Vibration-based plastic-gear crack detection system using a convolutional neural network - Robust evaluation and performance improvement by re-learning

Kien Huy BUI*, Daisuke IBA*, Yunosuke ISHII*, Yusuke TSUTSUI*, Nanako MIURA*, Takashi IIZUKA*, Arata MASUDA*, Akira SONE* and Ichiro MORIWAKI*

* Kyoto Institute of Technology
Gosyokaidocho, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
E-mail: kien@pml.mech.kit.ac.jp

Received: 12 June 2019; Revised: 26 August 2019; Accepted: 11 October 2019

Abstract
This paper evaluates the sensitivity of a proposed crack detection method of POM (Polyoxymethylene) gears using a deep convolutional neural network. The vibration signal was collected from an automatic data acquisition system for endurance tests of gears. The fast Fourier transform (FFT) of the measured vibration signals generated grayscale images for training input. A high-speed camera captured cracks at the tooth root, and the length of cracks was computed as a damage index for training labels. A convolutional neural network (CNN), called VGG16 ConvNet, which has 1000 classes in output, firstly was pre-learned from image data of ImageNet and then the weights of two layers, which were close to the output layer, was relearned from the created images of meshing vibration data with the transfer learning technique. The output layer was modified to fit two classifications problem related to the cracked or non-cracked situation of gears. The accuracy rate for the recognition of the gear fault reached 100%. However, the remained problem is whether the performance of the developed system is susceptible to the change of the working condition of gear, such as high rotational speed and torque, or not. Hence, the robustness of the crack detection performance of the developed system was investigated. The endurance tests of gears under some test conditions, such as high-low rotational speed and/or torque, were carried out to collect the different vibration signals. The accuracy rate of gear failure classification under various working condition was judged and the factors affected on the performance of the developed system under working condition changing was discussed. The results showed that the developed system learned from one testing condition incapably perform in varied testing conditions. In other words, the developed system must be learned from diversity data for a superior effectuation. In this case, our interest is to uncover how many experiments and images for re-learning are required in each experiment for better performance. The investigation of the re-learning in this paper showed the required number of images was 200 and a single endurance test under each condition was enough if an appropriate number of images were obtained.

Keywords: Plastic gear, Endurance test, VGG16 ConvNet, Robustness, Varied working condition

1. Introduction

Failure detection of gears in transmissions plays a critical role in the development of safety machine systems. Thus, as is well known, many studies of gear health monitoring have been carried out (Iba et al., 2013). Despite the accumulation of much knowledge about metal gear monitoring, failures of plastic gears remain largely unknown. Therefore, in order to obtain more information on the literature on health monitoring methods of plastic gear, a vibration acquisition system was constructed in our laboratory (Ishii et al., 2018). The long-term purpose of this project is to design and publish a large-scale dataset of failure modes on plastic gears. Until now, 30 endurance tests in stable testing condition were carried out and a data set of crack and vibration information was created. A large number of these collected vibration data enable to apply achievements of deep neural networks (DNNs) to monitor the situation of the plastic gear.

In recent years, DNNs attracts growing attention from researchers (Liu et al., 2017). A convolutional neural network (CNN), as one of the main types of DNNs, becomes the dominant approach for almost all recognition and detection tasks in image and speech analysis (Gu et al., 2018). From the evolution of the classic CNN architectures inspired by
Krizhevsky et al., a typical trend is that the CNNs are getting deeper with a sequence of convolution and some new CNN configurations (Krizhevsky et al., 2012). For example, VGG16 ConvNet has 16 layers in the network, which was designed by Simonyan and Zisserman at ILSVRC (ImageNet Large Scale Visual Recognition Challenge) held in 2014 for large-scale image recognition (Simonyan et al., 2015). This DCNN model gains significant improvements upon previous methods on the image classification task. Moreover, open-source software libraries, such as TensorFlow and Keras by Google, are available for the development of deep learning systems using VGG16. In addition, the “transfer learning” technique enables to use of a pre-learned VGG16 from ImageNet data set including 1.2 million images (Gopalakrishnan et al., 2017). In the technique, this model then was modified to fit own problem and relearned from new limited images. This technique does not require enormous images and computational time.

