Applied Mathematics, Modeling and Computer Simulation
C. Chen (Ed.)

doi:10.3233/ATDE220042

Bearing Fault Location Based on Feature Visualization of Convolutional Neural Network

Zhipeng Dong, Lei Chen,Weixing Song, Yangshuo Liu and Jianshe Kang[1]
Army Engineering University of PLA ,Shijiazhuang, China

Abstract. In industrial production, bearing is one of the most common and indispensable key components. It is of great significance to prevent safety accidents caused by bearing failure, timely fault type identification, and accurate fault location positioning. In this paper, we propose a method for bearing fault location based on feature visualization of convolutional neural network (CNN). In this method, the bearing vibration signals are transformed into images and the Grad-CAM feature visualization method is used to find the feature regions of the images. Thus, the points of the feature regions are traced back to the one-dimensional signals to roughly locate the faults. It has been proved that the experimental results of bearing fault location by this method are in good agreement with the theoretical values.

Keywords. CNN, visualization, bearing, failure, Grad-CAM

1. Introduction

Since Yann LeCun of New York University proposed Convolutional Neural Network (CNN) in 1998, this method has made great achievements in handwritten number recognition and face detection. As a multi-layer perceptron, CNN adopts the way of local connection and weight sharing, which reduces the number of weights and makes the network easy to be optimized. Meanwhile, it reduces the complexity of the model and reduces the risk of overfitting. The advantage performance in the input of the network is the image of the more obvious, the image directly as network input, to avoid the traditional recognition algorithm of the feature extraction and data reconstruction of the complex process. In the process of two-dimensional image processing, CNN has a lot of advantages, such as extracting features including color, texture, shape and topology structure. The identification of displacement, zooming or other forms of distortion invariance applications by CNN also have good robustness and operation efficiency.

[1] Corresponding Author, J S KANG, Army Engineering University of PLA, Shijiazhuang, China; Email: jskang201206@126.com.

At present, due to the limited understanding of the internal operation mechanism of CNN, it is still difficult to guarantee the high accuracy and interpretability of the model at the same time. For example, although the model based on expert system is a highly interpretable system, the accuracy of the model is usually low [1]. If the residual module is added, the depth of the network can be extended to several hundred layers, which ensures the excellent performance of the model and improves the accuracy of model prediction while avoiding overfitting [2]. However, the model will become too complex to be explained.

In order to understand deeply about how the functions of the middle feature layer in neural network changes with the training process, and to evaluate the model better, people introduced visualization technology in recent years. Zeiler et al. proposed a feature visualization technology based on deconvolution, which can adjust the network visually and improve the accuracy . B. Zhou et al proposed a technique called Class-Activated Mapping (CAM), which can find the region in the image for CNN to make classification basis, but the model does not include the full connection layer, which has certain requirements on the structure of the model [3]. By contrast, Ramprasaath R. Selvaraju proposed that based on the gradient of the positioning for visual interpretation from the depth of the network, no longer limit model structure, wider application range, cannot change the existing model, under the architecture of deep to explain model of decision-making under the guarantee the accuracy of the models, to achieve the interpretability of the model [4]. These methods are the development of the CAM, but no limit to model structure, suitable for multimodal input at the same time.

Visualization not only explains the inner work of CNN to some extent, but also has a broad application space. Chao Ma et al. proposed to use the rich hierarchical features of deep CNN to improve the accuracy and robustness of visual tracking [5]. Q.Wang et al. attempted to understand the deep learning network from the function of the primate visual cortex and separated the methods of extracting and perceiving similar features in V1 (primary visual cortex) and V2 (secondary visual cortex) regions by using multi-scale features of the convolutional layer and boundary ownership detection based on structured random forest [6]. Michael Elad et al. proposed a new solution to the style conversion problem based on texture synthesis, which can achieve effective style conversion [7]. Z.Gu et al. proposed a context-coding network (CE network) to capture the high-level information and retain the spatial information for 2D medical image segmentation [8]. Y. Liu and others introduced Rich Convolution Feature (RCF) as a new CNN structure, which made full use of feature hierarchy in CNN for edge detection [9]. E. Prasas et al. visualized the feature mapping of image transformation CNS, and briefly discussed the difficulties of visualization of image transformation networks and their differences with image classification models. They also combined several strategies to normalize the activation maximization algorithm [10]. Ramprasaath R.Selvaraju proposed a visual interpretation technique, namely grade-weighted class-activation mapping (Grad-CAM), which uses the information of the last convolutional layer of the network to locate the region in the input image that is used by the neural network to make the classification basis through the gradient weighting method [11-13]. Different from feature visualization and CAM based deconvolution, Grad-CAM provides completely different visualization results for the same image with different labels. This method can explain the model better, and does not need to specifically change the model structure or retrain the model, so it is applicable to a wider range.

