ENSEMBLE METHOD FOR UNSUPERVISED ANOMALOUS SOUND DETECTION

Technical Report

Jinhyuk Park, Sangwoong Yoon, Yonghyeon Lee, Minjun Son, Frank C. Park

Seoul National University, Robotics Laboratory
{jhpark, swyoon, YHLee, mjson}@robotics.snu.ac.kr, fcp@snu.ac.kr

ABSTRACT

We propose an anomalous sound detection method for DCASE2022 Challenge Task2. The method is basically an ensemble of multiple autoencoder-based approaches. The model reconstruct the input Mel spectrogram and decide it is an anomaly if the reconstruction error is higher than a threshold. The area under curve (AUC) performance achieved by the proposed approach is 53.35% on source domain and 43.48% on target domain, and the partial AUC is 48.00%.

Index Terms— Anomalous Sound Detection, Ensemble Method, Autoencoder

1. INTRODUCTION

The DCASE2022 Challenge Task2 [1] is to identify anomalous machine sound in general domain only with given normal machine sound data and so it is titled "Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques". The main issue of this challenge is that anomalous sound recordings are not given in training phase. This is a realistic constraint since the anomaly rarely occurs in real-world factories. The unsupervised learning setting makes it significantly challenging for the detection model to decide whether a sound recording is faulty or not. This unsupervised learning challenge succeed the DCASE2020 Challenge Task2 [2] and DCASE2021 Challenge Task2 [3].

The other important aspect of the challenge is the domain generalization. Domain generalization seeks to accomplish out-of-distribution (OOD) generalization by employing only source data in learning phase. A variety of methodologies have emerged as a result of the tremendous progress made in domain generalization research over the past ten years [4, 5, 6]. In challenge dataset, test data is different from training data with respect to the unknown factor other than abnormalities. For instance, operating circumstances of machine or ambient noise frequently vary in real-world situations and make difference between the training and testing phases. Specifically, the test data consists of samples that are unaffected by domain shifts (source domain data) and samples that are influenced by domain shifts (target domain data), and the domain of each sample is not stated. Note that the challenge from previous year specifies whether or not each recording in test data is from a shifted domain [3, 7].

We propose the anomaly detection method based on ensemble of multiple autoencoders. In the following sections, we describe the model architecture with its hyperparameters, training method, and the evaluated performance measure.

2. MODEL ARCHITECTURE

2.1. Autoencoder

The input of the model is log-Mel spectrogram with 128 Mel-frequency bands and 64 ms-long frame, which results in 128×64 intput dimension. The first model of the proposed method is an autoencoder composed of fully connected layers followed by batch normalization and ReLU activation function. The encoder network takes an input $x \in \mathbb{R}^{128\times64}$ and output latent vector $z \in \mathbb{R}^8$ through four fully connected layers which have $(128, 192, 192, 128)$ nodes respectively. The decoder network has the same four-layer structure as the encoder and takes as input the latent vector z and output $y \in \mathbb{R}^{128\times64}$ in order to reconstruct the original input sample x. Neither batch normalization nor ReLU activation is applied at the output layer of the model. The neural network model is trained by using Adam optimizer.

Similar to the first model, the second model is a fully connected autoencoder with more nodes: $(192, 256, 256, 192)$ for both encoder and decoder networks. The number of parameters of the first autoencoder model is 585,648 and the second model has 342,424 parameters. Whole structure of two models are shown in Table 1. Every models are implemented using Keras [8] and Tensorflow 2 [9].

When an input is a normal sound clip, the proposed model is able to reconstruct it perfectly since the feature of normal samples is learned by the model in training phase. However, when an anomalous sound clip is given as an input, the model fails to reconstruct the input and yields high reconstruction error. Therefore, by setting appropriate threshold θ, we could make a decision on each sample if it is normal or anomalous.

