNN) owns the ability of fusing a large amount of data, extracting high-level feature expressions related to classification, which provides a potential new technology to be applied in the GW-structural health monitoring method for crack evaluation. To address the influence of dispersion on reliable FCD, in this paper, a GW-CNN based FCD method is proposed. In this method, multiple damage indexes (DIs) from multiple GW exciting-acquisition channels are extracted. A CNN is designed and trained to further extract high-level features from the multiple DIs and implement feature fusion for crack evaluation. Fatigue tests on a typical kind of aircraft structure are performed to validate the proposed method. The results show that the proposed method can effectively reduce the influence of uncertainties on FCD, which is promising for real engineering applications.

Keywords: convolutional neural network; guided wave based monitoring; fatigue crack diagnosis; uncertainty; structural health monitoring

1. Introduction

Damage monitoring is a key issue for safety-critical systems such as aircraft, wind turbines, bridges, and nuclear plants [1,2]. As a common damage type, fatigue crack is one of the primary causes of structural failure, which is estimated to cause up to 90% of failures of metallic structures in service [3]. It is of considerable significance to monitor fatigue crack state to ensure the safety of structures in engineering applications.

The guided wave (GW) based structural health monitoring method [4,5] has been widely investigated to realize fatigue crack diagnosis (FCD) [6]. However, reliable FCD in real engineering applications still meets challenges because of uncertainties in real engineering applications, such as crystal structures in material, performances of sensors, bonding between sensors and structures, propagation paths, or internal inclination angle of fatigue cracks, etc. These uncertainties usually cause serious change both to the crack propagation itself and GW monitoring signals. Due to these uncertainties, FCD calibration by one or even a set of training structures usually does not work well for a monitoring structure in service.

Some developments have been reported on dealing with the influence of uncertainties on damage diagnosis. The method of extracting GW damage features that are sensitive to damage changes instead of external factors (such as environmental temperature, surface humidity) has been investigated [7–9]. However, the performance of these GW damage features has only been verified in simple structures such as aluminum plates. In real engineering applications, uncertainties will make the baseline signals dispersive, thus reducing the reliability of damage diagnosis. Furthermore, machine learning techniques are useful in modeling complex nonlinear phenomena in the presence of uncertainties [10], which can be introduced to construct a mathematical model of damage features to estimate the damage of structures. Yuan and Mei et al. [11] proposed the GW-Hidden Markov model based method to achieve a probabilistic evaluation of the propagation state of cracks. Uncertainties such as changing load and structural boundary conditions were considered in this research. Qiu et al. [12,13] proposed the GW-Gaussian mixture model based method to model the probability characteristic of GW features under changing loads. In the above methods, the number of Gaussian components, parameter initialization, and convergence of the model are still problems that need to be further researched. Moreover, the artificial neural network is an effective mathematical method that is widely used for damage estimating, which can learn the relevant features from data. Gu et al. [14] utilized an artificial neural network to study the change of natural frequency of structure caused by damages at different temperatures. It is shown that the proposed artificial neural network with novelty detection is able to distinguish damage occurrence and severity regardless of temperature variations and noise perturbations. However, the verification is only conducted based on a simple supported beam and finite element models. Other methods, such as support vector machine [15], fuzzy logic [16], and so on have also made some progress. However, these methods still have drawbacks, such as limited hierarchy, and the optimization process is prone to fall into the local optimum leading to a limited expression and limited generalization ability to data. These methods still face restrictions in engineering applications.

In recent years, deep learning has attracted significant attention [17]. As one of the deep learning methods, the convolutional neural network (CNN) has been successfully applied in many areas [18,19]. CNN has a deep hierarchy and complex structure, which makes it more capable of expressing data. Its unique structures, such as local connection, shared weights, and sub-sampling, allow it to automatically extract representations that are beneficial for classification by fusing a large amount of data [20]. These characteristics make CNN owns the potential to be used to improve the diagnosis accuracy of crack length under the influence of uncertainties. However, research on damage diagnosis of mechanical structures using CNN is rarely reported. Sun et al. [21] put forward a discriminant convolutional feature learning method. In this method, the CNN model is utilized to derive invariant and robust features, and vibration signals are used to monitor the fault conditions of the induction motor. Six different fault conditions are identified, but the severity of fault conditions is not distinguished. Guo et al. [22] proposed a hierarchical learning rate adaptive CNN based on an improved algorithm, in which the CNN model is based on the classical LeNet5 models proposed by LeCun. This method achieved both fault-pattern recognition and fault-size evaluation. Although showing potential in addressing the uncertainty issue, the current CNN based damage diagnosis are mainly performed based on vibration-based methods, which are not sensitive to small damages like crack. Until now, there has been little research on combining and taking advantages of the GW based structural health monitoring and CNN, which could be used for reliable FCD under the influence of uncertainties and is urgently needed.

In this paper, a GW-CNN based FCD method is proposed to address the influence of uncertainties on reliable FCD. In this method, different kinds of damage indexes (DIs) from multiple GW exciting-acquisition channels under one crack length were extracted to construct the input feature vector. Then, a designed CNN model was trained with feature vectors at different crack lengths extracted from historical GW data. Finally, the feature vector obtained from the monitored structure forward propagates through the trained CNN layer by layer. High-level features are automatically

extracted, which are beneficial for FCD. Reliable crack size can be directly obtained by the CNN output under the influence of uncertainties among different structures.

The rest of the article is organized as follows: in Section 2, details of the proposed method are shown. In Section 3, the feasibility of the proposed method is verified by attachment lug specimens, and the advantages are verified by comparative experiments. The conclusion is drawn in Section 4.

2. GW-CNN Based Fatigue Crack Diagnosis Method

In this section, the proposed GW-CNN based FCD method is introduced. Firstly, GW features that are extracted from multiple GW exciting-acquisition channels are used to construct the CNN input vector. Then, a CNN model is designed and trained for FCD. Finally, the trained CNN model can be used for reliable FCD of a monitored structure.

2.1. Multi-Channel and Multi-GW Features Extraction

GW is a kind of elastic waves that propagates in plate-like structures. It can travel a long distance in structures with small energy loss; hence, it can be used to monitor a relatively large structure area. Piezoelectric transducer (PZT) is a conventional kind of sensor used to excite and receive GW in structure. A typical sensor network configuration is shown in Figure 1 where J ($J \in N^+$) exciting-acquisition channels are formed. From one sensor network, the number of obtained GW signals is J . After the GW is excited in the structure by a PZT, the interaction of the GW with the crack can influence the GW propagation. By comparing the received signals under healthy and cracked conditions, the length of the crack can be estimated.

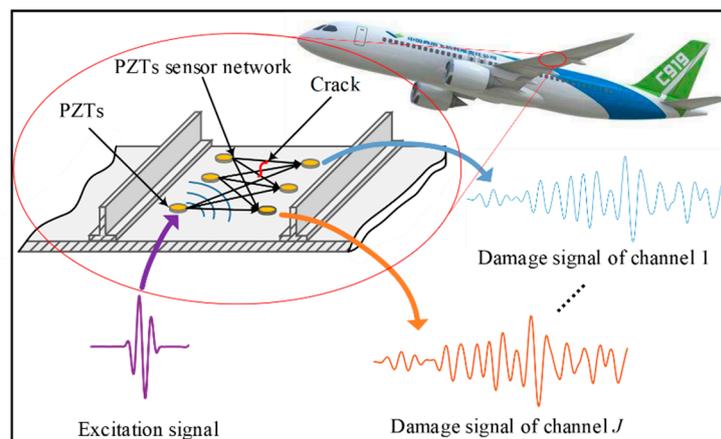


Figure 1. The guided wave (GW) based structural health monitoring method for on-line crack monitoring. PZT: Piezoelectric transducer.

Many kinds of DIs can be defined to evaluate the variations between the baseline signal collected when the structure is healthy and the monitoring signal when the crack propagates in the structure. These DIs are extracted from GW signals in the time domain, frequency domain, and time–frequency domain.

As mentioned above, various GW features can be obtained from one crack length, which provides the potential of reliable FCD through data fusion of CNN. Based on this, a multi-channel and multi-GW features extraction method is proposed as follows. Firstly, under the one crack length, GW signals from J exciting-acquisition channels are acquired, recorded as $\mathbf{S} = \{s^1, s^2, \dots, s^i, \dots, s^J\}$. Then, from one GW signal s^i , d kinds of DIs were chosen for GW features extraction, which were recorded as

$D^j = [D_1^j, D_2^j, \dots, D_d^j], j \in \{1, 2, \dots, J\}$. Finally, the extracted DIs were arranged into a feature vector of \mathbf{p} in order of channel as shown in Equation (1):

$$\mathbf{p} = [D_1^1, D_2^1, \dots, D_1^j, D_2^j, \dots, D_{d'}^j, \dots, D_{d-1}^J, D_d^J]_{(J \times d) \times 1} \quad (1)$$

The feature vector \mathbf{p} was standardized to \mathbf{p}^{norm} to obtain a similar data distribution [23] for each input feature to achieve a better learning efficiency and improve the accuracy of FCD.

