UNSUPERVISED DETECTION OF ANOMALOUS SOUND FOR MACHINE CONDITION MONITORING USING DIFFERENT AUTO-ENCODER METHODS

Technical Report

Truong V. Hoang[1], Hieu C. Nguyen[2]

[1]STU.HCM, FPT Software, Ho Chi Minh, Vietnam
[2]FHN.DCS, FPT Software, Hanoi, Vietnam
truonghv1@fsoft.com.vn, hieunc3@fsoft.com.vn

Giao N. Pham[3]

[3]FWI.AAA, FPT Software, Hanoi, Vietnam
giaopn@fsoft.com.vn

ABSTRACT

Anomaly detection from the sound of machines is an important task for monitoring machines. This paper presents four deep learning methods to detect anomalous sound for machine condition monitoring using Long-short term memory auto-encoder, U-Net auto-encoder, Interpolation deep neural network, and Fully-connected auto-encoder. With experiments on the same dataset with the baseline system, experimental results show that our methods out-perform the baseline system in terms of AUC and pAUC evaluation metrics.

***Index Terms*—** Anomaly Detection, Anomalous Sound, Auto-Encoder, U-Net, LSTM, and IDNN.

1. INTRODUCTION

Currently, anomaly detection in sound (ADS) has been used for various purposes including audio surveillance [1], [2], animal husbandry [3], [4], product inspection, and predictive maintenance [5]. Because ADS is used to indicate the symptoms of mistakes or malicious activities, their prompt detection can possibly prevent such problems.

In order to solve ADS tasks, we could use the supervised methods and unsupervised methods in machine learning. But it is hard to detect anomalies from machine sounds by using the supervised methods because it is difficult to collect a large volume of anomalous sounds. The frequency of equipment failure in real environments is very low, and the number of ways in which equipment can fail is also very large. Thus, it is not feasible to collect a sufficient amount of training sound data corresponding to anomalous operating states. Therefore, these approaches are unsuitable for anomaly detection in sound [6].

To solve the difficulty above, we would like to propose the unsupervised methods for anomaly detection in sound. We proposed and experimented with four methods to detect the anomalous sound of machines. All four methods are based on the Auto-Encoder architecture. The first method is using the Long Short Term Memory (LSTM) Auto-Encoder. The second method is using the U-Net [7] Auto-Encoder architecture. The third method is using the "interpolation deep neural network" (IDNN) architecture [8]. The fourth method is a custom Fully-Connected Auto-Encoder. To clarify the proposed algorithm, our paper is organized as follows: In Sec. 2, we show the proposed method in detail. Experimental results and the evaluation of the proposed scheme will be shown in Sec. 3. Sec. 4 shows the conclusion.

2. THE PROPOSED METHOD

2.1. Used Features

There are two types of features used in our experiments. The first type of feature is a vector constructed from six features as Mel Frequency Cepstral Coefficient, Short-Time Fourier Transform, Chroma Features, Mel Spectrogram, Spectral Contrast, and Tonnetz respectively as shown in Fig. 1. These features are then used to compute the average feature of six features, and then construct 1D feature vectors as the input layer of the proposed method. This feature vector is used for the LSTM Auto-Encoder and U-Net Auto-Encoder. The second type of feature is the log-mel spectrogram originally provided in the baseline system [11]. The second type of feature is used for IDNN, Fully-Connected Auto-Encoder, and U-Net Auto-Encoder.

