ANOMALOUS SOUND DETECTION SYSTEM BASED ON SIMILAR-PAIRS CONTRASTIVE LEARNING

Technical Report

Kong Dewei[1,2]*, Yu Hongjiang*[1,2]*, Wang Shuai*[1,2]*, Zhang Bo*[1,2]*,*

[1] Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, China
david.kong.fun@outlook.com yuhj0704@163.com
{2111892612,2030177804}@qq.com

ABSTRACT

This technical report presents our approach for Task 2 of the DCASE 2024 challenge, which focuses on unsupervised anomalous sound detection for machine condition monitoring. We constructed four subsystems based on similar-pairs contrastive learning, where the fisrt two are based on SMOTE, the third and the fourth subsystem are based on two augmentation method to get more generalizations. The difference between the first and second subsystem is which method is used to calculate the anomaly score, MSE or MAHALA. The same is true for the difference between the third and fourth systems.

***Index Terms*—** Unsupervised learning, SMOTE, Augmentation, MSE, MAHALA

1. INTRODUCTION

The DCASE 2024 Challenge Task 2 [1] aims at "First-shot unsupervised anomalous sound detection for machine condition monitoring". Compared with DCASE 2023 Challenge Task 2, the task want to further deepen the techniques that are useful for this problem setting grounded on real-world scenarios, and the evaluation dataset is updated with new machine types unseen in the previous DCASE ASD challenges, and that attribute information such as the machine operation conditions are concealed for several machine types.

For task 2, [2] provides two baseline methods. The first method utilizes a standard autoencoder, which performs well in unsupervised anomaly detection but faces challenges in domain generalization. The second method is based on a Mahalanobis distance autoencoder, which performs well on the source domain but has subpar performance on the target domain.

Classification-based self-supervised learning [3, 4] has been shown to work well in challenges over the past few years. However, similar-pairs contrastive learning has not yet been applied to machine condition monitoring. Different from constrastive learning, similar-pairs contrastive learning just use positive samples to get a latent feature. This is the first try. We use Autoencoder as our backbone network. To get better latent features, we introduced the projector after the encoder. In order to obtain better generalization, we adopt two data augmentation techniques. We also use MSE and MAHALA to calculate the anomaly score.

2. METHODS

2.1. SMOTE

Synthetic Minority Over-sampling Technique(SMOTE)[5] is an approach in machine learning used to address the problem of imbalanced datasets. In task 2, the data in the source domain far exceeds the data in the target domain. Therefore, we use SMOTE to increase the amount of data in the target domain.

2.2. Data Augmentation

To get better generalization, we use two data Augmentation methods, frequency masking and time masking[6].

Frequency masking is applied so that f consecutive mel frequency channels $[f_0, f_0 + f)$ are masked, where f is first chosen from a uniform distribution from 0 to the frequency mask parameter F , and f_0 is chosen from $[0, v - f)$. v is the number of mel frequency channels.

Time masking is applied so that t consecutive time steps $[t_0, t_0 + t)$are masked, where t is first chosen from a uniform distribution from 0 to the time mask parameter T, and t_0 is chosen from $[0, \tau - t)$.

2.3. NetWork

The framwork is inspired by Barlow Twins[7]. And we use two AEs as backbone network. We train separate models for each machine type. In order to get more generalization characteristics, we added the projector which is actually a fully connected layer after the encoder. Finally, we combine the reconstruction error and similarity error.

2.4. Datasets

The dataset used for this task is derived from the MIMII DG[8] and ToyADMOS2[9] dataset, consisting of normal and anomalous operation sounds from 14 types of toys/real machines. Each recording, which is generated by mixing machine sounds recorded at laboratories with environmental noise recorded at factories and in the suburbs, is a single-channel audio with a duration of 6 to 10 s and a sampling rate of 16 kHz. Each machine type has only one section included in both the development dataset and the additional dataset. In this report, we just use the development dataset, which consists

of seven machine types (fan, gearbox, bearing, slide rail, valve, ToyCar, ToyTrain). The performance of the model is evaluated on the testing data from the development dataset.

3. RESULTS AND DISCUSSIONS

The results are showed in Tab.1 and Tab.2. We find that using smote will improve the accuracy of the target, but using augmentation will not, but will worsen the results. The next work is to find a better network to extract latent features.

Table 1: **Anomaly detection results(%) for different machine types based on MSE**

	Methods	Baseline	Our SMOTE	Ours Aug
ToyCar	AUC(source)	66.98	66.24	66.12
	AUC(target)	33.75	37.32	34.23
	pAUC	48.77	48.42	48.89
ToyTrain	AUC(source)	76.63	76.88	76.08
	AUC(target)	46.92	48.22	46.24
	pAUC	47.95	48.31	47.89
bearing	AUC(source)	62.01	62.97	61.26
	AUC(target)	61.4	65	60.42
	pAUC	57.58	57.35	56.11
fan	AUC(source)	67.71	68.02	69.98
	AUC(target)	55.24	53.7	53.06
	pAUC	57.53	56	58.78
gearbox	AUC(source)	70.4	68.8	67.92
	AUC(target)	69.34	69.48	68.02
	pAUC	55.65	55.32	54.78
slider	AUC(source)	66.51	67.42	62.28
	AUC(target)	56.01	59.96	55.48
	pAUC	51.77	53.32	51.05
valve	AUC(source)	51.07	48.68	49.86
	AUC(target)	46.25	45.8	43.63
	pAUC	52.42	52.05	51.63

Table 2: **Anomaly detection results(%) for different machine types based on MAHALA**

	Methods	Baseline	Our SMOTE	Ours Aug
ToyCar	AUC(source)	63.01	59.28	58.58
	AUC(target)	37.35	40.44	42.28
	pAUC	51.04	51.21	49.84
ToyTrain	AUC(source)	61.99	61.24	62.4
	AUC(target)	39.99	40.06	41.5
	pAUC	48.21	47.89	48.26
bearing	AUC(source)	54.43	54.1	53.06
	AUC(target)	51.58	54.4	52.1
	pAUC	58.82	58.95	59.63
fan	AUC(source)	79.37	77.44	79.12
	AUC(target)	42.70	43.8	41.06
	pAUC	53.44	52.05	55.32
gearbox	AUC(source)	81.82	80.46	78.08
	AUC(target)	74.35	74.66	73.42
	pAUC	55.74	54.79	54.68
slider	AUC(source)	75.35	80.09	63.86
	AUC(target)	68.11	75.56	59.88
	pAUC	49.05	48.84	48.95
valve	AUC(source)	55.69	52.22	57
	AUC(target)	53.61	52.7	48.76
	pAUC	51.26	50.74	49.05

4. REFERENCES

[1] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2024 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," *In arXiv e-prints: 2406.07250*, 2024.

[2] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline," in *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 191–195.

[3] Y. Zhang, J. Liu, Y. Tian, H. Liu, and M. Li, "A dual-path framework with frequency-and-time excited network for anomalous sound detection," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1266–1270.

[4] K. Wilkinghoff and F. Kurth, "Why do angular margin losses work well for semi-supervised anomalous sound detection?" *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 608–622, 2024.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, June 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953

[6] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "Specaugment: A simple data augmentation method for automatic speech recognition," *Interspeech 2019*, Sep 2019. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-2680

[7] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," 2021.

[8] 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, November 2022.

[9] 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, November 2021, pp. 1–5.