2.2. CNN Based Fatigue Crack Diagnosis

Based on the above procedure of data preprocessing, DIs from multi-channels are formed into a one-dimensional (1D) feature vector \mathbf{p}^{norm} , which is designed as the CNN input. Crack lengths are divided into z ($z \in \mathbb{N}^+$) crack sizes for classification; each crack size is an interval that contains a certain crack length. Softmax classifier [24] is adopted as the output layer to express z crack sizes, which makes the CNN output a vector that has the form of $\mathbf{a} = [a_1, a_2, \dots, a_k, \dots, a_z]_{z \times 1}$. The classification result corresponds to the class that has the highest output value. For example, the desired output \mathbf{q} , which denotes the k th crack size, is defined as a vector with z elements as $\mathbf{q} = [q_1, q_2, \dots, q_k, \dots, q_z]_{z \times 1}$, in which the k th element value $q_k = 1$ and the rest are 0.

In this study, the GW signals corresponding to crack lengths are obtained from known structures, which are called historical data. The feature vector \mathbf{p}^{norm} obtained from historical data and its corresponding desired output \mathbf{q} is formed into an input sample $\{\mathbf{p}^{norm}, \mathbf{q}\}$. These historical input samples are utilized for CNN training.

CNN is designed to process data that come in the form of multiple arrays. For example, two-dimensional data for pictures or audio spectrograms and 1D data like signals and sequences [19]. There are four key ideas behind CNN that take advantage of the properties of input data: local connections, shared weights, pooling, and the use of multiple layers [16]. Typically, CNN for classification is composed of five parts [18]: convolutional layers (CL), pooling layers (PL), flatten layer (FL), full-connected layers (FCL), and classification layer, as shown in Figure 2. M and L represent the total layer number of convolutional layers and full-connected layers respectively. The convolutional and pooling layers are alternated layer by layer for extracting features from input data, the flatten layer is used to transform the outputs of pooling layer into the 1D feature set, and the last two parts are employed for classification from the learned features.

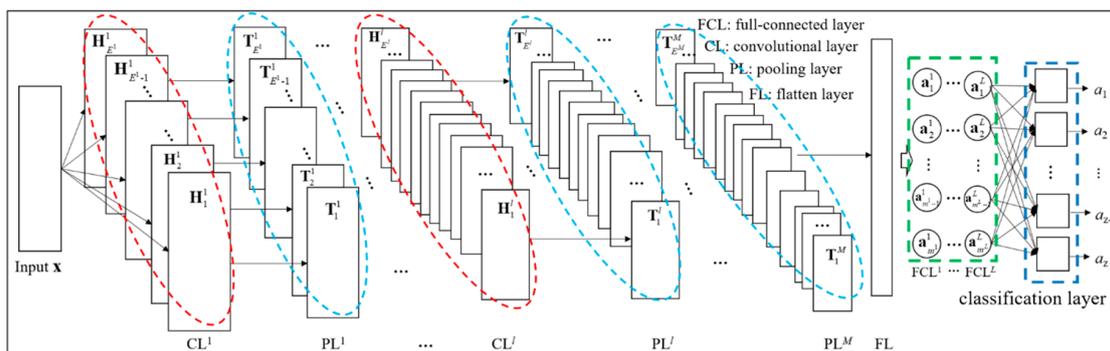


Figure 2. Architecture of a typical convolutional neural network (CNN).

The convolutional layer, as the name suggests, utilizes the operation of convolution to process input data. In one convolutional layer, there are several convolutional kernels (or filters), for example E^l convolutional kernels at the l th ($l \in [1, 2, \dots, L]$) layer. Each convolutional kernel consists of a certain number of trainable weights. One convolutional feature at l th layer is calculated as follows:

$$H_i^l(j) = f(\mathbf{p}^l(j) * \mathbf{w}_i^l + b_i^l) \quad (2)$$

where the symbol of asterisk * denotes the convolutional operation; $H_i^l(j)$ represents the convolutional feature of i th convolutional kernel that slides to the j th region of the input vector, where $i \in [1, 2, \dots, E^l]$; f represents the activation function; $\mathbf{p}^l(j)$ represents the vector composed of the elements in j th region of the input vector; $\mathbf{w}_i^l = [w_{i,1}^l, w_{i,2}^l, \dots, w_{i,r^l}^l]$; and b_i^l denote the i th convolutional kernel and its bias.

The activation function is defined as Rectified Linear Units (ReLU) function, which has more simple derivative result than traditional tanh and sigmoid functions leading to faster training when using the training algorithm. The mathematical expression of the ReLU function is shown as follows:

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

Each convolutional kernel slides over the input vector, then all the convolutional features constitute the convolutional output \mathbf{H}_i^l which is obtained by the i th convolutional kernel:

$$\mathbf{H}_i^l = [H_i^l(1), H_i^l(2), \dots, H_i^l(n^l - r^l + 1)] \quad (4)$$

Different convolutional kernels attain different convolutional outputs, which fuse information of the input vector from different ways. That is to say, different perspectives of features that are beneficial to classification can be extracted. Furthermore, the convolutional layer provides characteristics of local connections and shared weights, which can reduce the number of parameters, significantly reduce the computational costs, and have certain robustness to local noise [19].

The pooling layer is usually connected after the convolutional layer, which is used to sub-sampling features with maximum pooling, average pooling or other operation. Assuming the pooling size is c , the maximum pooling feature of the j th region at l th pooling layer can be expressed as:

$$T_i^l(j) = \max_j(\mathbf{H}_i^l(j)) \quad (5)$$

where $T_i^l(j)$ represents the pooling feature when slides to the j th region of the i th convolutional output and $\mathbf{H}_i^l(j)$ represents the j th region of the i th convolutional output.

The vector composed of all pooling features is the pooling output denotes as \mathbf{T}_i^l after sliding. The key idea of pooling is to reduce the amount of data transferred to the next layer. Moreover, the pooling takes typical features as its outputs, which owns the ability of invariance [19].

In consequence, the framework of the proposed CNN based FCD method can be described in Figure 3. It is noticeable that the proposed CNN model has a nine-layer configuration, which consists of one input layer, three convolutional layers followed with one pooling layer, a flatten layer, two full-connected layers, and an output layer. In the CNN model, to obtain features that are less affected by uncertainties, the first convolutional layer (CL1) is used to extract features from a GW signal and fuse crack information from different channels of GW signals at the same time. As mentioned above, DIs are arranged into an input vector in order of channel. Therefore, the crack information from the same channel is in d DIs that are arranged sequentially in the input feature vector. Accordingly, based on the above analysis, a set of kernels with the size of $d \times 1$ are utilized in CL1. In this study, each element of the feature vector contains useful crack information and the number of elements is really scarce. However, a pooling layer may lead to the loss of valuable information because it only chooses one typical feature as its output in a pooling area. Hence, there is no pooling layer followed by CL1, but two convolutional layers are added as CL2 and CL3. Taking the Pyramid Shape mentioned in Leslie's paper [25] into consideration, the kernel size decreases throughout the architecture to get better performance. Next, a pooling layer (PL) is added followed by a flatten layer which is used to transform the outputs of PL into the 1D feature set. Then, two full-connected layers (FCL) are added, denoted by FCL1 and FCL2. The outputs of the flatten layer are employed as the inputs of FCL1. These two full-connected layers are designed as hidden layers to map the features into CNN outputs and select

the features that are less affected by uncertainties. Most features that less affected by uncertainties will be found, which are in favor of leading to a higher FCD accuracy. In addition, the L2 regularization method [26] is operated in full-connected layers during the training process to reduce the possibility of overfitting.

