Unified Anomaly Detection for Machine Condition Monitoring: Handling Attribute-Rich and Attribute-Free Scenarios

Technical Report

Fan Chu[1], *Yuxuan Zhou*[1], *Mengui Qian*[1]
[1] National Intelligent Voice Innovation Center, Hefei, China
{fanchu, yxzhou15}@nivic.cn, qianmengui@my.swjtu.edu.cn

ABSTRACT

In this report, we present our solution to the DCASE 2024 Challenge Task 2, focusing on First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. The challenge this year involves handling machine types with varying levels of attribute and domain information. Our approach addresses this by categorizing the machines into two groups: those with attribute labels and those without. We respectively train two models that perform attribute classification for these two groups, both based on a pretrained model that performs domain classification. Our system achieves 62.001% 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, attribute classification, pretraining, machine condition monitoring

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 generalizationtask). 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 hyperparameter 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.

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 durations. 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 nine new machines (3DPrinter, AirCompressor, BrushlessMotor, HairDryer, HoveringDrone, RoboticArm, Scanner, ToothBrush, and ToyCircuit), 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. 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.

2.2. Pretraining

We use training data from all 16 machines to train a domain classification model as our pretrained model, with the target of

distinguishing 32 domain categories from the 16 machines. In addition, we used mixup [5] for data augmentation. Our backbone network is ResNet-34 [6], and pooling layer is Attentive Statistic Pooling [7], which calculates the first and second order statistics of each frame's audio features, obtains each frame's embedding through concatenation statistics, and introduces a self-attention mechanism to calculate the weight of each frame using each frame's audio features (output of the backbone network), and obtains the clip-level embedding through weighted averaging. The BCE loss function [8] is employed. Table 1 presents the architecture of our pretrained model.

Table 1: Architecture of pretrained model

Layer name	Operator
Conv1	7*7, 64, stride 2
Pooling1	3*3 Max Pooling, stride 2
Conv2_x	[3*3, 64] * 6
Conv3_x	[3*3, 128] *8
Conv4_x	[3*3, 256] *12
Conv5_x	[3*3, 512] *6
Pooling2	Attentive Statistic Pooling
FC1	1000-d
FC2	32-d

2.3. Classification-Based Model

For machines with available attribute and domain labels, we employ the model finetuned to classify all attributes associated with the five different machines, based on the pretrained model. We used the training data of nine machines that provided attribute information for finetuning, and the finetuned classification-based model is allowing for precise feature extraction from normal audio data. To avoid the impact of domain shifting, we used a domain balancing strategy to ensure that there is at least one target domain data in each batch. GMM [9] are used as inlier model [10] to fit the probability distribution of normal data, and then calculate anomaly scores. Finally, we obtain the results of five machines with attributes in the evaluation set.

For machines lacking attribute labels, we utilize another finetuned models to develop a generalized classification model, by training on the combined data from the four attribute-rich machines of the development set and all nine machines of evaluation set. We mix attribute labels and domain labels together for classification. Similarly, GMM are used as inlier model to fit the probability distribution of normal data and calculate anomaly scores. Finally, we obtain the results of the remaining five machines without attributes in the evaluation set.

This dual approach provides a flexible and effective solution for detecting anomalies in diverse machine condition monitoring scenarios. This enables a consistent and effective strategy for anomaly detection across both attribute-rich and attribute-free scenarios, ensuring robust performance whether attribute information is available or not.

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)	**66.98%**	62.02%
	AUC(target)	33.75%	**56.40%**
	pAUC	48.77%	**49.42%**
ToyTrain	AUC(source)	76.63%	**80.66%**
	AUC(target)	46.92%	**57.18%**
	pAUC	47.95%	**54.84%**
bearing	AUC(source)	62.01%	**69.10%**
	AUC(target)	61.40%	**70.46%**
	pAUC	57.58%	**58.47%**
fan	AUC(source)	67.71%	**76.92%**
	AUC(target)	**55.24%**	40.68%
	pAUC	**57.53%**	56.37%
gearbox	AUC(source)	**70.40%**	68.48%
	AUC(target)	**69.34%**	67.82%
	pAUC	**55.65%**	52.84%
slider	AUC(source)	66.51%	**78.20%**
	AUC(target)	56.01%	**64.24%**
	pAUC	51.77%	**54.89%**
valve	AUC(source)	51.07%	**86.66%**
	AUC(target)	46.25%	**71.22%**
	pAUC	52.42%	**66.74%**
All (hmean)	AUC(source)	65.00%	**73.74%**
	AUC(target)	50.28%	**59.15%**
	pAUC	52.84%	**55.80%**

3. CONCLUSIONS

We propose an approach for DCASE2024 task2, categorizing the machines into two groups: those with attribute labels and those without. We respectively train two models that perform attribute and domain classification for these two groups, both based on a pretrained model that performs domain classification. Our system achieves the score of 62.001% on the development set, and shows better performance than the baselines.

4. REFERENCES

[1] Nishida, Tomoya and Harada, Noboru and Niizumi, Daisuke and Albertini, Davide and Sannino, Roberto and Pradolini, Simone and Augusti, Filippo and Imoto, Keisuke and Dohi, Kota and Purohit, Harsh and Endo, Takashi and Kawaguchi, Yohei. 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] 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] Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arxiv preprint arxiv:1710.09412.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[7] Okabe, K., Koshinaka, T., & Shinoda, K. (2018). Attentive statistics pooling for deep speaker embedding. arxiv preprint arxiv:1803.10963.

[8] Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. Int. J. Adv. Trends Comput. Sci. Eng, 9(10).

[9] Bond, S. R., Hoeffler, A., & Temple, J. R. (2001). GMM estimation of empirical growth models. Available at SSRN 290522.

[10] 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.