2. Experimental Data Preprocessing

In order to validate the effectiveness of machine learning visualization technology in bearing fault diagnosis, two different types of data were selected from the motor bearing dataset of Case Western Reserve University. One is the 2K drive end bearing fault data, which contains 5 different types of faults, the other is normal baseline data on the 2K driver side. In the experiment, the sampling frequency is 12 kHz, the motor speed is about 1797 RPM, and the accelerometer is placed at the 12 o 'clock position of the drive end. Therefore, theoretically the acceleration sensor will collect about 400 sample points for each turn of the motor. The 400 sample points, the 100th point, the 200th point, and the 0th or 400th point approximately correspond to the direction of the bearing at 3 o 'clock, 6 o 'clock, and 12 o 'clock, respectively. The local data average decomposition method mentioned in ref 14 was selected to convert the original signal [14]. Since the signal arrangement was relatively regular and the original bearing vibration signal was periodic, image sizes of different fault positions of the same bearing fault type were selected in order to better present the characteristics of the transformed image. The data set is detailed in table 1.

Table 1. Faule types of bearing vibration singnals

	Dataset 1	**Dataset 2**
Bearing fault position information	@6:00； @3:00； @12:00	@6:00； @3:00； @12:00
The fault size	0.007mm	0.021mm
Operation status	hp=0	
speed	1797rpm	
Sampling frequency	12kHz	

As shown in figure 1, the one-dimensional vibration signals are normalized to 0 to 255 and successively filled into a two-dimensional matrix to transform into a two-dimensional image. After the transformation from one-dimensional time series to two-dimensional images is completed, pictures from different fault types are obtained. Then pictures of different sizes (such as 129*129, 99*99, etc.) are selected to obtain pictures with different sizes for the same fault type. For large-size images, the sequence length of the original vibration signal is fixed, i.e., the larger the image scale, the smaller the number of images were converted. In this way, we expanded the data set by rotating the image.

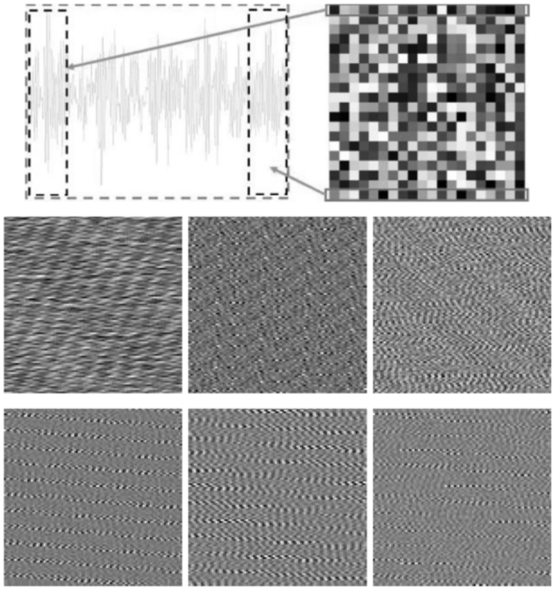


Figure 1. Transformed images.

3. Experimental Process

Fault diagnosis methods usually require the extraction of features from the original signal to integrate relevant information [15]. In this paper, Wen is used for reference [16]. A signal-to-image conversion method is proposed, and a new method is proposed as depicted in figure 2. In the experiment, the original image was rotated by 90 degrees to generate a new image, and the model in figure 3 was selected to train these images.