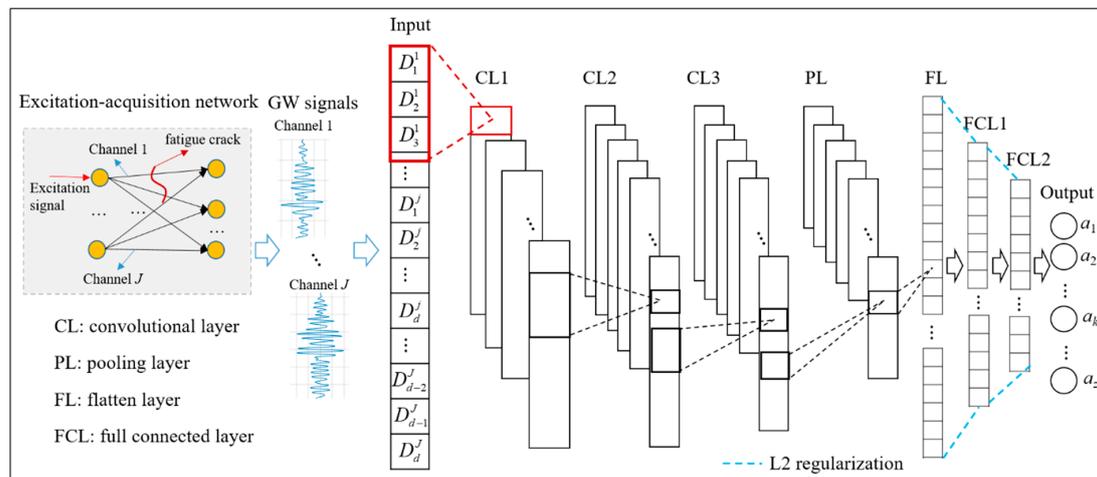


Figure 3. Structure of proposed fatigue crack diagnosis (FCD) CNN model.

The training of CNN is a procedure of optimizing the connected weights and bias, which are trained via minimizing the cost function with training data. In the proposed method, training data are input vectors at different crack length collected from historical data as mentioned above. The cost function of the proposed model is defined as the cross-entropy function in accordance with the softmax classifier which is the output layer of the model, given as the following equation:

$$G(w; q_k, a_k) = -\frac{1}{z} \sum_{k=1}^z [q_k \ln a_k + (1 - q_k) \ln(1 - a_k)] + \frac{\lambda}{2} \sum_{l=1}^L \sum_{j=1}^{m^l} \sum_{i=1}^{m^{l+1}} (w_{j,i}^{l+1})^2 \quad (6)$$

where $w_{j,i}^{l+1}$ denotes the weight that connects the j th neuron at the l th layer and the i th neuron at the $(l + 1)$ th layer and m^l denotes the neuron number at l th layer. In Equation (6), the first term is the cross-entropy between the CNN output and the desired output, and the second term is the L2 regularization part, in which $\lambda (0 < \lambda < 1)$ denotes the regularization coefficient. Here, weights are initialized with Xavier method [27], which makes the weights of each layer obey the following uniform distribution:

$$w \sim U\left[-\frac{\sqrt{6}}{\sqrt{m^l + m^{l+1}}}, \frac{\sqrt{6}}{\sqrt{m^l + m^{l+1}}}\right] \quad (7)$$

Bias is initialized to be $b^l = 0$. Then, weight and bias are updated with the Adam optimization method [28], in which independent adaptive learning rate is designed for different parameters by calculating the first and second order moment estimation of the gradient. In the Adam optimization method, the first-order moments of weight gradient and bias gradient are initialized to be $V_{dw^l} = 0, V_{db^l} = 0$ and the second-order moments are initialized to be $I_{dw^l} = 0, I_{db^l} = 0$ firstly. Next, based on the backpropagation algorithm, the recurrence relation for the sensitivity can be written as follows:

$$s_j^l = (f^l)'(x) \left(w_{j,i}^{l+1} \right)^T s_i^{l+1} \quad (8)$$

where s_j^l and s_i^{l+1} denote the sensitivity of j th neuron at the l th layer and the i th neuron at the $(l + 1)$ th layer and $(f^l)'(x)$ denotes the derivative of the activation function at l th layer. Then, the gradient can be calculated by the following expression:

$$dw^l = s^l a^l, db^l = s^l \quad (9)$$

After that, V_{dw^l} , V_{db^l} , I_{dw^l} , and I_{db^l} can be updated as follows:

$$V_{dw^l}(k+1) = \beta_1 V_{dw^l}(k) + (1-\beta_1) dw^l, V_{db^l}(k+1) = \beta_1 V_{db^l}(k) + (1-\beta_1) db^l \quad (10)$$

$$I_{dw^l}(k+1) = \beta_2 I_{dw^l}(k) + (1-\beta_2) (dw^l)^2, I_{db^l}(k+1) = \beta_2 I_{db^l}(k) + (1-\beta_2) (db^l)^2 \quad (11)$$

where k denotes the number of iterations and β_1 and β_2 are parameters that are generally preset as 0.9 and 0.999, respectively. Then, the updated values are corrected as follows:

$$V_{dw^l}^{\text{corrected}}(k+1) = \frac{V_{dw^l}(k+1)}{1-\beta_1(k)}, V_{db^l}^{\text{corrected}}(k+1) = \frac{V_{db^l}(k+1)}{1-\beta_1(k)} \quad (12)$$

$$I_{dw^l}^{\text{corrected}}(k+1) = \frac{I_{dw^l}(k+1)}{1-\beta_2(k)}, I_{db^l}^{\text{corrected}}(k+1) = \frac{I_{db^l}(k+1)}{1-\beta_2(k)} \quad (13)$$

where the superscript represents the corrected value. Finally, the weight and bias are updated as the following equations:

$$w^l(k+1) = w^l(k) - \alpha \frac{V_{dw^l}^{\text{corrected}}(k+1)}{\sqrt{I_{dw^l}^{\text{corrected}}(k+1) + \varepsilon}} \quad (14)$$

$$b^l(k+1) = b^l(k) - \alpha \frac{V_{db^l}^{\text{corrected}}(k+1)}{\sqrt{I_{db^l}^{\text{corrected}}(k+1) + \varepsilon}} \quad (15)$$

where α denotes the learning rate and ε is set to prevent the denominator from being 0, which can be $\varepsilon = 10^{-8}$.

The training of CNN is finished when the cost function is close to converging. Eventually, the input feature vector from a similar monitored structure is put into the trained model to get its diagnostic crack size.

3. Experimental Verification and Analysis

Attachment lugs are an important type of joint in aircraft structures. They are typically used as a connection between two components, and are quite susceptible to fatigue cracks due to stress concentrations. In this paper, fatigue tests are performed on attachment lug specimens to validate the proposed method. The uncertainty of GW signals and DIs are discussed. After that, the designed CNN model and its training process are given. Finally, the fatigue crack diagnosis results based on the proposed method are presented and discussed.

3.1. Fatigue Tests of Attachment Lug Specimens

Fatigue tests are performed on six attachment lug specimens, labeled from T1 to T6. To be close to the real engineering applications, the specimens are made of 5 mm-thick LY12 aluminum alloy, which have a hole with the diameter of 25 mm. Their geometry and sensor layout are shown in Figure 4. In real engineering applications, the attachment lug usually suffers from the axial tension load during service. Finite element method results for the lug show that stress concentration occurs at the edge of the hole. Since the stress concentration area often initiates crack, a 2 mm long notch is made at the edge of the lug hole to make the crack initiation and control the direction of crack growth, and this notch

was created by a wire-electrode cutting machine. The same sensor layout is chosen for each specimen, in which PZT sensors named PZT1–PZT3 are arranged on the front side and PZT4–PZT5 are arranged on the opposite side. As shown in Figure 5b, a fixture with a dowel pin is adopted to connect the lug, and transmits the axial tension load, which is a real load transmitting style in engineering applications.

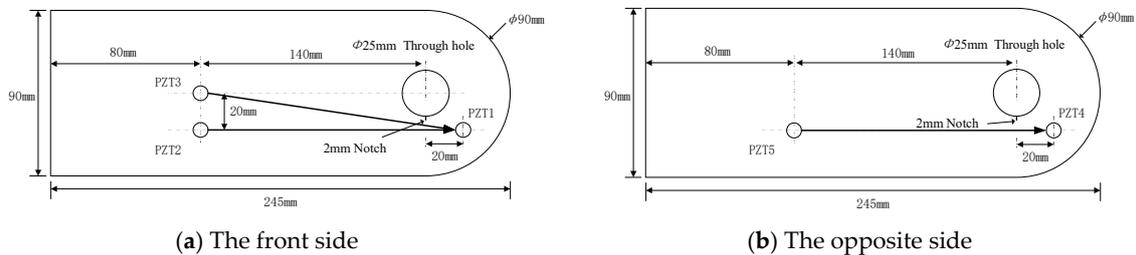


Figure 4. Geometry and sensor layout of the attachment lug.