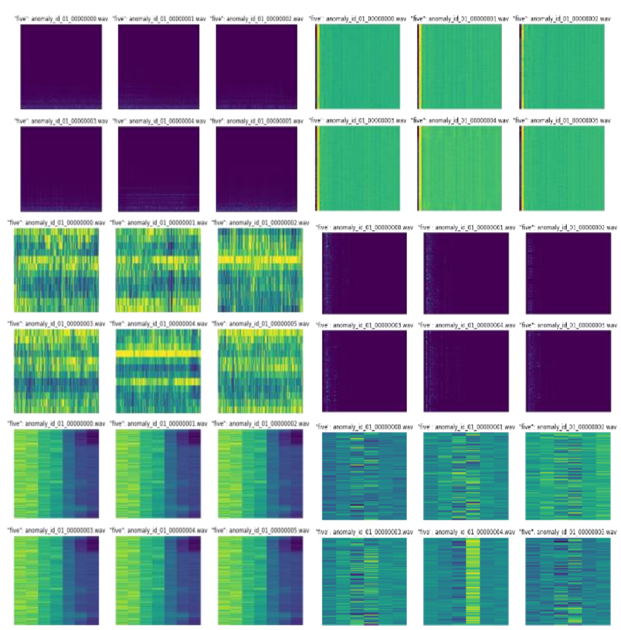

Figure 1. Used Features.

2.2. The Proposed Methods

2.2.1. LSTM Auto-Encoder Architecture

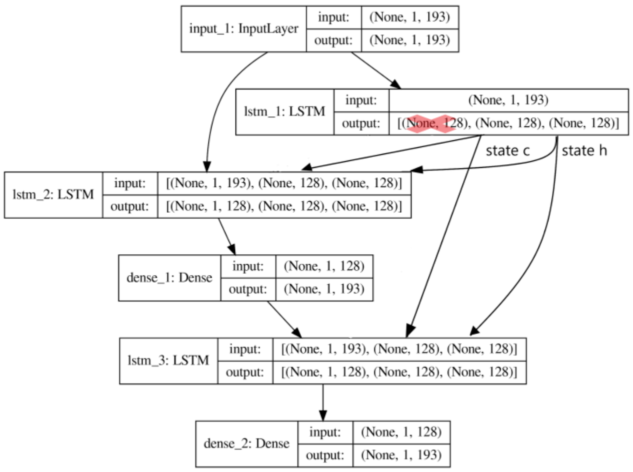

Figure 2. LSTM Auto-Encoder Method.

The first method is shown in Fig. 2. It includes six layers with one input layer, three LSTM layers, and two dense layers.

2.2.2. U-Net Auto-Encoder Architecture

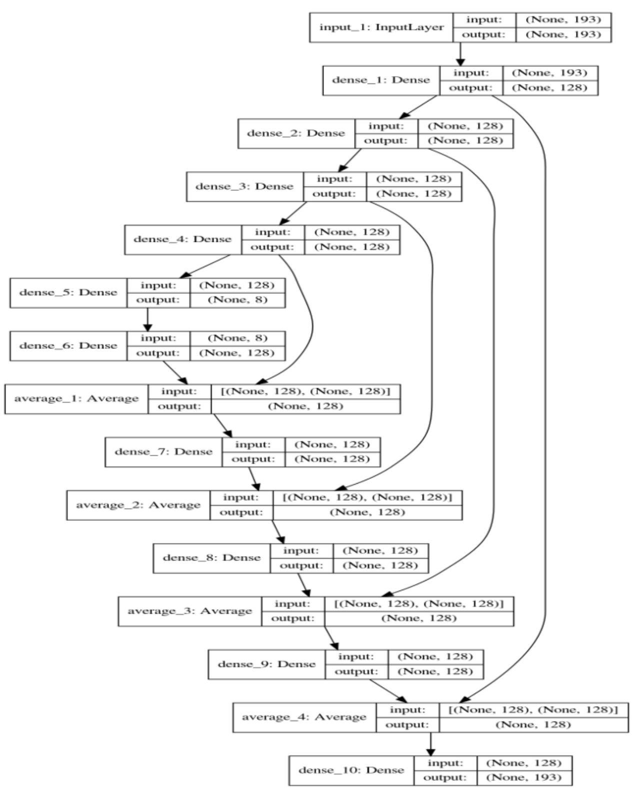

Figure 3. U-Net Auto-Encoder Architecture.

Figure 3 shows the second method with U-Net formed in Auto-Encoder structure. We build an Auto-Encoder based on the original U-Net [7]. The structure and number of units for each hidden layer is the same as the baseline auto-encoder [11], but with U-Net Auto-Encoder, we use Average layers instead of Concatenate layers to average between pairs of layers in encoder and decoder.

