PAPER • OPEN ACCESS

Audio Deformation based Data Augmentation for Convolution Neural Network in Vibration Analysis

To cite this article: M F M Esa *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012066

IOP Conf. Series: Materials Science and Engineering **551** (2019) 012066 doi:10.1088/1757-899X/551/1/012066

Audio Deformation based Data Augmentation for Convolution Neural Network in Vibration Analysis

M F M Esa[1] , N H Mustaffa[1] , N H M Radzi[1] and R Sallehuddin[1]

[1]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia 81310 UTM Johor Bahru, Johor, Malaysia

Email: pkharmony@yahoo.com

Abstract. Audio deformations in audio processing have proved ability in preserve semantic meaning for audio signal. Convolution Neural Network (CNN) is among deep learning model that requires huge dataset during training for excellence performance Thus, data augmentation (DA) method is used to overcome the problem of limited dataset number for vibration analysis. Several signal processing phases including segmentation and image converting need to be performed before the vibration signal can be used as input for CNN. In this research, audio-deformation based DA is proposed in generating the additional vibration signal dataset. The proses is start by encoding the raw vibration signal to audio signal format to enable the audio deformation process performing, then decoding back into new vibration signal. Speed and amplify transformation are selected for audio deformation process. The new vibration data set of bearing fault detection problem are used for training CNN to validate the proposed approach. The results obtained from 13 experiments setting have shown that the proposed DA able to increase the accuracy of training for CNN until 13% compared with the previous DA method.

1. Introduction

Bearing is significant mechanical component in rotary equipment and widely used in industrial application. The major fault in bearing may lead to the machinery malfunction and can interrupt the overall operation. The problem of bearing fault analysis has received increasing attention from the research community in recent years. The bearing fault analysis will identify the faulty bearing and its component in the system. Its applications range from rotating machine [1], helicopter gearbox and [2] electric locomotive [3] to wind turbine generator [4]. To date, variety of methods have been used for bearing fault analysis including the vibration analysis [5]. The vibration analysis is most commonly used due to its simple method of application [5].

In application of signal processing and machine learning, there are many techniques have been applied for vibration analysis and deep neural networks technique is the most recent technique used for the analysis [5]. In particular, Convolutional Neural Networks (CNN) is from the family of deep neural networks [6] and have advantages to solve the problem of bearing fault analysis. There are two significant characteristics of CNN that contributed to its successful. First, the capability of CNN in automatically extract complex feature representations [7] and second is the noise tolerance [8].

However CNN requires huge training dataset to be efficient, as they are composed of a lot of parameters [9]. Meanwhile, the collection of vibration signal data from defect bearing are usually limited because of expensive processing cost, laborious and depending to the device capabilities. These disadvantages will limit the application of CNN in vibration signal analysis for bearing fault detection. Fortunately, data augmentation (DA) able to solve these problem and been extensively discussed in literature [10]. The term DA refers to methods for sampling algorithms via the producing the unobserved data or latent variables in population data [11]. It is also, a common strategy adopted to increase the quantity of training data. Therefore, DA is simultaneously can generate extra size dataset and variant data at the same time preserves the data label. With these abilities, DA can worked excellently for image recognition and speech recognition [12].

Previously, DA also is a useful strategy to reduce over-fitting in CNN training for the image recognition problem [13] and in the same time achieved superior results at the time. In image processing, DA can be classified into geometric and photometric transformations [14, 15]. A geometric transformation is altering the geometry of the image to generate variant in position and orientation. For instance, flipping, cropping, scaling and rotating. Meanwhile, a photometric transformation is modification of the colour channels to generate variant in lighting and colour. Among the technique in this category are Fancy Principle Component Analysis (PCA) and colour jittering [13, 7].

In other domain, the other type of data format such as time series data has inspired DA technique from image processing community such as window slicing [16], overlap window slicing [8] and window warping [9]. In speech recognition, audio deformation is commonly been used as DA. Audio deformation is a method that artificially modified the raw sound signal to different tone or harmony of sound but preserve the semantic meaning of signal. The example of audio deformation are including time stretching, pitch shifting, dynamic range compression [17] and speed perturbation [18].

Generally, DA should be used indiscriminately [19]. Each type data requires different technique to suit with solution. Commonly, vibration signal is in the form of time-series data and previously, the data segmentation techniques are used to create maximum number of data record. Despite many DA technique have been developed in others domain, only sliding windows and overlap sliding windows techniques have been commonly implemented in vibration signal processing.

In this paper, we propose the use of audio-deformation based DA to overcome the problem of limited dataset number of vibration signal and explore different types of audio deformations and their influence on the CNN's training performance.