In order to obtain good visualization effect, the number of convolutional kernel and the network structure of CNN are slightly adjusted. Then, the model is trained in a way that not only requires the neural network meet the classification requirements, but also ensures the quality of the feature images obtained by the Grad-CAM algorithm. The first layers of the CNN contain the color and texture features of the image, the latter layers of the CNN feature information is relatively abstract, and the last layer of the CNN contains rich semantic information. Many literatures show a lot of research on semantic information extraction of the last convolutional layer. In our experiment, the semantic information of the last convolution layer is extracted by Grad-CAM. In most of the literatures, heat maps are chosen to highlight important areas, and both the heat maps and the original images are RGB images. Since the image converted from time series is gray scale image, the gray scale image of heat map is selected as the feature image in this paper. For the thermal map, the brighter the area, the closer of the pixels to 255, that is to say, the greater the influence of this area on the final classification, then these areas represent the characteristic areas of the image, that is, the area of the CNN as the classification basis.

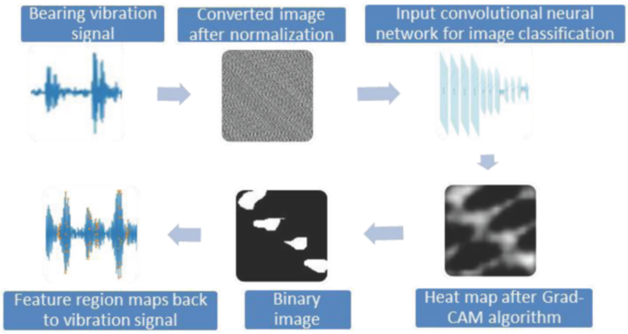


Figure 2. Fault location experiment flow chart based on feature visualization.

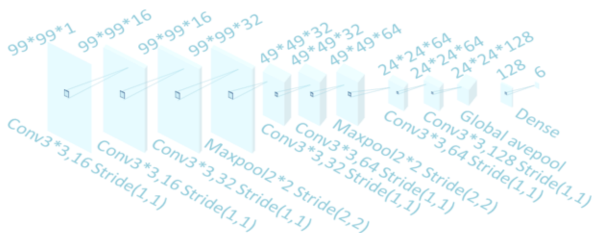


Figure 3. Schematic diagram of CNN structure.

Figure 4 shows the heat maps obtained by the Grad-CAM algorithm from the original images of different sizes converted by different bearing fault types. The white areas in the heat map also have a similar arrangement pattern to the original grayscale images. In order to obtain a better visual effect, an appropriate threshold is selected to convert these gray scale images to binary images. When selecting the threshold value, if the threshold value is too small, it will lead to too many white points, and there are a lot of redundant information in these white points, which is not conducive to the accurate fault location. On the contrary, if the threshold is set too high, there will be too few white points, and the feature points may be screened out. Therefore, the number of white dots in this experiment is set to be between one fourth and one third of the total pixel points. In addition, the isolated points in the binary image are removed to ensure that when these points are mapped back to the one-dimensional bearing vibration signals, all the signals are continuous sequence signals.