The CNNs have been proved effective in machinery health monitoring (Zhao et al., 2019). Various researchers have achieved success in applying CNNs to gear or gearbox health monitoring. For example, Wang et al. proposed a DCNN model learned from 2D time-frequency images using Wavelet transform (Wang et al., 2017). Four classes of steel gear faults are classified with the accuracy rate is 99.58% on dataset contained 480 images. Weifang Sun et al. proposed an enhanced CNN using features acquired from dual-tree complex Wavelet Transform to classify four classes of steel gear fault in gearbox (Sun et al., 2017). The accuracy of the method is 99.80% on dataset contained 2520 records of signals. In our work, crack detection of plastic gears system using a DCNN based on the pre-learned VGG16 was constructed. The VGG16 was modified to fit the two classes related to crack or healthy situation of plastic gears. The weights of the last two layers were relearned with transfer learning using labelled images which were visualized from vibration signal using fast Fourier transform (FFT). The accuracy rate for the recognition of the gear fault reached 100% for a single test condition. While almost published methods achieved high accuracy in the gear failure classification under unchanged working condition, the effect of working condition changes, such as high-low rotational speed and/or torque on the performance of the CNNs have been unknown. Furthermore, in order to effectively use the monitoring system in applications, it is required that the system is robust against the varied working condition. In other words, if the system had ineffectively worked, which measures would have been tackled, such as data quality improvement or construction modification.

Hence, the robustness of the crack detection performance of the developed system against various working condition is firstly investigated. The required number of endurance tests of POM gears under some testing conditions, such as high-low rotational speed and/or torque, are considered. Additionally, the mandatory number of images for efficient learning of the system is determined. The accuracy rate and cross-entropy in image classification tasks of the system are employed to assess the robustness of the system. Some effects of testing condition change on the accuracy are discussed. Finally, required measures are taken into account to improve the performance of the system under varied testing condition.

2. Experimental related works

In this section, an acquisition of meshing vibration data and images pictures of crack propagation at the tooth root of plastic gears has been explained. The endurance test of POM spur gears was carried out on a gear operating test rig under the unchanged testing condition of rotational speed and torque. The testing conditions were changed to make different vibration data and were evaluated to show which is good testing condition span for an endurance test. Acquired vibration signals collected from the acquisition system are processed by FFT and then visualized as grayscale images for learning by VGG16.

2.1 Gear operating test rig

The gear operating test rig has been described in Fig. 1. The endurance tests of the plastic gear were carried out under unchanged testing condition, which was monitored by a torque meter during the whole test. The vibration signal was collected from an accelerometer installed on the housing of bearings and the driven shaft phasing was counted by an angular velocity sensor. A high-speed camera observes cracks occurring at the root of the teeth, and the crack signs at the root are detected. A steel gear is the driving gear and the POM gear is the test gear. The dimension of gears is shown in Table 1.
Fig. 1 Gear operating test rig. In this figure, a torque meter ① is used for testing condition monitoring of rotational speed and torque. An accelerometer ② installed on the housing of bearings to collect vibration signal. An angular velocity sensor ③ counts the phase of the driven shaft. A high-speed camera ④ observes cracks occurring at the root of the teeth. The steel gear ⑤ is driving gear and the POM ⑥ gear is test gear.

Table 1 Specification of master and test gear.

|                         | Master gear | Test gear |
|-------------------------|-------------|-----------|
| Module (mm)             | 1.0         |           |
| Pres. angle (deg)       | 20.0        |           |
| Number of teeth         | 67          | 48        |
| Helix angle (deg)       | 0           | 0         |
| Prof. shift coef.       | 0           | 0         |
| Facewidth (mm)          | 10.0        | 8.0       |
| Tip diameter (mm)       | 69.0        | 50.0      |
| Root diameter (mm)      | 64.5        | 45.5      |
| Material                | Steel       | POM       |

2.2 Data acquisition system

A data acquisition system developed in the previous study (Ishii et al., 2018) was used to collect vibration and image data. Every 1 minute, the vibration signals are collected at 100kHz of sampling frequency and the situation of gear are captured by the high-speed camera. Figure 2 shows an example of the captured images of crack growth at the tooth root of plastic gear in an endurance test. The existence of cracks in the captured images has been detected, and the label of crack or healthy situation of gears was generated and combined with the vibration signal to visualize as labelled images for learning data set.

