Abnormal Sound Detection Based on Domain Generalization

Technical Report

Mengyuan Wan[1], Yong Sun[2], Junjie Wang[1], Jiajun Wang[2], Shengbing Chen[2], Mengyuan Liu[1]

[1]Industrial Equipment States Evaluation and Fault Prediction Technology R&D Lab
[2]Multimodal Information Processing and Modeling for Industrial Equipment R&D Lab
School of Artificial Intelligence and Big Data
Hefei University, Hefei, China
18861195122@163.com, 2032336570@qq.com, 15656793081@163.com
2350846451@qq.com, shbchen@hfuu.edu.cn, 2863845001@qq.com

ABSTRACT

This technical report describes the abnormal sound detection system we submitted for DCASE 2024 Task 2. Compared to previous challenges, this task not only focuses on the first shot issue in 2023, but also requires the system to run well when attribute information is available and unavailable. We submit four methods for machine state anomaly sound detection. The first and second methods are based on self-supervised learning, using the feature vector extracted from the convolutional neural network, using the outlier detection algorithm to identify abnormal sounds. The third method uses the mixup method for probability mixing, and the fourth method uses the combination of SMOTE and mixup. Experiments on the development set show that the performance of the four methods is better than that of the baseline model.

***Index Terms*—**DCASE, anomalous sound detection, domain generalization

1. INTRODUCTION

This report describes the Anomalous Sound Detection (ASD) method developed for DCASE 2024 Challenge Task 2. For Task 2, the task organizer [1] provided a baseline system based on an autoencoder (AE) with two different operating modes. Simple autoencoder mode and selective Mahalanobis mode. This task is mainly set for the first-shot problem under the domain generalization requirement. First, the primary goal of the first problem is to rapidly deploy an ASD system for a new type of machine without the need for machine-specific hyperparameter tuning. Second, there may be a limited number of machines of the machine type, in which case the system should be able to use several machines of the machine type to train the model. Finally, the system must function well when new machine type attribute information is available and not.

In anomaly detection, the proportion of source domain and target domain data in the data set is seriously unbalanced. When the model trained on such data is applied to the target domain data, the model detection performance will be degraded due to the domain deviation caused by factors other than the anomaly. In order to solve the problem of imbalance between source domain and target domain data, we introduced SMOTE method and Mixup method on the basis of the baseline model to balance and mix the data of source domain and target domain. The experimental results show that these methods are effective in developing sets.

2. EXPERIMENTAL SETUP

2.1. Dataset

Our system uses a dataset consisting of MIMII DG［3］and ToyADMOS2[2]. The development dataset includes normal and abnormal sounds from 7 real machines, including fans, transmissions, bearings, sliders, ToyCar, ToyTrain, and Valve. It is worth noting that for the four types of machines (fans, bearings, valves, toy cars), the file name and attributes csv also provide attributes representing the operation or environmental conditions. For the other three types of machines, the attributes are hidden. In the evaluation dataset, the collection of machine types is completely different from the development dataset. There are nine novel types of machines (3D printers, air compressors, brushless motors, hair dryers, hovering drones, robotic arms, scanners, toothbrushes, ToyCircuit). For five types of machines (3D printers, hair dryers, robotic arms, scanners, ToyCircuit), this dataset provides attributes. For the other four machine types, the attributes are hidden.

2.2. Features

In order to extract information from raw audio data for abnormal sound detection, we extracted log Mel spectral features as inputs to the network.The sample rate used in the experiments is 16KHz, and we use 128 Mel-bins and nframes of 64. In addition, the 5 consecutive frames areconcatenated and 640 dimensions are input to model. In addition, in order to enhance the data, pitch enhancement, clipping transformation and random perturbation are also applied to this feature.

3. PROPOSED METHOD

3.1. Self-supervised learning

For task 2, the difficulty of the task increases when there is an acoustic difference (domain shift) between the training data set and the test data set. The first system (method1) and the second system (method2) we propose is an ASD model based on self-supervised learning. Firstly, we design an effective attention module, Multidimensional Attention module (MDAM)[5]. Given a shallow feature map of a sound, the module inferences attention along three independent dimensions: time, frequency, and channel. It focuses on specific frequency bands that contain semantically relevant discriminant information and time frames, thereby enhancing the representation learning ability of network models.