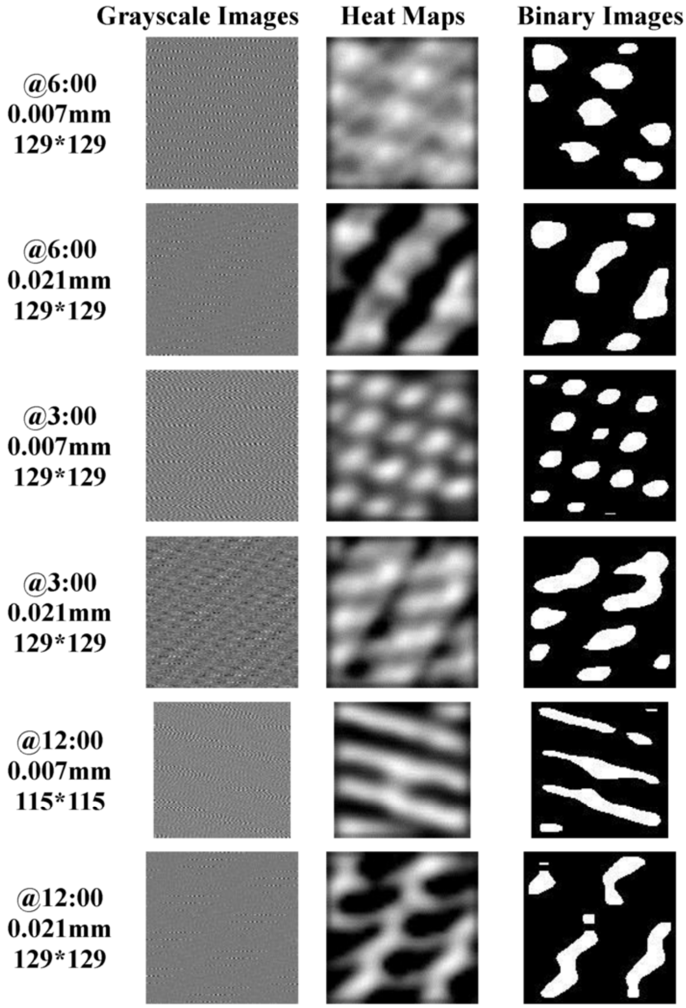


Figure 4. Heat maps and binary images of different sizes.

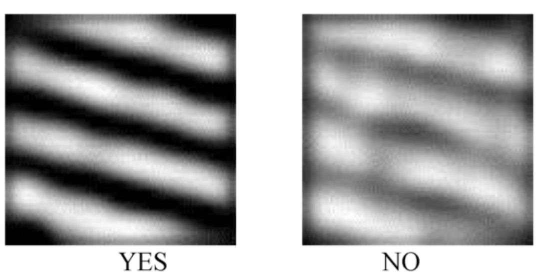


Figure 5. Selection of the appropriate heat map.

After the binary image is obtained and the sequence of signals in the converted image is known, it is possible to obtain the bearing vibration signal points that have a great influence on the final classification of the CNN. Firstly, the rough fault location was carried out with these points (figure 5 and figure 6). In order to recognize the texture features of the image by naked eyes, we selected a large-size image (129*129) as shown in figure 4. Secondly, the number of convolutional kernels in the convolutional layer of neural network was adjusted, the structure of the model was fine-tuned, and different models were trained. Several heat maps were obtained by Grad-CAM algorithm. Most of the texture distributions of these heat maps are similar to the texture features of the original image observed by the naked eye, which indicates that these heat maps have high reliability and can be used to extract feature information. The position of the points corresponding to the feature region in the image is not necessarily the position where the fault occurs, but the feature points must be hidden in it. The resulted sequence appears periodically according to certain rules, but it cannot be explained by the characteristic sequence. This method can indeed find the feature sequence, because in the cases that the bearing vibrates at a fixed speed, the feature sequence appears periodically in accordance with certain rules. Next, we use the special sequence points obtained from the above experiments to roughly locate the faults.

In the experiment, theoretically speaking, under the premise that the fault position of the bearing and the bearing speed remain unchanged, the corresponding fault feature sequence should appear once for each turn of the bearing. It means that the fault feature sequence should appear once every 400 points. With this assumption, after we found the previous feature points and identified their positions on the one-dimensional vibration signal, we tried to find the top 30 points with the highest occurrence frequency of 400 points per time (e.g., 1, 80, 280, 380, 401, 480, and 1201). Point 1 is more likely to be the failure point than the other points. In this paper, three kinds of fault locations in different positions are given. The experimental results are shown in table 2.

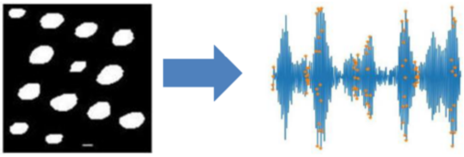


Figure 6. The points on the binary image map back to the original vibration signal.