Fig. 2 Captured images of crack growth. The cracks occurring at the root of the tooth gear can be captured by a high-speed camera every 1 minute. The figure shows an example of 3 situations of plastic gear, such as healthy, initial crack and heavy crack. However, there are two situations (with or without cracks) are used as labels for training data.
2.3 Endurance test under varied testing condition

In our previous research on plastic gears, we have performed the endurance test under a single testing condition, 7Nm of torque and 1000rpm of rotational speed to monitor the situation of gears during testing time from healthy to completely broken. In order to reveal the potential of the developed crack detection system working under varied testing conditions, the other endurance tests under various testing conditions were carried out. Table 2 shows the number and average testing time of the endurance tests under 9 divergent testing conditions.

Table 2  Endurance test situation.

| Torque and rotational speed [Nm] & [rpm] | Number of endurance test | Average testing time [minute] |
|----------------------------------------|--------------------------|-----------------------------|
| 4Nm, 500rpm                            | 1                        | >20000                      |
| 4Nm, 1000rpm                           | 2                        | 6731                        |
| 4Nm, 1500rpm                           | 4                        | 6440                        |
| 7Nm, 500rpm                            | 2                        | 1752                        |
| 7Nm, 1000rpm                           | 30                       | 296                         |
| 7Nm, 1500rpm                           | 2                        | 212                         |
| 10Nm, 500rpm                           | 5                        | 221                         |
| 10Nm, 1000rpm                          | 2                        | 87                          |
| 10Nm, 1500rpm                          | 3                        | 36                          |

In this work, we firstly carried out 30 endurance tests under the testing condition with 7Nm of torque and 1000rpm of rotational speed. The average testing time in this testing condition is 296 minutes. Then, the neighbouring testing conditions are set up to perform the other limited number of endurance tests. The relationship between the testing condition and testing time of plastic gear in Fig. 3 shows the challenge of carrying out endurance tests under varied testing condition. The endurance tests under high torque and rotational speed consume short testing time, e.g., the endurance test under 10Nm and 1500rpm finished after 36 minutes. Otherwise, the endurance tests under low torque and rotational speed required a long time of endurance testing, e.g., the endurance test under 4Nm and 500rpm finished after more than 20,000 minutes of testing time. This trouble limited us to collect big data under varied testing condition. However, the assessment of the learning of the system with different numbers of images in the next section will prove that the limited number of endurance tests is sufficient.

Fig. 3  The relationship between testing condition and testing time of plastic gear in the endurance test. The vertical axis shows 9 testing conditions, which were set up to carry out endurance tests. The horizontal axis shows the average testing time, excluding 4Nm of torque and 500 rpm of rotational speed with the average testing time is more than 20,000 minutes.
3. Crack or non-crack detection system

In this section, a combination of FFT and a DCNN is explained for crack detection system of plastic gears. The method to visualize vibration signals as images and the construction of a DCNN based on pre-trained VGG16 are presented, followed by the result of learning on crack or non-crack images classification of trained DCNN.

3.1 Visualized image from vibration signal

Fault detection directly from the time series, e.g., raw vibration signals measured from the endurance test of gear, is difficult. Various methods are used to process these time series into a frequency or time-frequency domain. A traditional method like FFT has proved sufficient to extract features from vibration data and reduce computational timing requirements. Hence, in this study, FFT is employed to visualize vibration signals as grayscale images for learning by the FFT spectrum peak picking method as can be seen in Fig. 4 (Ishii et al., 2018).

![FFT spectrum peak picking method](image-url)

Fig. 4  The FFT spectrum peak picking method. In this figure, a raw vibration signal was converted into a frequency spectrum. The peak amplitudes of frequencies on the frequency spectrum are picked to a matrix, which is visualised as grayscale images. The crack or non-crack labels of vibration signal detected by high-speed camera are assigned for grayscale images respectively.

