ENSEMBLE SYSTEMS WITH GAN AND AUTO-ENCODER MODELS FOR ANOMALOUS SOUND DETECTION

Technical Report

Zhonghao Zhao[1,2], *Yang Tan*[1,2], *Kun Qian*[1,2*], *Kele Xu*[3*], *and Bin Hu*[1,2*]

[1] Key Laboratory of Brain Health Intelligent Evaluation and Intervention, Ministry of Education (Beijing Institute of Technology), P. R. China
[2] School of Medical Technology, Beijing Institute of Technology, P. R. China,
{zhonghao.zhao, yang_tan, qian, bh}@bit.edu.cn
[3] National University of Defense Technology
{xukelele}@163.com

ABSTRACT

In this paper, we describe our submissions for DCASE 2023 Challenge Task 2. For solving anomalous sound detection problem, an ensemble system with gan and auto-encoder model are proposed. Spectrograms and log-mel energies are used to train models. As a result, the proposed systems achieved a better performance than the baseline models.

***Index Terms*—** DCASE, Anomalous Sound Detection, Auto-Encoder, Log Mel Energies

1. INTRODUCTION

Error-free operation of the machines guarantees good production quality and efficiency in industry. The timely detection and thus maintenance of the machines prevent human casualties and huge economic losses to companies [1]. In addition, real-time detection for machines greatly reduces the input of manpower. The main signs of a fault devices condition are abnormal sound, abnormal vibration and abnormal overheating. Abnormal sound detection (ASD) has been widely used in the industrial [2], medical [3] and environmental fields [4]. Mechanical operation is often accompanied by acoustic events. Tanglesome and heavy sound indicates possible equipment malfunction. In [5], acoustic features performed better than vibration features for fault detection of machines, while performed as well as heat at a lower cost for data acquisition with microphone. Therefore, the sound provides significant and inexpensive evidence for detecting the condition of machines.

The DCASE 2023 Challenge Task 2 has released a task to detect anomalous state of the target machine using the sound data [6]. Different from other tasks, this is a unsupervised task. Development dataset contains only seven types of normal machine sound (fan, gearbox, bearing, slide rail, toy car, toy train, valve) and is completely different with additional training dataset and evaluation dataset (Vacuum, ToyTank, ToyNscale, ToyDrone, bandsaw, grinder and shaker). There are four requirements that we need to comply with, which makes the task more challenge. First, the model is trained on the development dataset as rare abnormal sounds exist in the real world. Second, because machine anomalies occur at any time, we need to detect anomalies regardless of domain shifts. Third, hyperparameters are obtained during training stage and are forbidden adjust in the later stages due to the first requirement. Last, machine types and sound data are limited.

The official has proposed a simple auto-encoder (AE) based implementation combined with selective Mahalanobis metric as the baseline system [7]. Moreover, Jiang *et al.* [8] has released a AEGAN-AD model in which the generator (also an auto-encoder) was trained to reconstruct input spectrograms. The model performed best among all generative models provided the official with a general improvement of 3.84 %. To this end, we develop three methods based on the above two models to deal with this challenge.

The remainder of this article will be organised as follows: Firstly, we give the information of the used database in Section 2. Then, the experimental setup and results are presented in Section 3 and Section 4, respectively. Finally, we draw a conclusion in Section 5.

2. DATASET

All experiments are conducted on the databases provided by the DCASE 2023 Challenge Task 2, which contains development dataset for obtaining the hyperparameters, the additional training dataset for training the model and evaluation dataset for testing the model [9, 10]. The development dataset are completely different from the additional training and evaluation dataset. The data of fan, gearbox, bearing, slide rail and valve in the development dataset are from the MIMII DG dataset [9]. ToyCar and ToyTrain are from the ToyADMOS2 dataset [10]. The additional training dataset and evaluation dataset are required on the DCASE 2023 Challenge website, which includes seven types of sound, i. e. Vacuum, ToyTank, ToyNscale, ToyDrone, bandsaw, grinder and shaker. All sounds were sampled at a sampling rate of 16 kHz. Different from past challenges, this task is need to be performed under the conditions that the acoustic characteristics of the training data and the test data are different (i. e., domain shift).

*This work was partially supported by the Ministry of Science and Technology of the People's Republic of China with the STI2030-Major Projects (No. 2021ZD0201900), the National Natural Science Foundation of China (No. 62227807 and 62272044), and the Teli Young Fellow Program from the Beijing Institute of Technology, China. (*Corresponding authors: K. Qian, K. Xu, and B. Hu.*)

3. EXPERIMENT SETUP

3.1. Neural network architecture

We experiment three models to handle the task. First, we use the auto-encoder proposed by the official to do the classification task, which outperform in some audios than other auto-encoder networks. Then, an auto-encoder networks is improved to achieve better classification results based on the model. In addition, we also experiment a latest network, AEGAN-AD, which also has been designed and used for DCASE 2022 Challege Task 2 proposed by Jiang *et al.* [8].

3.2. Training setup

All experiments implementations are based on pytorch. In training stage, the AEGAN-AD model and AE model have different settings, respectively. We display the settings in Table 1.

Table 1: The training settings for AEGAN-AD and Auto-Encoder models.

	AEGAN-AD	Auto-Encoder
epoch	60	100
batch size	512	256
learning rate	0.0002	0.001
feature	spectrogram	log-mel energies

3.3. Submissions

For the different characteristics, we adopt the ensemble strategy to improve the overall performance. In our experiments, we divide the audios from different categories into different groups, which can be seen in Table 2. The details of the submitted systems for evaluation dataset are shown in Table 3.