3.2. Unsupervised learning

We introduced SMOTE (Synthetic Minority Oversampling Technique) [6] and Mixup[7] methods on the basis of the autoencoder baseline system[4], as the third and fourth systems we submitted, and the fourth system is a combination of SMOTE and mixup methods.We compared the standard AE with the Mahalanobis distancebased AE (MHLAE) and observed that MHLAE performs better overall on the development set but exhibits poorer performance on the target domain. This could be attributed to the limited number of samples in the target domain. To address this issue, we employed the SMOTEoversampling technique to alleviate the sample imbalance between the source and target domains. Additionally, we utilized Mixup, which involves interpolating between samples from the target and source domains to generate new samples that simulate a new domain. This approach enhanced the model's ability to generalize to unseen data.

4. RESULTS

Table 1 shows the results of Methods 1, 2, 3, and 4 on the development dataset, including AUC and pac for source and target domains. As shown in the table, our results outperformed the baseline results on most machine types.

5. REFERENCES

[1] Nishida T, Harada N, Niizumi D, et al. Description and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring[J]. arxiv preprint arxiv:2406.07250, 2024.

[2] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, “ToyADMOS2: Another dataset of miniaturemachine operating sounds for anomalous sound detection under domain shift conditions,” in DCASE Workshop, 2021, pp.1–5

[3] 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 DCASE Workshop, 2022.

[4] N. Harada, N. Daisuke, T. Daiki, O. Yasunori, and Y. Masahiro, “First-shot anomaly detection for machine condition monitoring: a domain generalization baseline,” in EUSIPCO, 2023, pp. 191–195.

[5] Chen, S., Wang, J., Wang, J., Xu, Z. (2024). MDAM: Multi-Dimensional Attention Module for Anomalous Sound Detection. In: Luo, B., Cheng, L., Wu, ZG., Li, H., Li, C. (eds) Neural Information Processing. ICONIP 2023. Communications in Computer and Information Science, vol 1967. Springer, Singapore. https://doi.org/10.1007/978-981-99-8178-6_4

[6] Fernández A, Garcia S, Herrera F, et al. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary[J]. Journal of artificial intelligence research, 2018, 61: 863-905.

[7] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.

Table 1: Anomaly detection results for different machine types

Model	Score	ToyCar	ToyTrain	Bearing	Fan	Gearbox	Slider	Valve
Baseline (Mahala)	s_AUC	63.1	61.99	**79.37**	**81.82**	54.43	75.35	55.69
	t_AUC	37.35	39.99	42.7	74.35	51.58	68.11	53.61
	pAUC	**51.04**	48.21	53.44	55.74	58.82	49.05	51.26
Method1	s_AUC	52.0	77.88	61.6	63.23	**73.72**	95.76	**94.64**
	t_AUC	**50.2**	63.0	67.2	62.36	**77.64**	87.72	**59.0**
	pAUC	49.2	52.57	**57.57**	55.42	54.31	70.84	**63.21**
Method2	s_AUC	50.16	**78.16**	60.68	63.6	69.19	**95.96**	92.39
	t_AUC	47.44	**65.52**	**65.72**	77.84	68.24	**88.6**	54.12
	pAUC	49.36	**55.21**	55.15	58.78	54.47	**75.0**	61.47
Method3	s_AUC	**64.24**	66.12	78.58	78.9	59.96	80.86	58.28
	t_AUC	42.44	34.54	45.64	**78.92**	56.6	73.8	53.02
	pAUC	48.89	47.52	57.42	**61.10**	58.73	55.10	50.78
Method4	s_AUC	56.6	56.5	76.88	81.56	52.68	80.32	53.02
	t_AUC	44.0	42.92	45.84	73.62	58.78	73.92	49.84
	pAUC	50.47	47.78	53.0	54.36	**59.42**	48.47	52.94