In this study, every 1 minute, the vibration acquisition system can collect 10-second long vibration signal which is divided into tenths. Each divided 1-second vibration signal with 100,000Hz of sampling frequency is converted to the frequency domain. The peaks of amplitudes of frequencies were picked every one rotational speed frequency (RSF) from 0Hz to the 2nd order gear meshing frequency (GMF) to a matrix in a zig-zag way. Besides, the phases of the driven shaft are also collected to another matrix. The combined matrix including amplitudes and phases information was visualized as a grayscale image. The created images, which was labelled as “crack” or “non-crack” based on checking the situation of gear by high-speed camera, are learned by DCNN. For example, we consider an image created under 7Nm and 1000rpm of testing condition in Fig. 4. As the rotational speed is 1000rpm and the number of teeth of plastic test gear is 48, the 2nd order GMF is 1600Hz and the RSF is 16.67Hz. In this method, the first pixel of the image, which located in the top-left corner of the image, expresses the amplitude of frequency at 16.67Hz. Likewise, amplitudes of next frequencies were arranged in zig-zag way following the direction of arrows. It is noted that the amplitude of the 1st order GMF at 800Hz, which is the highest value, was represented by the white pixel as can be seen in Fig. 4. Finally, the amplitude of the 2nd order GMF at 1600Hz was represented by the grey pixel located the top-right corner of the amplitude area in the image.
3.2 Deep convolutional neural network

In this subsection, a “crack” or “non-crack” image classification system based on a pre-trained VGG16 model and transfer learning technique is described. The learning result of the system from labelled grayscale images is examined.

3.2.1 VGG16

VGG16 model is a deep convolutional neural network proposed by K. Simonyan and A. Zisserman at ILSVRC (ImageNet Large Scale Visual Recognition Challenge) held in 2014 [6]. This famous model achieves 92.7% top-5 test accuracy in this challenge. It makes the improvement over previous neural networks on the image classification task. This VGG16 model now publicly available in the Keras and allowed users to apply in their own works. The architecture of VGG16 was shown in Fig. 5.

![Fig. 5] The architecture of VGG16 (Simonyan et al., 2015). This DCNN model has 13 convolution layers (orange blocks), five max-pooling layers (blue blocks), three fully connected layers (green blocks) and one softmax layer (output layer). The maximum size of input images is 224x224 pixels with 3 colour channels. The resulting number at last layer determine the class of input images.

3.2.2 Transfer learning technique

DCNN models have achieved high accuracy on image classification, however, required enormous computational requirements. For example, VGG16 including 138 million parameters was trained for weeks and was using NVIDIA Titan Black GPU’s. In practice, very few users train entire DCNN from scratch (with random initialization). Instead, it is common to pre-train a DCNN on a very large data set and then using the models with obtained weights and architecture. This work means “transfer the learning” of the pre-trained model to solve specific problems. Furthermore, the “fine-tuning” technique allowed to modify the pre-existing model to fit own problem. The combination of “transfer learning and fine-tuning” to modify VGG16 to crack detection system is described in Fig. 6. By this means, the learning time and required image number decrease dramatically.

![Fig. 6] VGG16 with “transfer learning and fine turning” (Gopalakrishnan et al., 2017). With this technique, the fixed weights have been pre-trained from ImageNet data set and the updated weights can be trained by labelled images visualized from vibration signal. The input layer was modified with a maximum image size of 50x50 pixels. The output layer was modified to two classes related to “crack” and “non-crack” images.
3.2.3 Result of re-learning

The constructed system for classification of the crack of plastic gear was re-learned from 1000 images (500 images of each class) created from 10 appropriate endurance tests under a stable testing condition of 7Nm and 1000rpm. The time taken to train the network was 1 second per epoch and around 3 minutes for the whole training procedure on an i9 7900X processor at 3.31 GHz with 64 GB memory. The transition of accuracy and error during relearning are shown in Fig. 7. The training accuracy reached 100% after 50 epochs and training cross-entropy downed to 0.0194 after 200 epochs. The results proved that the developed system can be learned effectively from grayscale images created from vibration signals.

![Fig. 7](image1.png)

**Fig. 7** The transition of accuracy and error during re-learning. The training accuracy achieved 100% after 50 learning epochs. The training cross-entropy, which compares the model’s training with the true probability distribution of training data, finished at 0.0194 after 200 epochs.

