Anomaly Sound Detection Method Based on Training Attribute Classification Models

Technical Report

Fan Chu[1], *Mengui Qian*[1]
[1] National Intelligent Voice Innovation Center, Hefei, China
fanchu@nivic.cn, qianmengui@my.swjtu.edu.cn

ABSTRACT

In this report, we present our solution to the DCASE 2025 Challenge Task 2, focusing on First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. In this year's challenge, some machines still lack attribute and domain labels, while clean machine sound or background noise are provided. Our approach involves using clustering or feature frequency analysis algorithms to assign artificial labels to samples without attribute labels, then training an attribute classification model together with other machines that have attributes. Additionally, we introduce a data augmentation strategy by mixing clean machine sound with background noise to generate simulated data. Finally, we employ the model's embedding to train a KNN model for obtaining anomaly scores. Our system achieves 63.25% in the harmonic mean of AUC and pAUC (p = 0.1) across all machine types and domains on the development set.

Index Terms**—** anomalous sound detection, clustering, feature frequency analysis, attribute classification

1. INTRODUCTION

Anomalous sound detection (ASD) is the task of identifying whether the sound emitted from a target machine is normal or anomalous. Automatic detection of mechanical failure is an essential technology in the fourth industrial revolution, which involves artificial-intelligence-based factory automation. Prompt detection of machine anomalies by observing sounds is useful for monitoring the condition of machines.

The task this year is to develop an ASD system that meets the following five requirements [1]:

Train a model using only normal sound (unsupervised learning scenario). Because anomalies rarely occur and are highly diverse in real-world factories, it can be difficult to collect exhaustive patterns of anomalous sounds. Therefore, the system must detect unknown types of anomalous sounds that are not provided in the training data.

Detect anomalies regardless of domain shifts (domain generalization task). In real-world cases, the operational states of a machine or the environmental noise can change to cause domain shifts. Domain-generalization techniques can be useful for handling domain shifts that occur frequently or are hard-to-notice. In this task, the system is required to use domain-generalization techniques for handling these domain shifts.

Train a model for a completely new machine type. For a completely new machine type, hyperparameters of the trained model cannot be tuned. Therefore, the system should have the ability to train models without additional hyperparameters tuning.

Train a model using a limited number of machines from its machine type. While sounds from multiple machines of the same machine type can be used to enhance the detection performance, it is often the case that only a limited number of machines are available for a machine type. In such a case, the system should be able to train models using a few machines from a machine type.

Train a model both with or without attribute information. While additional attribute information can help enhance the detection performance, we cannot always obtain such information. Therefore, the system must work well both when attribute information is available and when it is not.

Train a model with additional clean machine data or noise-only data (optional) Although the primary training data consists of machine sounds recorded under noisy conditions, in some situations it may be possible to collect clean machine data when the factory is idle or gather noise recordings when the machine itself is not running. Participants are free to incorporate these additional data sources to enhance the accuracy of their models.

In the following, we describe our approach and experimental results in detail. Each sound used in this challenge is a single channel audio, and different machines have different audio duration. The development set includes seven machines (ToyCar, ToyTrain, Fan, Gearbox, Bearing, Slide rail and Valve), of which three machines do not provide attribute information and only provide domain information, while the other four machines provide attribute and domain information. Furthermore, the additional training set and evaluation set include eight new machines (AutoTrash, HomeCamera, ToyPet, ToyRCCar, Polisher, ScrewFeeder, CoffeeGrinder and BandSealer), of which four machines do not provide attribute information and only provide domain information, while the other five machines provide attribute and domain information. [2, 3]

2. PROPOSED APPROACH

2.1. Automatic Attribute Annotation

For non-attributed machines, we primarily employ two methodologies for automatic attribute annotation: one based on unsupervised clustering of acoustic features, and the other leveraging characteristic frequencies. The specific approaches are as follows:

1. Unsupervised Clustering: Analysis has revealed significant disparities in the operational acoustic features of certain machines. Consequently, we perform unsupervised clustering on spectrograms or frequency spectra of such machines to derive attribute annotations.
2. Attribute Annotation via Characteristic Frequencies: As most machines are predominantly motor-driven, their operation exhibits distinctive characteristic frequencies that vary according

to machine attributes.Therefore, we conduct envelope spectrum analysis on audio signals to extract characteristic frequencies within the target sound frequency bands. These frequencies subsequently inform machine attribute annotation.

2.2. Feature Extraction

For shorter audio below 10 seconds, we repeatedly concatenate to ensure that all audio has a duration of 10 seconds. Then we transformed all audio clip into LogMel feature. We found that the characteristics of certain machines are mainly concentrated in high frequencies through observing the spectrum. Based on this discovery, we first pass through a high pass filter before passing through the Mel filter. Experiments have shown that for certain types of machines, filtered features are more suitable for anomaly detection task. The maximum frequency is set to 8000Hz, and the minimum frequency is 200Hz.