Table 2: The divided groups of all audios. The audios from development dataset and evaluation dataset are grouped, respectively.

Development dataset	Group1	bearing, fan, gearbox, slider
	Group2	ToyCar, ToyTrain, valve
Evaluation dataset	Group3	bandsaw, grinder, shaker
	Group4	ToyNscale
	Group5	ToyTank, Vacuum

3.4. Anomaly score

In our experiment, we find that the results calculated by different anomaly scores are various for different audios. Because only normal samples are available in the training dataset and the type in the training dataset are completely distinct with the evaluation dataset, the mean square error (MSE) and the Mahalanobis distance (MAHALA) were used to calculate the anomaly scores to detect abnormal samples for auto-encoder models in DCASE 2022 Challenge Task 2. In order to get the best performance of our system, we apply the two metrics to different groups.

Table 3: The ensemble systems details. Model1: AEGAN-AD (metric: G_z_cos_max). Model2: AEGAN-AD (metric: G_x_1_min). Model3: Baseline model with LeakyReLU (metric: the Mahalanobis distance). Model4: A layer is added to the baseline model and all of the connection dimensions are replaced by 128 (metric: the mean square error).

	Model1	Model2	Model3	Model4
system1	Group3 Group5	Group4		
system2	Group3	Group4 Group5		
system3			Group3 Group5	Group4
system4			Group3	Group4 Group5

4. RESULTS

According to the requirements in DCASE 2023 Challenge Task 2, we only present the results of the development dataset. Table 4 shows the AUC of source domain and target domain of all seven machines. Simultaneously, we display the pAUC ($p = 0.1$) for all machines across all domains. As a comparison, we provide the baseline results experimented in our computers.

From the table 4, it can be seen that AEGAN-AD obtains better overall metrics than the other three systems on most machine classification tasks, but lacks the ability to classify valve effectively. The second system is AE, which performs best to classify the value particularly. In summary, both AEGAN-AD and AE are more robust systems than baseline.

5. CONCLUSION

In this work, we have presented four systems to solve ASD problem for DCASE 2023 Challenge Task 2. Then, the ensemble systems are used to achieve the best results according to experience. We compared the performances of our systems with the official benchmark model. The results showed that our systems outperformed the baseline. Combining all the results, AEGAN-AD is the most steady system for this challenge.

6. REFERENCES

[1] D. Wang, C. Shen, and W. T. Peter, "A novel adaptive wavelet stripping algorithm for extracting the transients caused by bearing localized faults," *Journal of Sound and Vibration*, vol. 332, no. 25, pp. 6871–6890, 2013.

[2] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, "Mimii due: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions," in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz, NY, USA: IEEE, 2021, pp. 21–25.

[3] K. Qian, Z. Zhang, Y. Yamamoto, and B. W. Schuller, "Artificial intelligence internet of things for the elderly: from assisted living to health-care monitoring," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 78–88, 2021.

[4] K. Qian, Z. Zhang, A. Baird, and B. Schuller, "Active learning for bird sound classification via a kernel-based extreme learning machine," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1796–1804, 2017.

[5] Y. E. Karabacak, N. G. Özmen, and L. Gümüşel, "Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features," *Applied Acoustics*, vol. 186, p. 108463, 2022.

[6] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Description and discussion on dcase 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," *arXiv preprint arXiv:2305.07828*, 2023.

[7] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline," *arXiv preprint arXiv:2303.00455*, 2023.

[8] A. Jiang, W.-Q. Zhang, Y. Deng, P. Fan, and J. Liu, "Unsupervised anomaly detection and localization of machine audio: A gan-based approach," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE, 2023, pp. 1–5.

[9] 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," in *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*. Nancy, France: DCASE, 2022.

[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 Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*. Barcelona, Spain: DCASE, 2021, pp. 1–5.

Table 4: Anomaly detection results of different models on the evaluation dataset. AUC (source): The AUC of the source domain (%). AUC (target): The AUC of the target domain (%). pAUC (src & tgt): The pAUC (partial AUC) is calculated as the AUC over a low false-positive-rate range [0, 0.1] of source domain and target domain (%).

system	metrics	bearing	fan	gearbox	slider	ToyCar	ToyTrain	valve
baseline (MSE)	AUC (source)	65.92	80.19	60.31	70.31	70.10	57.93	55.35
	AUC (target)	55.75	36.18	60.69	48.77	46.89	57.02	50.69
	pAUC (src & tgt)	50.42	59.04	53.22	56.37	**52.47**	48.57	51.18
baseline (MAHALA)	AUC (source)	65.16	**87.10**	71.88	84.02	74.53	55.98	56.31
	AUC (target)	55.28	45.98	70.78	73.29	43.42	42.45	51.40
	pAUC (src & tgt)	51.37	59.33	54.34	54.72	49.18	48.13	51.08
AEGAN-AD	AUC (source)	**75.48**	81.32	**73.80**	**89.10**	**78.22**	**70.66**	43.18
	AUC (target)	**67.70**	**62.56**	69.74	67.38	**54.44**	59.04	43.04
	pAUC (src & tgt)	**58.00**	59.42	**59.84**	**64.11**	49.68	**51.26**	49.05
AE	AUC (source)	63.24	84.60	72.86	82.04	70.64	61.48	**67.48**
	AUC (target)	56.08	59.32	**74.06**	**75.74**	48.74	**59.62**	**62.62**
	pAUC (src & tgt)	51.21	**64.32**	58.89	59.42	51.16	48.84	**55.00**