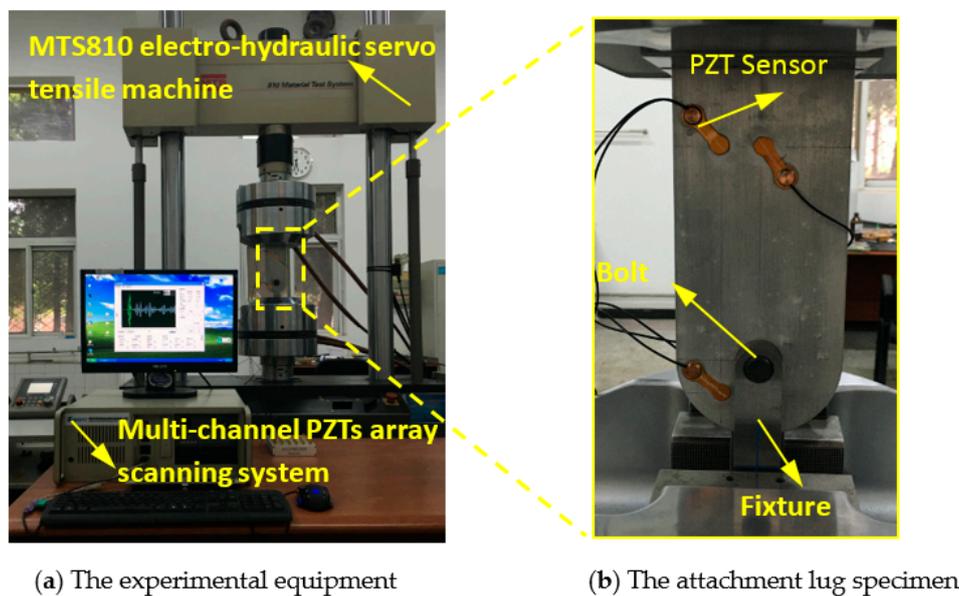


Figure 5. The fatigue test setup.

An MTS810 electro-hydraulic servo tensile machine is used to apply the fatigue load, as shown in Figure 5a. According to engineering experience, a sinusoidal load with a peak value of $F_{max} = 18$ kN is chosen for fatigue tests, which is 25% of the fracture load of 72kN. The load frequency is chosen as 10Hz and the stress ratio is $R = F_{min}/F_{max} = 0.1$. The crack lengths were measured with a digital magnifier and scale lines on one surface of the specimen. The crack growth processes of specimens T1–T6 are shown in Figure 6. It is noteworthy that crack growth of different specimens is different; this will introduce uncertainties for FCD.

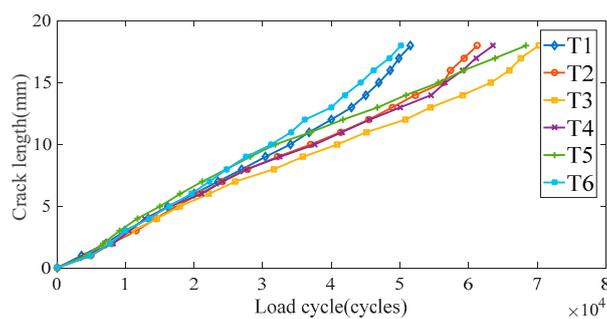


Figure 6. Crack growth of specimens T1–T6.

During the fatigue tests, the multi-channel PZTs array scanning system developed by the authors' group is employed to perform the GW based fatigue crack monitoring [29]. Here, the three-cycle sine burst signal with central frequency of 160 kHz and the exciting voltage of ± 70 V is adopted as exciting signal, and the sampling frequency is set as 50MHz. Moreover, PZT2, PZT3, and PZT5 are chosen to excite signals, while PZT1 and PZT4 are used to acquire signals. As a result, three effective exciting-acquisition channels are obtained denoted as 2-1 channel, 3-1 channel, and 5-4 channel as shown in Figure 4. The fatigue load is suspended to collect two sets of GW signals when the fatigue crack grows every 1 mm for the crack length of 0 mm to 18 mm, where each set contains three channels of GW signals.

As a consequence, 38 sets of signals are collected from each specimen, and 228 set of signals for six specimens in total. Figures 7 and 8 show the GW signals of specimens T1 and T4 under partial crack length during the fatigue crack monitoring in the 2-1 channel.

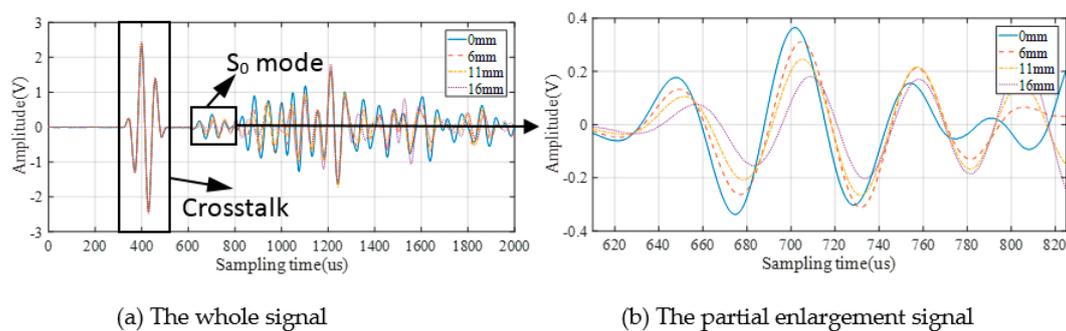


Figure 7. GW signals of 2-1 channel in specimen T1.

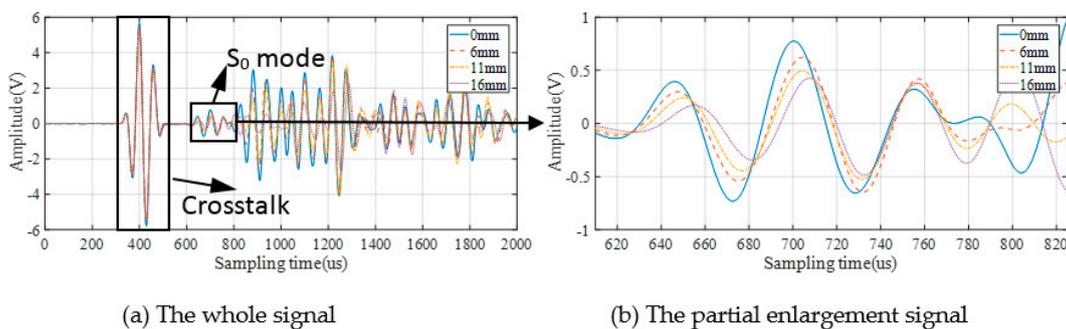


Figure 8. GW signals of 2-1 channel in specimen T4.

It can be seen from the above figures that GW signals obtained from different specimens (T1/T4) are different in both amplitude and phase. These are the reflection of uncertainties among different structures on original GW signals.

3.2. GW Monitoring and DIs Extraction

According to Figures 7 and 8, the S_0 mode which has a clear wave packet is chosen to calculate DIs. However, the other part of the signal is affected by complex mode, boundary reflection, and so on, which makes it difficult to extract the crack information from the mix of these waves. In addition, for different specimens, the difference of signal behind S_0 in different specimens is much greater than that of S_0 mode in different specimens. That will make the crack quantification problem more difficult. Seven kinds of DIs are used for GW features extraction as shown in Table 1, where $B(t)$ and $D(t)$ are the baseline signal and the monitoring signal, respectively; the baseline is the GW signal acquired after the notch is machined and it is unique to each sample; t_1 and t_2 are the start time and the end time of the selected wave packet; and ω denotes the signal frequency obtained by the Fourier transform.

Table 1. DI extraction algorithms.