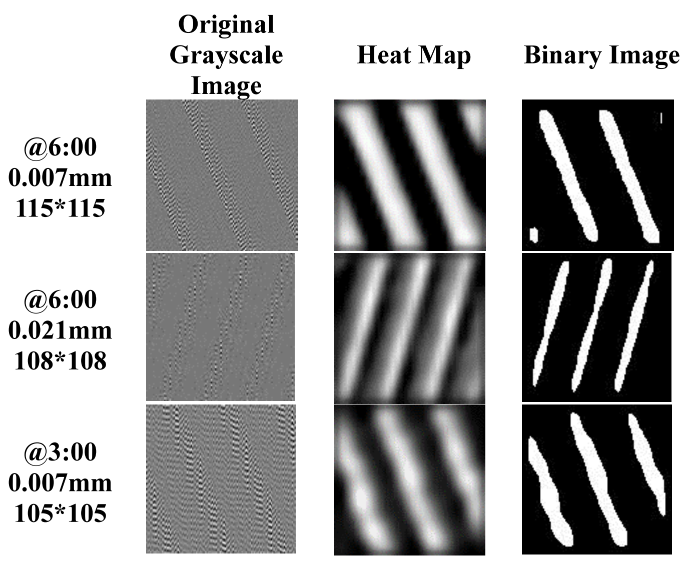
As can be seen from table 2, there are several potential failure points, which are the 130 points apart, 129 being exactly the size of the image, which is caused by the experimental method. If a point in the image is a feature point (i.e., occurs periodically every 400 points), then it is possible that the points above and below are also feature points. In addition, when one changes the size of the image, the image will appear a different texture as shown in figure 7.

Therefore, in order to locate the fault more accurately, images of different sizes were selected when the one-dimensional vibration signals were converted into two-dimensional images, and the difference of image sizes was ensured to be large (for example, 129 and 99).Use the same method to find these feature points and their locations. The experimental results are shown in table 3.

Taking the orthogonal (@3:00) (0.007) fault as an example, table 2 shows that the potential fault location is around 100, 230, and 360, while table 3 shows that the potential fault location is around 100, 200, and 310. According to the results from the two experiment, we concluded that the probability of the fault location is about 100 is large, which is in consistent with the actual value.

Table 2. Suspected fault locations

	Centered(@6:00) (0.007)	**Orthogonal(@3:00) (0.007)**	**Opposite(@12:00) (0.021)**
Suspected fault location	[79-82], [204-210], [328-340]	[102-106], [113-115], [232-235], [374-375]	[5-20], [137-148]
	[50-57], [190-194], [318-323]	[96-102], [225-231], [354-372]	[396-397], [0-8], [132-138], [260-272]
	[59-63]+128, [276-304]	[87-96], [10-108], [228-234], [358-356]	[114-126]-127, [248-257], [383-394]
	[45-55], [192-199], [315-321]	[87-92], [219-220], [225-234], [357-365]	[112-119], [250-256], [382-394]
Picture number	[1,6,11,14]	[1,6,11,18]	[1,6,9,10]
Image size	129*129	129*129	129*129



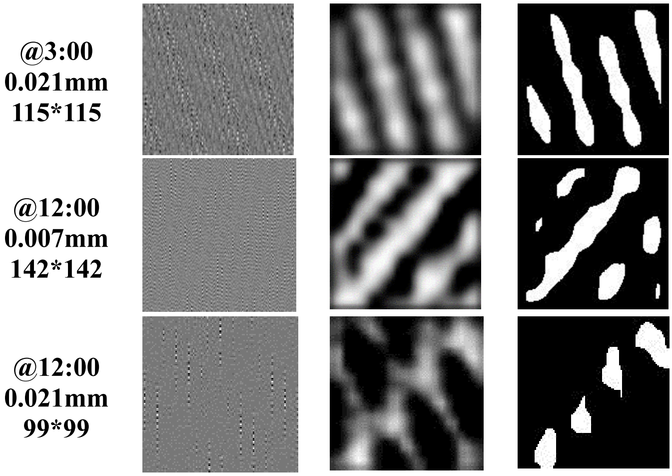


Figure 7. Heat maps and Binary images of different sizes.

Table 3. Suspected location of fault

	Centered(@6:00) (0.007)	**Orthogonal(@3:00) (0.007)**	**Opposite(@12:00) (0.021)**
	[30-35],	[24-28],	**[3-6]**,
	[92-98]+112,	**[88-94]**,	[104-110],
	[295-302],	[200-210],	[203-211],
	[320-330]	[310-320]	[302-308]
	[42-45],	[15-19],	**[94-100]-100**,
	[94-100],	**[103-109]**,	[192-201],
	[212-218],	[185-191],	[291-312],
Suspected fault	[324-330]	[302-306]	[396-398]
location	[28-36],	[22-27],	**[82-94]-99**,
	[142-149],	**[84-90]**,	[182-185],
	[204-212],	[196-201],	[286-292],
	[314-326]	[310-314]	[386-393]
	[68-70],	[46-55],	**[81-88]-100**,
	[84-92]+112,	**[96-198]**,	[182-193],
	[175-182],	[208-210],	[280-288],
	[290-296]	[334-344]	[382-388]
Picture number	[3,8,15,24]	[3,7,13,20]	[1,8,11,15]
Image size	114*114	114*114	99*99