2.3. Classification-Based Model

We utilize a SeResNet18-based feature extractor [5], trained with the AdaCos loss function, which demonstrates greater robustness to label noise compared to traditional softmax-based losses. For classification, the total number of classes encompasses all unique combinations of machine IDs, domain labels, and attribute labels present in the dataset. In cases where attribute information is unavailable for some machines, we apply the approach outlined in Section 2.1 to generate artificial attribute labels.

Kevin Wilkinghoff introduced an effective self-supervised learning (SSL) method for audio scene disentanglement (ASD) [6], integrating mixup [7], StatEx [8], and FeatEx to augment new categories. This approach enhances model generalization and improves audio feature extraction. To further diversify the categories, we curated clean machine data or noise-only data from supplemental dataset and combined them with DCASE Task2 data for training.

2.4. KNN based Backend

A K-nearest neighbors (KNN) algorithm was employed as an inlier model [9] to fit the probability distribution of normal data, after which anomaly scores were computed. Ultimately, we evaluated the performance of the model on machines with original or artificial attribute labels in the evaluation set.

Table 2 presents the results of all our models. Compared to the baselines [4], our system shows better performance both on the source domain and the target domain.

Table 2: Anomaly detection results for different machine types

	Method	Baseline	Our system
ToyCar	AUC(source)	71.05%	**94.3%**
	AUC(target)	53.52%	**77.02%**
	pAUC	49.7%	**57.32%**
ToyTrain	AUC(source)	**61.76%**	59.18%
	AUC(target)	56.46%	**63.24%**
	pAUC	**50.19%**	50.05%
bearing	AUC(source)	**66.53%**	57.08%
	AUC(target)	53.15%	**62.68%**
	pAUC	**61.12%**	52.47%
fan	AUC(source)	**70.96%**	54.78%
	AUC(target)	38.75%	**53.70%**
	pAUC	49.46%	**50.74%**
gearbox	AUC(source)	64.8%	**79.92%**
	AUC(target)	50.49%	**78.44%**
	pAUC	52.49%	**64.68%**
slider	AUC(source)	70.1%	**74.24%**
	AUC(target)	48.77%	**61.68%**
	pAUC	52.32%	**52.89%**
valve	AUC(source)	63.53%	**79.28%**
	AUC(target)	67.18%	**84.04%**
	pAUC	57.35%	**65.11%**
All (hmean)	AUC(source)	66.78%	**68.71%**
	AUC(target)	51.39%	**67.16%**
	pAUC	52.94%	**55.59%**

3. CONCLUSIONS

We propose an approach for DCASE2025 task2, using artificial labels to samples without attribute labels, then training an attribute classification model together with other machines that have attributes and training a KNN model for obtaining anomaly scores. Our system achieves the score of 63.25% on the development set, and shows better performance than the baselines.

4. REFERENCES

[1] Tomoya Nishida, Noboru Harada, Daisuke Niizumi, Davide Albertini, Roberto Sannino, Simone Pradolini, Filippo Augusti, Keisuke Imoto, Kota Dohi, Harsh Purohit, Takashi Endo, and Yohei Kawaguchi. Description and discussion on DCASE 2025 challenge task 2: first-shot unsupervised anomalous sound detection for machine condition monitoring. In arXiv e-prints: 2506.10097, 2025.

[2] Harada, Noboru and Niizumi, Daisuke and Takeuchi, Daiki and Ohishi, Yasunori and Yasuda, Masahiro and Saito, Shoichiro. 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), 1–5. Barcelona, Spain, November 2021.

[3] Dohi, Kota and Nishida, Tomoya and Purohit, Harsh and Tanabe, Ryo and Endo, Takashi and Yamamoto, Masaaki and Nikaido, Yuki and Kawaguchi, Yohei. 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.

[4] Harada, Noboru and Niizumi, Daisuke and Ohishi, Yasunori and Takeuchi, Daiki and Yasuda, Masahiro. First-Shot Anomaly Sound Detection for Machine Condition

Monitoring: A Domain Generalization Baseline. In 2023 31st European Signal Processing Conference (EUSIPCO), 191-195. 2023.

[5] Hu, Jingcao, Liangzhi Shen, and Saad Albanie. "SeResNet: Deep Spatial Feature Reconstruction for Image Recognition." IEEE Transactions on Multimedia, vol. 24, no. 3, 2021, pp. 985-996.

[6] Wilkinghoff, Kevin. "Self-supervised learning for anomalous sound detection." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

[7] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz, " Mixup: Beyond empirical risk minimization," in ICLR, 2018.

[8] Han Chen, Yan Song, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, and Ian McLoughlin, " An effective anomalous sound detection method based on representation learning with simulated anomalies," in ICASSP. IEEE, 2023.

[9] Fujimura, T., Kuroyanagi, I., Hayashi, T., & Toda, T. (2023). Anomalous sound detection by end-to-end training of outlier exposure and normalizing flow with domain generalization techniques. DCASE2023 Challenge, Tech. Rep.