Damage Index DI	Extraction Algorithm
Cross correlation [6]	$DI = 1 - \sqrt{\frac{[\int_{t_1}^{t_2} B(t)D(t)dt]^2}{\int_{t_1}^{t_2} B^2(t)dt \int_{t_1}^{t_2} D^2(t)dt}}$
Spatial phase difference	$DI = \int_{t_1}^{t_2} (\tilde{D}(t) - \alpha B(t))^2 dt, \tilde{D}(t) = \frac{D(t)}{\sqrt{\int_{t_1}^{t_2} D^2(t)dt}}, \alpha = \frac{\int_{t_1}^{t_2} \tilde{D}(t)B(t)dt}{\int_{t_1}^{t_2} B^2(t)dt}$
Spectrum loss	$DI = \frac{\int_{\omega_1}^{\omega_N} B(\omega) - D(\omega) d\omega}{\int_{\omega_1}^{\omega_N} B(\omega) d\omega}$
Central spectrum loss	$DI = \frac{a(\omega) - b(\omega)}{a(\omega)}, a(\omega) = \max(B(\omega)), b(\omega) = \max(D(\omega))$
Differential curve energy [7]	$DI = \frac{\sum_{n=2}^N [b(t_n) - b(t_{n-1})]^2}{\sum_{n=2}^N [B(t_n) - B(t_{n-1})]^2}, b(n) = B(n) - D(n)$
Normalized Correlation Moment [8]	$DI = \frac{\int_{\tau=t_1}^{\tau=t_2} \tau^k r_{HH}(\tau) d\tau - \int_{\tau=t_1}^{\tau=t_2} \tau^k r_{HD}(\tau) d\tau}{\int_{\tau=t_1}^{\tau=t_2} \tau^k r_{HH}(\tau) d\tau}, r_{HD}(\tau) = \frac{\int_{-\infty}^{+\infty} B(t)D(t-\tau)dt}{\int_{-\infty}^{+\infty} B(t)B(t-\tau)dt}$
Differential signal energy	$DI = \int_{t_1}^{t_2} (\tilde{B}(t) - \tilde{D}(t))^2 dt, \tilde{B}(t) = \frac{B(t)}{\sqrt{\int_{t_1}^{t_2} B^2(t)dt}}, \tilde{D}(t) = \frac{D(t)}{\sqrt{\int_{t_1}^{t_2} D^2(t)dt}}$

As is shown in the table, These DIs reflect the crack from the time domain and frequency domain, which can be more comprehensive response to crack length.

1. The Cross correlation is a DI in the time domain, which is describe the degree of correlation between the normalized baseline signal and the monitoring signal;
2. The Spatial phase difference is also a DI in the time domain that describes the size of the angle between the normalized baseline signal and the monitoring signal;
3. The Spectrum loss is a DI in the frequency domain, which describes the difference value in spectrum between two signals;
4. The Central spectrum loss is also a DI in the frequency domain, which measures the change of central spectrum between baseline and monitoring signal;
5. The Differential curve energy is a DI that measured the variation of waveform curve of the difference of two signal;
6. The Normalized Correlation Moment is a DI that based on local statistical features of the waveform; the energy and phase change of the signal has been taken into consideration;
7. The differential signal energy is a DI that measured the variation of signal energy.

To give an intuition of what the uncertainty of DIs is and show how the multi-channel DIs fusing owns the potential of improving the accuracy of FCD. Firstly, typical time-domain DI (cross correlation) and frequency-domain DI (central spectrum loss) varying with crack growth in the same typical 2-1 channel are shown in Figure 9.

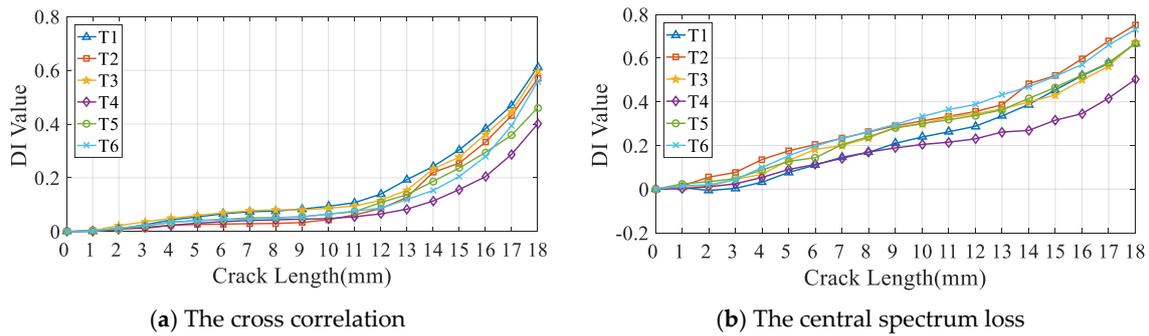


Figure 9. DI of 2-1 channel in specimens T1–T6.

It is apparent that these DIs can reflect the fatigue crack growth effectively, but there is a large uncertainty of the same DI values under the same fatigue crack length for different specimens, which means the same DI value corresponds to multiple crack lengths. Meanwhile, numerical distributions of different DIs are altered during the fatigue crack growth process for the same specimen, which indicates that the influence of crack on GW signal is described from different perspectives by different DIs. This provides a basis for data fusion with CNN.

Then, the cross-correlation DI is chosen to show its value varying with fatigue crack growth for typical specimen T1 and T4 under three different exciting-acquisition channels. As shown in Figure 10, the distribution of DIs in different exciting-acquisition channels are different during crack growth, which indicates that different channels of DIs provide different perspectives of crack information for CNN to learn.

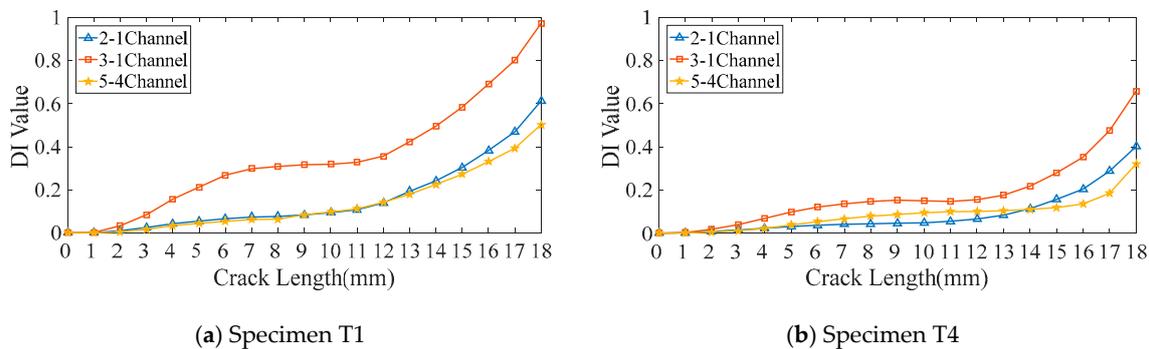


Figure 10. Different channels of cross-correlation DI.

As is shown from the above analysis, three channels of GW signals can be acquired under one crack length during the fatigue tests. As for each GW signal, 7 kinds of DIs are extracted. Finally, $7 \times 3 \times 1 = 21$ DIs are obtained under one crack length to construct a 1D feature vector \mathbf{p} .

3.3. GW-CNN Based Diagnosis Training

In this study, all input vectors \mathbf{p} that correspond to different crack length are standardized to \mathbf{p}^{norm} firstly. Then, the fatigue cracks are classified into 19 crack sizes according to its crack length: 1 crack size per 1mm, named C1–C19, as shown in Table 2. Crack size C_k ($k \in [1, 2, \dots, 19]$) is transformed into the desired output form of $\mathbf{q} = [0, 0, \dots, 1, \dots, 0]$ in which the k th element is 1 and the rest are 0. The standardized input \mathbf{p}^{norm} and its corresponding desired output \mathbf{q} are formed to input-output pair: $\{\mathbf{p}_{21 \times 1}^{norm}, \mathbf{q}_{19 \times 1}\}$ to be an input sample.

Table 2. Fatigue crack sizes corresponding to crack length.

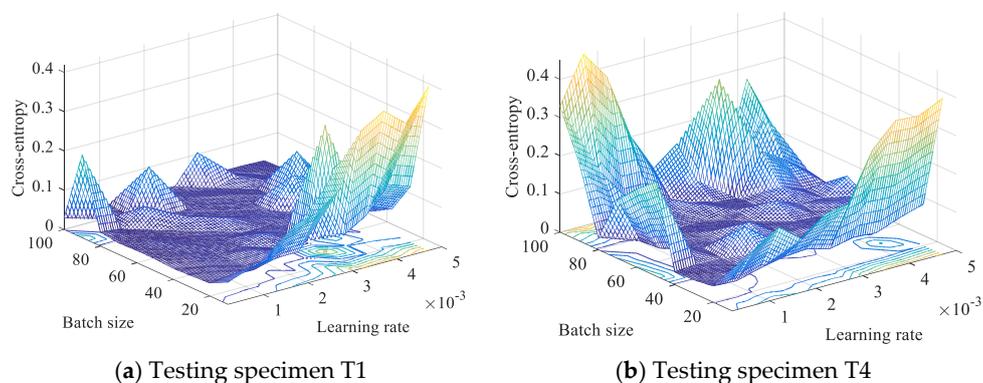
Crack Size	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Crack length (mm)	0	1	2	3	4	5	6	7	8	9
Crack size	C11	C12	C13	C14	C15	C16	C17	C18	C19	
Crack length (mm)	10	11	12	13	14	15	16	17	18	

In real engineering applications, the fatigue crack information obtained from existing structures is often used to diagnose fatigue crack of unknown structures. Thus, samples from one of the specimens among T1–T6 are chosen as the testing samples, and the rest samples are used for training, as shown in Table 3. Consequently, as for each testing specimen, 38 and 190 samples are utilized for testing and training.