3.2.4 Impact of the number of images on learning

In this subsection, we discuss how many images is sufficient for learning by the constructed system. A data set including 1000 images created from 10 other endurance tests was prepared for training. 6 packets of image data contained a specific number of images, such as 20, 50, 100, 200, 300, 400, were chosen randomly from the data set. 100 images (50 images for each class) were used to validate the accuracy of the classification of the system after learning. The accuracy rates of learning from each limited number of images were evaluated in Fig. 8. The results showed that a limited number of images is sufficient for efficient learning of the constructed system. 200 images are enough for our case.

![Fig. 8](image2.png)

**Fig. 8** Accuracy rates of learning under a different number of images. The dashed line shows the training accuracy from each packet of image data, which included 20, 50, 100, 200, 300, 400 images. The solid line shows the validation accuracy of trained system on testing data set, which included 50 images for each class. The validation accuracy is low with learning from a small or large number of images. In this figure, the system got the highest 90% accuracy on classification task with learning from 200 images.
4. Robustness evaluation of developed system

In this section, the trained system, which learned from 1000 images collected under 7Nm and 1000rpm of testing condition in subsection 3.2.3, was employed to perform the classification on diversity data collected under varied testing condition. The validation accuracy rates show the developed system with learning from an unchanged testing condition like 7Nm and 1000rpm is robust against varied testing conditions or not. The effect of the testing condition changes, such as rotational speed or torque, on the performance of the system is discussed.

4.1 Performance evaluation

900 images (50 images of each class under every one testing condition) created from 9 endurance tests under 9 testing conditions, three torque conditions, 4Nm, 7Nm, 10Nm and three rotation speed conditions, 500rpm, 1000rpm, 1500rpm. The robust performance of the developed system trained by the single condition, 7Nm and 1000rpm, was evaluated through these data. The accuracy rate and cross-entropy on “crack” and “non-crack” classifying task under each testing condition are shown in Fig. 9. The results indicate that the developed system with learning from the unchanged testing condition is not robust against varied testing conditions.

![Fig. 9 Accuracy and cross-entropy of validation under varied testing condition.](image-url)

Overall, the performance of the system is high under the high testing condition of torque and rotational speed and vice versa under low testing condition. With unchanged rotational speed, the accuracy rates tend to increase with growing torque. Moreover, with unchanged torque, the accuracy rates show an upward trend with growing rotational speed. The accuracy rate in 7Nm and 1000rpm achieved an outstanding result. These results of the system recommend us to assess effects of testing condition on image quality.

4.2 Impact of testing condition on image quality

In general, image quality affects tremendously the performance of DCNNs (Dodge et al. 2016). Many research results show the effects of image quality such as blur, noise, contrast, etc, on the image classification system. However, the constant resolution quality of images, which is 12x16 pixels created from the proposed visualized image method, has no image distortion. The primary cause is the useful information contained in an image, which is closely related to the failure features. Iba et al. pointed out that the failure features of plastic gears appear not only at the gear meshing frequency and the higher-order frequencies but also at lower frequencies (Iba et al., 2013). In particular, the amplitudes of low frequencies are smaller than of the meshing frequencies, but many failure features appear in a low-frequency span (Iba et al., 2013). We named this low-frequency range, which is from 0Hz to the 1st order GMF, as sensitive frequency span (SFS). The ratio between the SFS and the monitoring frequency span (MFS), which is from 0Hz to the 2nd order GMF in the image creating method proposed in our previous study (Ishii et al., 2018), was compared to validate the training image quality in varied testing condition, as can be seen in Fig. 10. The ratio of useful information related to the SFS increases with the rise of rotation speed in a training image.
Fig. 10  The ratio of gear failure feature in a training image. In this figure, the captured area by the red-dash line is monitoring frequency span (MFS) of training image. Because of learning from 7Nm and 1000rpm, the MFS of training image is [0 1600]Hz. The captured area by the green-dash line is sensitive frequency span (SFS) in each testing condition. The SFS is [0 400]Hz, [0 800]Hz and [0 1200]Hz corresponding to 500rpm, 1000rpm and 1500rpm. The ratios of useful information in a training image are 25%, 50% and 75% respectively. It is clear that the information on failure features is most in high rotational speed (75%) and at least in low rotational speed (25%).