2.2.3. IDNN Architecture

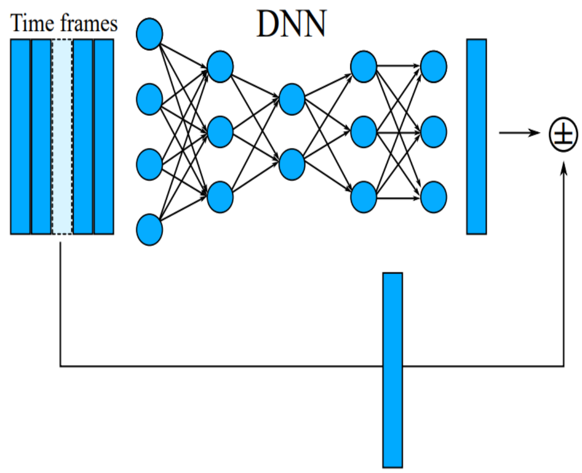

Figure 4. IDNN Structure.

The third method is the IDNN architecture [8], which we replicated from the authors' paper, described in Fig. 4. Each fully-connected layer is followed by a batch norm layer and a ReLU activation function.

2.2.4. Fully-Connected Auto-Encoder

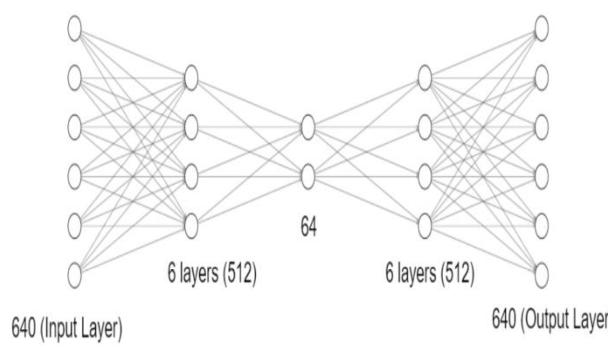

Figure 5. Fully-Connected Auto-Encoder Architecture.

The fourth method is our Fully-Connected Auto-Encoder architecture, which is described in Fig. 5. The encoder and decoder each consist of 6 hidden layers. Each fully-connected layer is followed by a batch norm layer and a ReLU activation function.

3. EXPERIMENTAL RESULTS

3.1. Dataset and Baseline System

In our experiments, we used the ToyADMOS [9] and the MIMII Dataset [10] consisting of the normal/anomalous operating sounds of six types of toy/real machines. Each recording is a single-channel 10-sec length audio that includes both a target machine's operating sound and environmental noise. The following six types of toy/real machines are used: Toy-car (ToyADMOS), Toy-conveyor (ToyADMOS), Valve (MIMII Dataset), Pump (MIMII Dataset), Fan (MIMII Dataset), and Slide rail (MIMII Dataset).

Our experimental results are used to compare to the result of baseline system. The baseline system is a simple auto-encoder-based anomaly score calculator [11]. The anomaly score is calculated as the reconstruction error of the observed sound. To obtain small anomaly scores for normal sounds, the AE is trained to minimize the reconstruction error of the normal training data.

3.2. Experimental Results

In Sec. 2.2.1 to 2.2.3, we used only the development dataset, with 90-10 train-validation split. In the training process, we used the Reduce LR On Plateau callback with factor 0.5, minimum learning rate of 10^{-4}, and patience 30. We also used the Early Stopping with patience 50. We trained 10,000 epochs, with the batch size of 512, the Adam optimizer with the learning rate of 10^{-3}, and in Sec. 2.2.3 we used the learning rate of 0.01. In Sec. 2.2.4, we used the development plus additional dataset, with 100% of training data. We used the Adam optimizer with the learning rate of 10^{-3}, with batch size of 4096.The loss function for all methods is the default Mean Square Error.