Table 3. Testing specimen and its corresponding training specimens.

Testing Specimen	T1	T2	T3	T4	T5	T6
Training specimens	T2–T6	T1, T3–T6	T1–T2, T4–T6	T1–T3, T5–T6	T1–T4, T6	T1–T5

The algorithm implementation of the designed CNN is under Python on a desktop with Nvidia graphics processing unit (GPU) GTX960. Before the model training, several important hyper-parameters, including learning rate, mini-batch size, kernel size and kernel numbers in convolutional layers and pooling layer, neuron numbers of full-connected layers, and regularization coefficient λ , needed to be determined in advance. In this work, the CNN configuration was set to CL1(16)-CL2(32)-CL3(32)-PL(32)-FCL1(1024)-FCL2(512), and the kernel size was set to CL1(7)-CL2(5)-CL3(5) according to previous studies. The L2 regularization was added to FCL1 and FCL2, and the regularization coefficients were set to be 0.01 and 0.05, respectively. The optimal learning rate and batch size were obtained via a trial-and-error method. Here, various batch sizes were considered, ranging from 10 to 100 with an interval of 10, while the learning rate varies from 0.0002 to 0.005 with an interval of 0.0005. Figure 11 shows the cross-entropy of CNN model for testing specimens T1 and T4. It is noticeable that small values of batch size and large values of learning rate will lead to poor performance (big costs value) of the trained model. With the increase of batch size and decrease of the learning rate, the cross-entropy error of the CNN model will be gradually decreased and then tends to be stable or will be increased a little bit after the costs arrive at its minimum value. According to the result in Figure 11 and a more subtle adjustment, 24 and 0.001 were adopted as the optimal batch size and learning rate for CNN training.

**Figure 11.** Cross-entropy of the CNN model with different batch size and learning rates.

Then, the CNN model was trained based on optimal model parameters using training samples. In the situation where specimen T1 was used as the testing specimen, the training samples from T2–T6

were put into CNN model to train the model with Adam algorithm. As shown in Figure 12, the CNN model at the iteration of 22 was chosen as the trained CNN model because the costs started to converge and too much training may lead to overfitting. When another specimen was selected as the testing specimen, the corresponding CNN model was trained in the same way above.

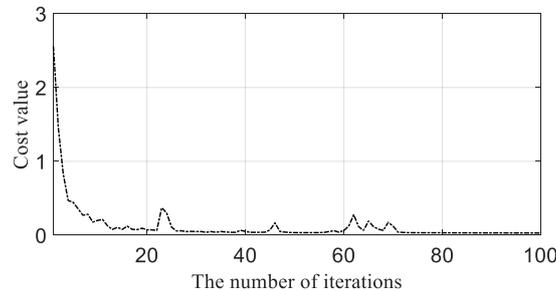


Figure 12. Costs value varying with iteration when the testing specimen is T1.

3.4. Diagnosis Results and Discussion

Testing samples from the testing specimen are put into the trained CNN model to obtain diagnostic crack size. After that, the testing accuracy shown as Table 4 can be calculated as the following equation:

$$\text{Testing accuracy} = \frac{NCC}{NT} \times 100\% \tag{16}$$

where *NCC* denotes the number of correctly classified crack size by CNN and *NT* denotes the total number of testing samples. The CNN diagnostic crack size and its real crack size are given in Figure 13.

Table 4. Diagnostic results of different testing specimen.

Testing Specimen	T1	T2	T3	T4	T5	T6
Accuracy	100%	86.84%	97.37%	86.84%	100%	94.74%

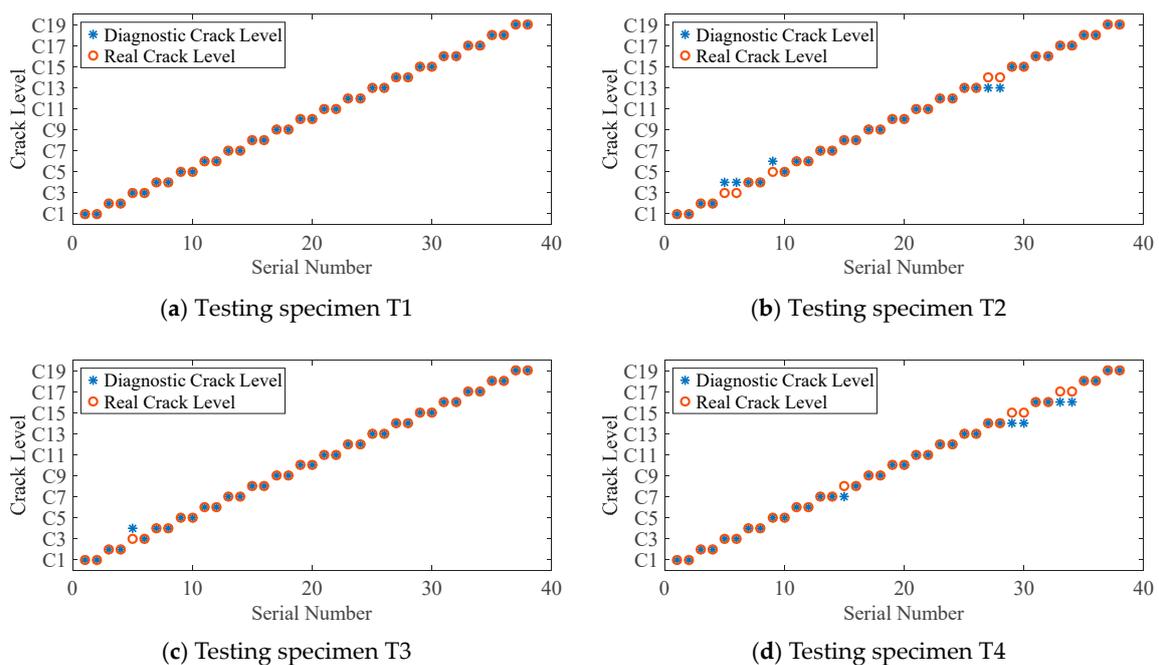


Figure 13. Cont.

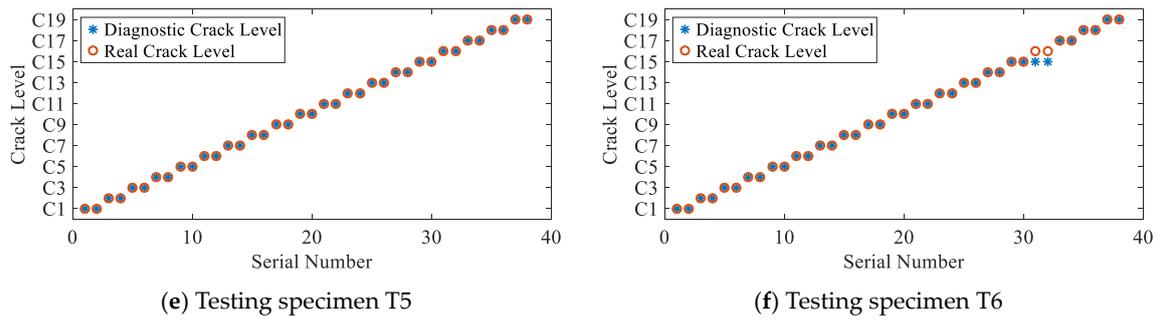


Figure 13. Diagnostic results of different testing specimen.

Seen from the above results, the diagnostic accuracies of testing specimens T1 and T5 are 100%; as for testing specimens T2 and T4, the accuracies are 86.84%. To display what happened in T2 and T4, the distributions of typical DI (cross-correlation) in the testing data (red triangle) and training data (blue circle) of the 2-1 channel varying with crack sizes are shown in Figure 14, respectively. In Figure 14a, DI values of testing specimen T2 are basically smaller than those of training DI values before C14, but the values become larger after C15. In Figure 14b, DI values of the testing specimen of T4 are basically smaller than DI values of training specimens. These phenomena also happen in other kinds of DIs or other channels, which make it difficult to estimate the fatigue crack size by the general rule of the DI values varying with crack sizes of training data. However, the diagnostic error is only 1 mm as shown in Figure 14, which is allowed in real engineering applications. This indicates that the proposed method is promising in the presence of uncertainties.