Furthermore, differences between predictive image quality and testing image quality additionally cause accuracy degradation. After training, the system uses the trained parameters to predict input images. If the input image were largely different from predictive images, the accuracy rates of classification would have reduced. The differences between predictive images and testing images under varied testing condition were shown in Fig. 11. The 100% ratio indicates that the testing images in 1000rpm of rotational speed are similar to the predictive images. The differences with 50% or 150% ratio cause accuracy degradation in the two remain rotational speed.

Fig. 11  Differences between predictive images and testing images. With learning from 1000rpm of rotational speed, the MFS of predictive images captured by the green-dash line is [0 1600]Hz. However, the MFS of testing images captured by the purple-dash line is [0 800]Hz, [0 1600]Hz or [0 2400]Hz corresponding to 500rpm, 1000rpm and 1500rpm. The ratios of differences between MFS of predictive images and MFS of testing image are 50%, 100% and 150% respectively. The 100% ratio indicates that the testing images in 1000rpm of rotational speed are similar to the predictive images. The differences with 50% or 150% ratio cause accuracy degradation in the two remain rotational speeds.
Besides the influence of rotational speed, torque change affects image quality by the amplitude of frequencies. In order to evaluate the effect of torque on image quality, we considered the amplitude change of the 1st order GMF during endurance tests. We compared endurance tests with constant rotational speed and varied torque. The amplitude changes of the 1st order GMF every 1 minute was presented in Fig. 12. The figure indicates that the difference of amplitude between 4Nm of torque and the two remain in our tests. This dissimilarity causes accuracy degradation in the testing condition of low torque.

![Fig. 12](image_url)  
**Fig. 12** The amplitude change of the 1st order GMF in endurance test under varied testing condition. In this figure, the amplitude of the 1st order GMF in the endurance tests under constant rotational speed and varied torque was compared. With 500rpm or 1000rpm of rotational speed, the amplitude changes in 7Nm and 10Nm are similar. With 1500rpm of rotational speed, the amplitude changes are similar in whole torque. The figure indicates the difference of amplitude change between 4Nm of torque and the two remain, especially with 500rpm or 1000rpm of rotational speed. This dissimilarity causes the accuracy degradation in testing condition of low torque.

### 5. Evaluation of learning from varied data

#### 5.1 Performance improvement

The developed system was not robust against the varied testing conditions when learning from data created under unchanged testing condition. However, the proposed system achieved an impressive result when learning from data created under varied testing condition. 1800 images (100 images of each class under every one testing condition) created from 9 endurance tests under 9 testing conditions were carried out to re-train the developed system. The same testing data set, which has 900 images from varied testing conditions, was used to evaluate the robustness of the new learned system. The result in Fig. 13 proved that the performance of failure detection by the developed system was dramatically improved when learning from the varied data. These results show that the proposed system ensures excellent performance against every test conditions after re-learning of the limited number of images (200 images) in each test condition.

![Fig. 13](image_url)  
**Fig. 13** Performance of a new learning system under varied testing conditions. The figure shows the accuracy rate and cross-entropy of a new learning system under 9 testing condition respected to 4Nm, 7Nm, 10Nm of torque, and 500rpm, 1000rpm, 1500rpm of rotational speed. In the testing condition of 10Nm and 1500rpm, the accuracy rate is minimum with 80%, while almost accuracy rates in the remain testing condition are high. This imperfect accuracy caused by endurance test quality would be improved in the next subsection.
5.2 Endurance test assessment

As can be seen in Fig. 13, the accuracy rate in 10Nm and 1500rpm was 80%, which is lower than the others. We discuss the cause in this subsection. Endurance test quality keeps a vital role in image quality improvement in our proposed method. The “good” endurance test can create high-quality images and vice versa. In order to assess the endurance test quality, we carried out 10 endurance tests under the unchanged testing condition of 10Nm and 1500rpm. Firstly, the amplitude changes of 1st order GMF were checked to determine “good” or “not good” endurance test, as can be seen in Fig. 14. The results show that almost endurance tests were “good” tests which have smooth amplitude change during operation. However, some “not good” tests have many dramatic amplitude changes. Currently, we were not able to identify the cause of the dramatic amplitude changes, the output of the accelerometer, however, included large changes like step response in the time history. These abnormal changes lead to mislabeling data or noise for learning. Thus, in addition to learning from varied data, the system must be learned from “clean” data created from “good” endurance test to achieve high performance.