Table 1. Experimental Results

Machine Type	**AUC**		***p*AUC**	
	Baseline	**Proposed**	Baseline	**Proposed**
ToyCar	78.77	**89.78**	67.58	**75.60**
ToyConveyor	72.53	**76.57**	60.43	**62.29**
Fan	65.83	**75.05**	52.45	**60.01**
pump	72.89	**86.86**	59.99	**71.62**
Slider	84.76	**91.96**	66.53	**76.71**
Valve	66.28	**87.77**	50.98	**69.74**
Average	73.51	**84.67**	59.66	**69.33**

For this task, the evaluation metrics used are the area under the operating receiver characteristic (ROC) curve (AUC) and the partial-AUC (*p*AUC) [11]. From our experiments, LSTM Auto-Encoder is effective towards the ToyCar machine type; U-Net Auto-Encoder is effective towards the Fan, Pump, and Slider machine types; IDNN is effective towards the Valve machine types; Fully-Connected Auto-Encoder is effective towards the ToyConveyor and Slider development test dataset. Tab. 1 shows that all the AUC and *p*AUC scores of our different methods are higher than the AUC and *p*AUC of the baseline system.

4. CONCLUSION

In this paper, we proposed four methods to detect the anomalous sound of machines. The proposed methods are experimented on the normal datasets without abnormal data. The contributions of this paper are as follows:

- Proposed and experimented method LSTM Auto-Encoder for ADS
- Proposed and experimented method U-Net Auto-Encoder for ADS
- Proposed and experimented method IDNN for ADS
- Proposed and experimented method Fully-Connected Auto-Encoder for ADS.

Experimental results verified that our methods are better than the baseline system mentioned previously. In the future, we will try to develop algorithms that can even detect anomalous sounds that would be difficult, and we will tackle the following remaining issues of ADS systems in real environments.

ACKNOWLEDGMENT

This work is supported by FPT Software AI Committee, FPT Software, Hanoi, Vietnam.

REFERENCES

[1] Y. Li, X. Li, Y. Zhang, M. Liu, W. Wang, “Anomalous Sound Detection Using Deep Audio Representation and a BLSTM Network for Audio Surveillance of Roads,” *IEEE Access*, pp. 1-1, Oct. 2018.

[2] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, and M. Vento, “Audio Surveillance of Roads: A System for Detecting Anomalous Sounds,” *IEEE Trans. ITS*, pp.279–288, 2016.

[3] P. Coucke, B. De. Ketelaere, and J. De. Baerdemaeker, “Experimental analysis of the dynamic, mechanical behavior of a chicken egg,” *Journal of Sound and Vibration*, vol. 266, pp.711–721, 2003.

[4] Y. Chung, S. Oh, J. Lee, D. Park, H. H. Chang and S. Kim, “Automatic Detection and Recognition of Pig Wasting Diseases Using Sound Data in Audio Surveillance Systems,” *Sensors*, pp.12929–12942, 2013.

[5] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, “Optimizing Acoustic Feature Extractor for Anomalous Sound Detection Based on Neyman-Pearson Lemma,” in *Proc. of 25th European Signal Processing Conference*, pp. 728-732, Sept. 2017.

[6] Y. Koizumi, S. Saito, and H. Uematsu, “Anomalous Sound Detection for Machine Operating Sounds using Deep Neural Networks,” in *Proc. 2017 Spring Meeting of the Acoustical Society of Japan*, pp. 473–476, Kawasaki, Japan, Mar. 2017.

[7] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional Networks for Biomedical Image Segmentation,” in

Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241, Springer, 2015.

[8] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, “Anomalous sound detection based on interpolation deep neural network,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 271-275, 2020.

[9] Y. Koizumi, S. Saito, H.Uematsu, N. Harada, and K. Imoto, “ToyADMOS: A Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection,” in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 308–312, Nov. 2019.

[10] H.Purohit, R. Tanabe, T.Ichige, T. Endo, and Y.Nikaido, K. Suefusa, and Y. Kawaguchi, “MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation & Inspection,” in *Proc. of the Detection & Classification of Acoustic Scenes & Events 2019 Workshop*, pp. 209–213, Nov. 2019.

[11] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, “Description and discussion on DCASE2020 challenge task2: unsupervised anomalous sound detection for machine condition monitoring,” in *arXiv* e-prints: 2006.05822, 1-4. June 2020.