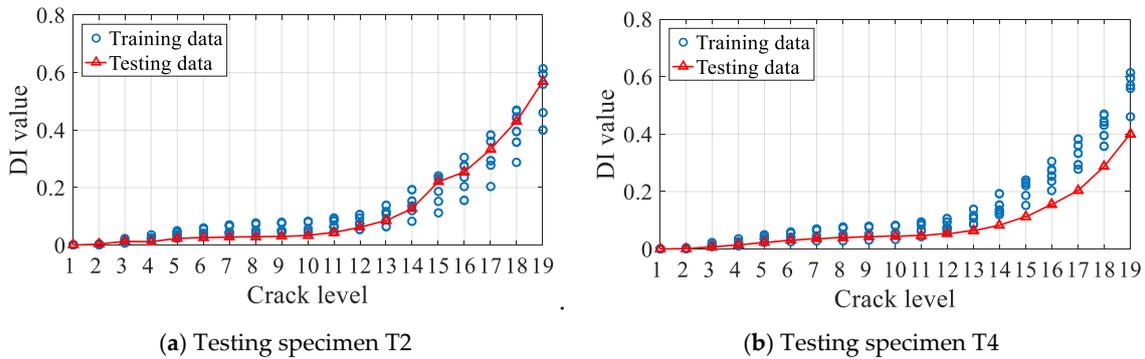


Figure 14. The distribution of cross correlation DI for testing and training data.

To illustrate the effectiveness of the proposed method in solving the problem of uncertainties, the Unified Euclidean Distance is adopted to quantitatively analyze the similarity of input feature vector and features extracted by CNN between two different specimens. Assuming there are two kinds of variables $\mathbf{X} = [x_1, x_2, \dots, x_i, \dots]$ and $\mathbf{Y} = [y_1, y_2, \dots, y_j, \dots]$, in which $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ and $y_j = [y_{j1}, y_{j2}, \dots, y_{jm}]$. These two variables are different in variable length ($n \neq m$) and numerical distribution. Taking the Unified Euclidean Distance between specimen 1 and specimen 2 as an example, the Unified Euclidean Distance of D_{12}^X between $x_1 = [x_{11}, x_{12}, \dots, x_{1n}]$ and $x_2 = [x_{21}, x_{22}, \dots, x_{2n}]$ was calculated as follows: firstly, x_1 and x_2 were normalized to reduce the influence of different numerical distributions expressed as Equation (17); then, the results were divided by the length of the variable n to avoid the influence of length as expressed in Equation (18). The Unified Euclidean Distance D_{12}^Y between $y_1 = [y_{11}, y_{12}, \dots, y_{1m}]$ and $y_2 = [y_{21}, y_{22}, \dots, y_{2m}]$ was calculated in the same way as D_{12}^X . After that, D_{12}^X and D_{12}^Y can be used to compare which is more similar between two specimens. The larger the Unified Euclidean Distance is, the less similar the data will be.

$$x'_{1i} = x_{1i} / \max(\mathbf{x}_1), x'_{2i} = x_{2i} / \max(\mathbf{x}_2), i \in [1, 2, 3, \dots, n] \quad (17)$$

$$D_{12}^X = \sqrt{\sum_{k=1}^n (x'_{1k} - x'_{2k})^2 / n} \quad (18)$$

Suppose that T1 is the testing specimen, the typical crack sizes C7 and C18 are selected to compare two kinds of Unified Euclidean Distance (a. input feature vector, b. output features from the FCL2, denoted as the output feature vector) between two specimens. Results are shown in Table 5, where D^{in} and D^{out} denotes the Unified Euclidean Distance of the input feature vector and output feature vector respectively.

Table 5. Unified Euclidean Distance between different specimens.

Crack Size	Input Samples			Output Features		
	$D_{T2,T4}^{\text{in}}$	$D_{T2,T1}^{\text{in}}$	$D_{T4,T1}^{\text{in}}$	$D_{T2,T4}^{\text{out}}$	$D_{T2,T1}^{\text{out}}$	$D_{T4,T1}^{\text{out}}$
C7	2.0159	0.2142	1.8326	0.0002	0.0016	0.0011
C18	0.0204	0.0253	0.0137	0.0002	0.0011	0.0015

In Table 5, it is apparent that, under the same crack size, Unified Euclidean Distance of output features are smaller than Unified Euclidean Distance of input features between the identical specimens. This shows that output features between training specimens (T2 and T4) or between testing specimen and training specimen (T2 and T1, T4 and T1) are more similar than input features under the same crack size. That is to say, after features fusion and extraction by CNN, the invariance of different specimens under the same crack size is found, which reduces the uncertainties and effectively improves the reliability of FCD.

Furthermore, two sets of comparative experiments (I and II) were carried out to verify the advantages of the proposed method for reliable FCD under the influence of uncertainties.

(I). To verify advantages of multi-channel DIs extraction, DIs extracted from a single-channel (2-1 channel, 3-1 channel, and 5-4 channel) were chosen to form the input sample, which has the dimension of 7×1 to be the comparison modes. Notably, the form of the CNN input is the only difference in this comparative experiment.

Diagnostic accuracies of specimens T1 and T4 for three single-channel input forms and multi-channel input form are shown in Figure 15, respectively. Seen from the figures, diagnostic accuracies for single-channel input forms are lower than 80% for T1 and lower than 60% for T4. This indicates that the uncertainty of DIs seriously affects the diagnostic accuracies. Compared with the single-channel input forms, the diagnostic accuracies for multi-channel input form are 100% and 86.84% for T1 and T4. The multi-channel input vector contains more crack information for the designed CNN to extract features that are beneficial to classification, leading to the improvement of the diagnostic accuracy effectively.

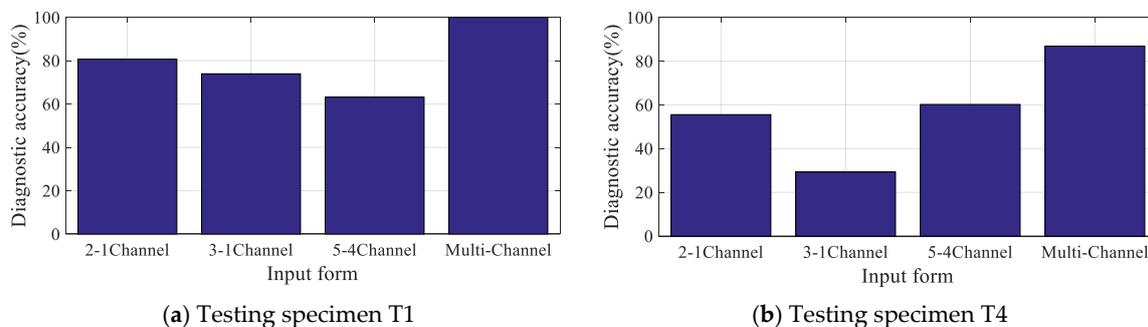


Figure 15. Diagnostic accuracies for typical specimens with different input forms.

(II). Traditional neural networks (backpropagation networks) and the softmax classifier are selected as comparative classification methods to show the advantages of designed CNN model. For one type, the neural number of single hidden layer are set which has the same neural number (NN(1024)) and different neural number (NN(100)) as FCL1, respectively. For another type, the neural number of multiple hidden layer are set which has the same neural number (NN(1024-512)) and different neural number (NN(512-100)) as FCL1 and FCL2, respectively. The purpose of these settings is to compare the effectiveness of convolutional layers. The neural number of the softmax classifier is 19. It is worth noting that different classification methods are trained and tested with the same training and testing input samples.

The diagnostic accuracies and training time of testing specimens T1 and T4 are shown in Figure 16 and Table 6, respectively. In Figure 16, the proposed method achieves the highest diagnostic accuracy in both T1 and T4. That is to say, the reliable FCD is difficult to be realized only by full-connected neural networks because of the lack of expression ability. Table 6 shows that the training consumption of the proposed method has been dramatically reduced compared with traditional methods indicating high diagnostic efficiency of the proposed method.