![Fig. 14](image_url)  
Fig. 14  The amplitude change of 1st GMF in endurance tests under 10Nm and 1500rpm of testing condition. The amplitude of 1st order GMF was calculated every 1 minute during an endurance test. This figure shows an example of “good” or “not good” endurance tests. In good endurance tests, the amplitude changed smoothly from 0 to 1.2. In not good endurance test, the amplitude changed dramatically many times.

The performance of the system can get 100% with learning from “good” endurance tests, as can be seen in Fig. 15. In this work, 240 images (100 images for training and 20 for validation in each class) from the selected “good” endurance tests under 10Nm and 1500rpm. The training accuracy gained 100% and the validation accuracy achieved 100% after 15 epochs. These results show that the elimination of the “not good” data avoids the degradation problem after re-learning.

![Fig. 15](image_url)  
Fig. 15  Result of learning from “good” endurance tests. The training accuracy achieved 100% after 15 epochs and the training cross-entropy downed to 0.0014 after 50 epochs. The validation accuracy achieved 100% after 15 epochs and the training cross-entropy downed to 0.0015 after 38 epochs.
6. Conclusion

In this paper, an evaluation of vibration-based gear crack detection system using a convolutional neural network was concerned. Firstly, the construction of the proposed system was explained and the endurance tests under varied testing condition were carried out to collect diversity data for robustness evaluation. Then, the performance of the system on “crack” or “non-crack” image classification was evaluated through the accuracy rate and cross-entropy of image classification. The results showed that the system learning from data created from the unchanged testing condition was not robust against the varied testing condition. The reason was discussed and the impact of torque and rotation speed on the visualized images was explained. Then, we showed that the re-learning in each test condition improved the failure detection performance. The required number of images was 200 in each condition. Hence, a single endurance test under each condition was enough if an appropriate number of images were obtained.

References

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, Vol. 25 (2012), pp.1106-1114.
Daisuke Iba, Shuhei Ohmori, Junichi Hongu, Morimasa Nakamura, Ichiro Moriwaki, Failure detection of plastic gears based on comparison of fourier coefficients of a gear mesh vibration model by rectangle pulse train and frequency analysis of acceleration response, Transactions of the Japan Society of Mechanical Engineers, Series C, Vol. 79, Issue 808 (2013), pp.5138-5148 (in Japanese).
Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai, Tsuhan Chen, Recent advances in convolutional neural networks. Pattern Recognition, Vol. 77 (2018), pp.354-377.
Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, Proceedings of International Conference on Learning Representation (ICLR) (2015), pp.1-14.
Kasthurirangan Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhary, Ankit Agrawal, Deep convolutional neural networks with transfer learning for computer vision-based data driven pavement distress detection, Journal of construction and building materials, Vol. 157 (2017), pp.322-330.
Peng Wang, Ananya, Ruqiang Yan, Robert X. Gao, Virtualization and deep recognition for system fault classification, Journal of Manufacturing Systems, Vol. 44 (2017), pp.310-316.
Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, Robert X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing, Vol. 115 (2019), pp.213-237.
Samuel Dodge, Lina Karam, Understanding how image quality affects deep neural networks, 8th International Conference on Quality of Multimedia Experience (QoMEX) (2016), pp.1-6.
Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, Fuad E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing, Vol. 234 (2017), pp.11-26.
Weifang Sun, Bin Yao, Nianyin Zeng, Binqiang Chen, Yuchao He, Xincheng Cao, Wangpeng He, An intelligent gear fault diagnosis methodology using a complex Wavelet enhanced convolutional neural network, Materials, Vol. 10, No. 7 (2017), p.790.
Yunosuke Ishii, Daisuke Iba, Satoshi Miyamoto, Nanako Miura, Takashi Iizuka, Arata Masuda, Akira Sone, Ichiro Moriwaki, Automatic acquisition system of meshing vibration data and image pictures of crack propagation, Proceedings of the Machine Design and Tribology division meeting in JSME, Vol. 2018.18 IB1-6 (2018), pp.1-4 (in Japanese).