2. Methodology

The traditional vibration analysis consisting several sequence processes which start with preparing the vibration raw signal dataset. The long length of vibration signal will be segmented into shorter length than original. Furthermore, after segmentation signal and domain transformation process, these sets of 1-D time series are converted to image. Thus, these images will be used as input to CNN model and produces the classification result.

2.1. CNN Model

CNN is a multi-layer artificial neural network (ANN) that is can be consisting with of several filter layer and ending with classification layer [20]. The filter layer is used to extract features representation from the inputs data, that is mainly contains of three basic operations of kinds of layers, the convolutional layer, activation layer and the pooling layer. The classification stage is a multi-layer perceptron, which is composed of fully connected layers.

The experiments implement the Matlab2018b software. The CNN training parameters setting in experiments are as follows. For the learning algorithm, the stochastic gradient descent with momentum algorithm is applied. For every 5 epochs, the learning rate will be reduced by a 0.2 factor and the epochs maximum number is 20 at each iteration. Meanwhile, the CNN architecture setting are as follows. A First convolution layer with single filter with size [256,1] is used, continued by size 2 of max pooling. Then, a second convolution layer is design of 2 filters with the size [32,1] and followed by max pooling of size 4 is used. Later, a third convolution layer is make by 3 filters with the size [16,1] and consecutively a max pooling of size 4 is used. All activation layer using RELU activation function.

2.2. Audio Deformation based Data Augmentation

This paper proposes an audio deformation approach to generate DA for CNN in vibration analysis. With the appealing capability preserve semantic meaning after the signal was modified by the approach, the limited of data sources signal can be solved. Transformation at data level is achieved by encode the raw data of vibration signal into an audio file format. Followed decode audio file back to vibration signal after performed the audio transformation. Figure 1 shows the flowchart of the proposed audio deformation for DA method at raw data level.

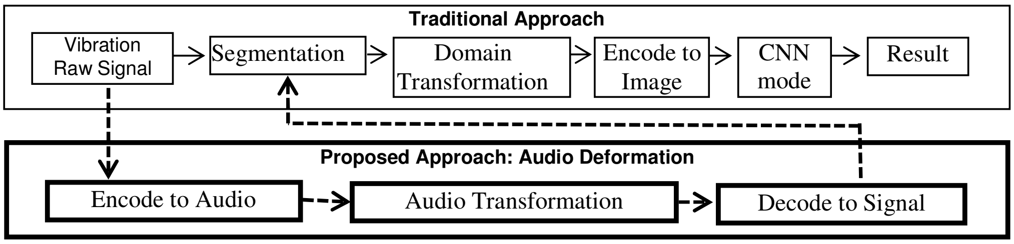

Figure 1. Flow chart of the proposed audio-deformation based DA.

Audacity® Software is used as tool to transform a raw signal in this study. Two audio effect function are selected to modify the raw signal are speed and amplify. Selection of the speed deformation due to its ability produce variant signal at frequency angle (axis x) meanwhile amplify deformation able to create variant signal at decibal(db) angle (axis y) of audio spectrum graph.

Basically, signal will be resampling in speed function meanwhile in amplify function will adjust the volume of audio signal. Five factor of deformation ware selected for each transformation in range 1% until 5%. Finally, five additional copies of the original training data were created by modifying each transformation function based on factor of the original rate.

2.3. Experimental Evaluation

To evaluate the effectiveness of the proposed approach for vibration analysis, the vibration signal dataset collected by Case Western Reserve University (CWRU) is analyzed. The dataset contains 9 defect and single good condition in various frequency, motor load, speed and bearing size. This study was selected a dataset with the properties are 48khz frequency, 1 hp load, 1772 rpm and 10 vibration signal that representing 10 defects condition bearing. From the raw signal, each signal will be executed the proposed technique and produces 12 additional signal where the first 10 additional are based on 5 different audio deformation factor and the rest are combination for all transformation factor for each audio deformation technique.

Then each signal included raw and additional signal are segmented using fix window technique with 2048 size window which was produced 200 sub signal for each signal. However, for mix transformation factor, total of sub signal for these signals are bigger 5 time than others.

All the generated signals are then divided into 2 part with ratios 70% training and 30% validation. Meanwhile, testing dataset are collected at 10% of an original raw data. Moreover, these setting are running 30 time each.

3. Results and Discussion

A similar CNN training and architecture parameter are used for all experiments in this paper. Figure 2 shows the result of 13 experiments of CNN training with speed deformation based DA and amplify deformation based DA in average.