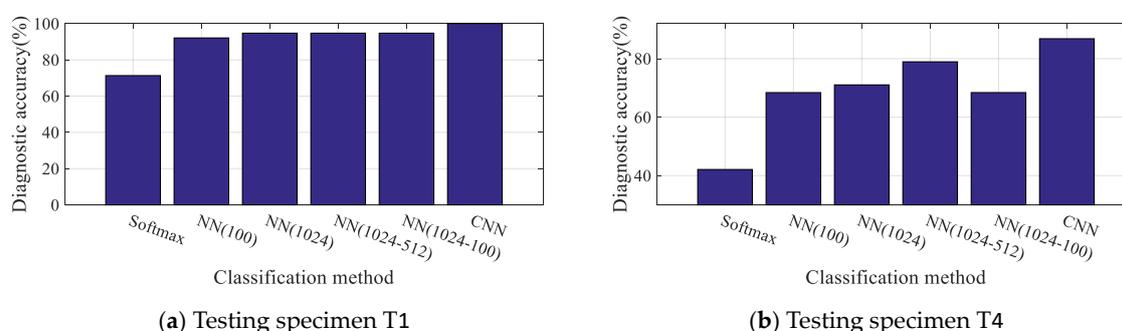


Figure 16. Diagnostic accuracies for typical specimens with different classification methods.

Table 6. Training time for typical specimens with different classification methods.

Classification Method	Softmax	NN (100)	NN (1024)	NN (1024-512)	NN (512-100)	CNN
T1	89 s	59 s	33 s	89 s	50 s	14 s
T4	35 s	117 s	117 s	86 s	57 s	9 s

4. Conclusions

This article puts forward a GW-CNN based approach to perform FCD under the influence of uncertainties due to different structures in real engineering applications. Firstly, multi-channel DIs under one crack length were extracted to construct an input feature vector to obtain more crack information for CNN to learn. Then, a CNN was designed to fuse multi-perspective fatigue crack information from the feature vector to realize reliable FCD. Moreover, data standardization, L2 regularization, and an Adam algorithm were adopted to improve the performance of CNN. Finally, CNN was trained with input vectors at different crack length from historical data. What is more, the proposed method was verified with attachment lug specimens. The lowest diagnostic accuracy was 86.84%, which has a diagnostic error of only 1 mm. Results show that the proposed method effectively reduced the influence of uncertainties on FCD between different structures. Finally, two sets of comparative experiments were carried out. Compared with the single-channel input sample form, the proposed method obtained much better diagnostic robustness and accuracy. Compared with the traditional methods, the designed CNN not only improves the accuracy of FCD, but also reduces the training consumption. There are some points for ongoing work to take verification on real structures or in operating conditions, considering the time-varying temperature and loading conditions.

Author Contributions: Conceptualization, L.X. and S.Y.; methodology, L.X.; software, L.X.; validation, L.X. and J.C.; formal analysis, L.X. and S.Y.; investigation, L.X.; resources, S.Y.; data curation, J.C.; writing—original draft preparation, L.X.; writing—review and editing, L.X., J.C., Y.R. and S.Y.; visualization, J.C. and Y.R.; supervision, S.Y.; project administration, S.Y.; funding acquisition, S.Y.

Funding: This paper is supported by the Key Program of National Natural Science Foundation of China (Grant No.51635008), National Natural Science Foundation of China (Grant No.51575263), Jiangsu Provincial Key Research and Development Program of China (Grant No. BE2018123), and the Research Fund of State Key Laboratory of Mechanics and Control of Mechanical Structures (Nanjing University of Aeronautics and astronautics) (Grant No. MCMS-0517K01), Priority Academic Program Development of Jiangsu Higher Education Institutions of China.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, L.; Yuan, S.F.; Chen, J.; Bao, Q. Deep learning based fatigue crack diagnosis of aircraft structures. In Proceedings of the 7th Asia-Pacific Workshop on Structural Health Monitoring 2018, Hong Kong, China, 12–15 November 2018.
2. Giurgiutiu, V.; Zagrai, A.; Bao, J.J. Piezoelectric Wafer Embedded Active Sensors for Aging Aircraft Structural Health Monitoring. *Struct. Health Monit.* **2002**, *1*, 41–61. [[CrossRef](#)]
3. Duffour, P.; Morbidini, M.; Cawley, P. A study of the vibro-acoustic modulation technique for the detection of cracks in metals. *J. Acoust. Soc. Am.* **2006**, *119*, 1463. [[CrossRef](#)]
4. Yuan, S.; Ren, Y.; Qiu, L.; Mei, H. A Multi-Response-Based Wireless Impact Monitoring Network for Aircraft Composite Structures. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7712–7722. [[CrossRef](#)]
5. Qiu, L.; Liu, B.; Yuan, S.; Su, Z.Q. Impact imaging of aircraft composite structure based on a model-independent spatial-wavenumber filter. *Ultrasonics* **2016**, *64*, 10–24. [[CrossRef](#)] [[PubMed](#)]
6. Su, Z.Q.; Zhou, C.; Hong, M.; Cheng, L.; Wang, Q.; Qing, X. Acousto-ultrasonics-based fatigue damage characterization: Linear versus nonlinear signal features. *Mech. Syst. Signal Process.* **2014**, *45*, 225–239. [[CrossRef](#)]
7. Michaels, J.; Michaels, T. Detection of structural damage from the local temporal coherence of diffuse ultrasonic signals. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **2005**, *52*, 1769–1782. [[CrossRef](#)] [[PubMed](#)]
8. Lu, Y.; Michaels, J. Feature Extraction and Sensor Fusion for Ultrasonic Structural Health Monitoring Under Changing Environmental Conditions. *IEEE Sens. J.* **2009**, *9*, 1462–1471.
9. Torkamani, S.; Roy, S.; Barkey, M.E.; Sazonov, E.; Burkett, S.; Kotru, S. A novel damage index for damage identification using guided waves with application in laminated composites. *Smart Mater. Struct.* **2014**, *23*, 95015. [[CrossRef](#)]
10. Mitra, M.; Gopalakrishnan, S. Guided wave based structural health monitoring: A review. *Smart Mater. Struct.* **2016**, *25*, 53001. [[CrossRef](#)]
11. Mei, H.; Yuan, S.; Qiu, L.; Zhang, J. Damage evaluation by a guided wave-hidden Markov model based method. *Smart Mater. Struct.* **2016**, *25*, 25021. [[CrossRef](#)]
12. Qiu, L.; Yuan, S.; Chang, F.; Bao, Q.; Mei, H. On-line updating Gaussian mixture model for aircraft wing spar damage evaluation under time-varying boundary condition. *Smart Mater. Struct.* **2014**, *23*, 125001. [[CrossRef](#)]
13. Qiu, L.; Yuan, S.; Boller, C. An adaptive guided wave-Gaussian mixture model for damage monitoring under time-varying conditions: Validation in a full-scale aircraft fatigue test. *Struct. Health Monit.* **2017**, *16*, 501–517. [[CrossRef](#)]
14. Gu, J.; Gul, M.; Wu, X. Damage detection under varying temperature using artificial neural networks. *Struct. Control. Health Monit.* **2017**, *24*, e1998. [[CrossRef](#)]
15. Gui, G.; Pan, H.; Lin, Z.; Li, Y.; Yuan, Z. Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. *KSCE J. Civ. Eng.* **2017**, *21*, 523–534. [[CrossRef](#)]
16. Taha, M.R.; Lucero, J. Damage identification for structural health monitoring using fuzzy pattern recognition. *Eng. Struct.* **2005**, *27*, 1774–1783. [[CrossRef](#)]
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
18. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.R.; Jaitly, N.; Sainath, T. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]

19. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
20. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
21. Shao, S.; Sun, W.; Zhao, R.; Yan, R.; Chen, X. Convolutional Discriminative Feature Learning for Induction Motor Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1350–1359.
22. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **2016**, *93*, 490–502. [[CrossRef](#)]
23. Grus, J. *Data Science from Scratch: First Principles with Python*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015; pp. 99–100.
24. Liao, B.; Xu, J.; Lv, J.; Zhou, S. An Image Retrieval Method for Binary Images Based on DBN and Softmax Classifier. *IETE Tech. Rev.* **2015**, *32*, 1–10. [[CrossRef](#)]
25. Smith, L.N.; Topin, N. Deep Convolutional Neural Network Design Patterns 2016. *arXiv* **2016**, arXiv:1611.00847.
26. Nielsen, M. *Neural Networks and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015.
27. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
28. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
29. Qiu, L.; Yuan, S. On development of a multi-channel PZT array scanning system and its evaluating application on UAV wing box. *Sens. Actuators A Phys.* **2009**, *151*, 220–230. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).