Table 1: Autoencoder model configuration

		Model 1	**Model 2**
	layer	node number	node number
	input	128×64	128×64
Encoder	fc1	128	192
	fc2	192	256
	fc3	192	256
	fc4	128	192
	fc5	8	8
Decoder	fc6	128	192
	fc7	192	256
	fc8	192	256
	fc9	128	192
	output	128×64	128×64

2.2. Ensemble method

Ensemble methods include building multiple models and combining them to achieve better outcomes. Generally speaking, ensemble approaches yield more precise results than a single model would. The ensemble technique has been commonly employed in the winning solution of a number of machine learning contests. We randomly initialize each autoencoder model several times in training and select best-performing three models. A total of six neural networks are combined so that the average predictions of all networks are calculated as final outputs. Multiple autoencoders complement each other and prevent the worst-case scenarios where overfitting or underfitting occurs in training phase.

3. EXPERIMENTS

3.1. Dataset

The audio data for DCASE Challenge Task2 is basically from ToyADMOS2 [10] and MIMII Dataset [11]. The operation sound of seven different types of machines—fan, gearbox, bearing, slide rail, toy car, toy train, valve—are selected for the challenge. Each recording is a 10-sec-long, single-channel audio clip including noises from the target machine and its surroundings. For more details, please refer to [1].

3.2. Results

The evaluated performance of the proposed algorithm for development dataset is shown in Table 2. The performance measure is given as the area under the receiver operating characteristic curve (AUC) and partial AUC (pAUC) with $p = 0.1$. We found that ensemble model shows better performance than single models.

4. CONCLUSIONS

In this report, we propose an anomalous sound detection algorithm for DCASE2022 Challenge Task2, which is based on ensemble of autoencoders. Each autoencoder is composed of fully connected encoder and decoder with four layers that reconstruct Mel spectrogram samples. The anomalous machine operating sound can be classified using reconstruction error of the proposed model. The performance measure of the proposed method in harmonic mean of AUC is 53.35% and 43.48% on source and target domain, respectively. The harmonic mean of partial AUC (pAUC) is 48.00%.

5. ACKNOWLEDGMENT

This work was supported by Daewoo Shipbuilding Marine Engineering, SNUIAMD, SNU BK21+ Program in Mechanical Engineering, and the SNU Institute for Engineering Research.

6. REFERENCES

[1] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, T. Endo, M. Yamamoto, and Y. Kawaguchi, "Description and discussion on DCASE 2022 challenge task 2: unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," *arXiv e-prints:2206.05876*, 2022.

[2] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, *et al.*, "Description and discussion on DCASE 2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," *arXiv preprint arXiv:2006.05822*, 2020.

[3] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on DCASE 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions," *arXiv preprint arXiv:2106.04492*, 2021.

[4] G. Blanchard, G. Lee, and C. Scott, "Generalizing from several related classification tasks to a new unlabeled sample," *Advances in neural information processing systems*, vol. 24, 2011.

[5] S. Shankar, V. Piratla, S. Chakrabarti, S. Chaudhuri, P. Jyothi, and S. Sarawagi, "Generalizing across domains via cross-gradient training," *arXiv preprint arXiv:1804.10745*, 2018.

[6] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *CoRR*, vol. abs/2103.02503, 2021. [Online]. Available: https://arxiv.org/abs/2103.02503

[7] J. A. Lopez, G. Stemmer, P. Lopez-Meyer, P. Singh, J. A. del Hoyo Ontiveros, and H. A. Cordourier, "Ensemble of complementary anomaly detectors under domain shifted conditions," in *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*.

[8] F. Chollet *et al.* (2015) Keras. [Online]. Available: https://github.com/fchollet/keras

[9] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[10] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions," in *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, Barcelona, Spain, November 2021.

[11] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "MIMII DG: sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task," *arXiv e-prints:2205.13879*, 2022.

Table 2: Autoencoder model configuration

	ToyCar	**ToyTrain**	**fan**	**gearbox**	**bearing**	**slider**	**valve**	**total**
h-mean AUC source	62.54	49.67	60.14	56.45	47.95	52.16	48.27	53.35
h-mean AUC target	35.76	38.84	50.14	55.55	46.64	37.95	46.41	43.48
h-mean pAUC	45.50	43.59	54.69	56.00	47.28	43.94	47.91	48.00