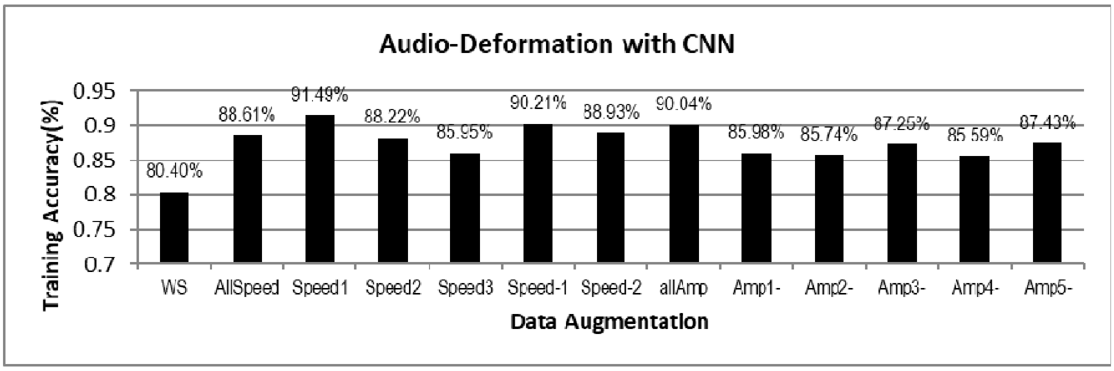

Figure 2. Comparison of CNN training accuracy with different factor of proposed approach.

The result shows that the proposed approach achieves higher training accuracy for all signals compared with window sliding (WS) DA due to ability of proposed approach to producing additional number and variant of dataset. The analysis has shown that the additional dataset using audio-deformation based DA has improved the efficiency of vibration analysis.

The result also shows that the fusion factor of speed deformation based DA is better than the fusion factor of amplify deformation based DA. However, a single speed deformation based signals perform better compared to the combination of all factor where the highest accuracy is achieved by speed with small factor either in positive or negative value. It is indicates that a fusion with various factor cannot guarantee a better result than the single factor.

4. Conclusion

In this paper, the audio deformation based DA for training CNN was proposed for vibration analysis in bearing fault detection. The idea is inspired from advantages of audio processing preserves sound meaning in various tone. Audio deformation is performed at raw data level to increase the CNN training accuracy by combining the raw signals to form the input of the CNN-based model.

Experimental studies on 10 kind of bearing defects verified the bearing fault analysis performance using vibration signal of the proposed approach for fault classification. The comparison between the proposed method and the traditional approaches shows that the proposed CNN-based method could achieve higher and more reliable diagnosis performance.

In addition, the capabilities of the proposed approach will allow for its extensive application in signal processing of various types of signals despite have limited in data sources and data generator methods. In the future, the fusion of different audio transformation will be considered in the analysis.

References

[1] Barszcz T *et al* (2018) *Advances in Condition Monitoring of Machinery in Non-Stationary Operations* (Springer,Cham) 235-224
[2] Zhou L *et al* (2019) *Appl. Acoustics* **147** 4-14
[3] Shao H *et al* (2018) *IEEE Trans. Ind. Electron.* ***65***(3) 2727
[4] Lu S *et al* (2017) *Mech. Sys. and Signal Process.* **85** 82
[5] Tandon N *et al* (1999) *Tribology Int.* **32** (8)
[6] Park C S (2013) *et al* (2013) *Mech. Syst. Signal Process.* **38(**2) 534–548
[7] Howard A (2013) Some improvements on deep CNN based image classification arXiv preprint arXiv:1312.5402
[8] Zhang W *et al* (2017) *Sensors* **17**(2) 425
[9] Guennec L A *et al* (2006) *ECML/PKDD workshop on advanced analytics and learning on temporal data* (Italy:HAL) 1-9
[10] LeCun Y *et al* (1990) *Adv. in Neural Inf. Process. Sys.* **2** 396–404

[11] Van Dyk D A *et al* (2001) *J. of Comp. and Graphical Stat.* **10**(1) 1-50.
[12] Rebai I *et al* (2017) *Procedia Comp. Sc.* **112** 316-322
[13] Russakovsky O *et al* (2016) *Adv. in Neural Inf. Process. Sys.* 3108
[14] He K *et al* (2014) *Comp. Vision ECCV 2014* (Berlin: Springer) 346
[15] Krizhevsky A *et al* (2012) *Adv. in Neural Inf. Process. Sys.* 1097
[16] Cui Z *et al* (2016) Multi-scale CNNs for time series classification arXiv:1603.06995
[17] Justin S *et al* (2017) *IEEE Signal Processing Letters* **24**(3) 279-283
[18] Ko T *et al* (2015) *16th Ann.l Conf. of the Int. Speech Comm. Ass.* 3586-3589
[19] Lemley J *et al* (2017) *IEEE Access* **5** 5858-5869.
[20] Lu W *et al* (2016) *IEEE Trans. Ind. Electron.* **64**(3) 2296-2305