4. Experimental Conclusions

Characteristics are proposed in this paper, based on the convolution neural network visualization technology of bearing fault location method, case western reserve university bearing data as data sources, the Grad - CAM feature visualization method, find the CNN features in the image area, through the characteristic points of the area of the back to one dimensional signal, the fault of rough localization, the experimental result is consistent with the theoretical value, improve the efficiency and precision of the signal analysis. The limitation of this method is that it is more sensitive to feature area larger images. When it comes to the smaller image of the region characteristic, this method is not so accurate for positioning. Also, since Grad-CAM method does not feature area accurately to the pixel, the fault location is still unable to obtain higher accuracy. Further improvements are needed in the follow-up research work.

References

[1] P. Jackson, Introduction to Expert Systems, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1998. https://www.osti.gov/biblio/5675197

[2] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[3] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba, "Learning deep features for discriminative localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 2921-2929. https://doi.org/10.1109/cvpr.2016.319

[4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), 2020; 128(2): 336-359. https://doi.org/10.1007/s11263-019-01228-7

[5] Chao Ma, Jia-Bin Huang, Xiaokang Yang, Ming-Hsuan Yang, "Robust Visual Tracking via Hierarchical Convolutional Features," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(11): 2709–2723. https://doi.org/10.1109/tpami.2018.2865311

[6] Qiurui Wang, Chun Yuan, Yan Liu, "Learning Deep Conditional Neural Network for Image Segmentation," in IEEE Transactions on Multimedia, 2019 July, 21(7): 1839–1852. https://doi.org/10.1109/tmm.2018.2890360

[7] Michael Elad, Peyman Milanfar, "Style transfer via texture synthesis," in IEEE Transactions on image processing, 2017 May, 26(5): 2338–2351. https://doi.org/10.1109/tip.2017.2678168

[8] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, Jiang Liu, "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," in IEEE Transactions on Medical Imaging, 2019 Oct, 38(10): 2281–92. https://doi.org/10.1109/tmi.2019.2903562

[9] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Jia-Wang Bian, Le Zhang, Xiang Bai, Jinhui Tang. "Richer Convolutional Features for Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence,. 2019 Aug, 41(8):1939-1946. https://doi.org/10.1109/tpami.2018.2878849

[10] Eglen Protas, Jose Douglas Bratti, Joel F. O. Gaya, Paulo Drews, Silvia S. C. Botelho, "Visualization methods for image transformation convolutional neural networks," in IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(7): 2231-2243. https://doi.org/10.1109/tnnls.2018.2881194

[11] Chuang Gan, Naiyan Wang, Yi Yang, Dit-Yan Yeung, Alexander G. Hauptmann, "Devnet: A deep event network for multimedia event detection and evidence recounting," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015 June, 2568-2577, https://doi.org/10.1109/cvpr.2015.7298872

[12] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. CoRR, abs/1312.6034, 2013. https://arxiv.org/abs/1312.6034v2

[13] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. Striving for simplicity: The all convolutional net. CoRR, abs/1412.6806, 2014. https://arxiv.org/abs/1412.6806

[14] Baojia Chen, Xuefeng Chen, Zhengjia He, Jiyong Tan, "Mechanical Fault Diagnosis Based on Local Mean Decomposition Method," 2009 International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, China, 2009, 681-684. https://doi.org/10.1109/icmtma.2009.373

[15] Rui Zhao, Dongzhe Wang, Ruqiang Yan, Kezhi Mao, Fei Shen, Jinjiang Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," in IEEE Transactions on Industrial Electronics, 2018 Feb; 65(2): 1539–1548. https://doi.org/10.1109/tie.2017.2733438

[16] Long Wen, Xinyu Li, Liang Gao, Yuyan Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," in IEEE Transactions on Industrial Electronics, 2018 Jul;65(7):5990-5998, https://doi.org/10.1109/tie.2017.2774777