HYBRID ANOMALY DETECTION APPROACH FOR DCASE 2024 TASK 2

Technical Report

Shuxian Wang[1], *Guirui Zhong*[1], *Qing Wang*[1], *Jun Du*[1]

[1] University of Science and Technology of China, Hefei, China
sxwang21@mail.ustc.edu.cn, grzhong2002@163.com,{qingwang2, jundu}@ustc.edu.cn

ABSTRACT

Addressing the unique challenge of the DCASE 2024 Task 2, where the availability of attribute information varies, we propose a hybrid anomaly detection approach that combines generative and discriminative techniques. Leveraging both autoencoder (AE) for unsupervised learning and attribute classification for supervised learning, our system is designed to perform effectively under diverse conditions. The AE is trained to reconstruct normal sound data and detect anomalies, providing robustness in scenarios where attribute information is unavailable. Simultaneously, the attribute classification component enhances detection performance when attribute information is present. By seamlessly integrating these approaches, our system achieves a balanced performance across different conditions, ensuring reliable anomaly detection in machine condition monitoring applications.

***Index Terms*—** anomaly detection, autoencoder, attribute classification, hybrid approach, machine condition monitoring

1. INTRODUCTION

In DCASE challenge 2024 Task 2 "*First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring*" [1], it is required to detect anomalous sounds of machines. In real-world conditions, it is often easier for us to obtain the sound of the machine working normally, while the anomalies are rare and highly diverse. Therefore, we need to use the normal sounds in the training data to detect anomalous sounds in the test data. Furthermore, the operational states of a machine or the environmental noise can change to cause domain shifts. The system needs to use domain generalization techniques to handle frequent or hard-to-notice domain shifts. In the DCASE 2024 task, consistent with last year, first-shot problem is still introduced, whose two main features are training a model for a completely new machine type, and using a limited number of machines from its machine type when train a model. Moreover, unlike last year [2], one new requirement is introduced in the DCASE 2024 task, that is, we need to train a model both with or without attribute information because we cannot always obtain such information in real-world applications.

Our submission includes two major approaches for anomalous sound detection. To begin with, for those machines with attribute information, the detection method is based on machine attribute classification. Secondly, when the machine attribute information is unavailable, the corresponding approach is using the autoencoder to detect anomalies, and the Mahalanobis distance is used to calculate the anomaly score.

Each recording used in this challenge is a single-channel. For the case that the recording duration of part different machine types is inconsistent, we process the duration of all audio to 10 seconds by filling. The development set includes seven machines: ToyCar, ToyTrain, Fan, Gearbox, Bearing, Slide rail and Valve, and the evaluation set includes nine new machines: 3DPrinter, AirCompressor, BrushlessMotor, HairDryer, HoveringDrone, RoboticArm, Scanner, ToothBrush and ToyCircuit [3, 4]. In the following, we will describe each approach and our experimental results in detail.

2. PROPOSED APPROACH

2.1. Attribute Classification

Since the attribute information of some machine types is available, we can train a classifier with machine attribute information. Such anomalous sound detection methods based on self-supervised classification have been used in previous challenges [2, 5, 6, 7] and achieved good results [8, 9, 10, 11]. In our submission system, specifically, in order to get a more robust anomaly detector, first, we use the training data of all machines in the development set to train a 8-category domain classifier, and then we fine-tune the model parameters to get the attribute classifier for each machine. Each attribute of the machine has a classification head, and there is also a classification head for distinguishing positive machines from negative machines (the other three types of machines). Then, based on the attribute classifier, we extract embeddings of training data to train inlier models (IM) to model the probability distribution of normal data. In the inference stage, after each test data embedding is extracted, data that deviates from the probability distribution of normal data is detected as abnormal.

Firstly, we transformed all audio clip into spectrograms with a Mel transformation. In the choices of classifier architectures, we choose EfficientNet-B0 [12] as the network structure for domain classification and attribute classification, and mixup [13] is used for data augmentation. Further, a domain generalization strategy is applied, that is, when creating a mini-batch, we sample normal data in the target domain to ensure that there are two target domain samples in the mini-batch. AdamW [14] optimizer is used with the OneCycleLR scheduler for 300 epochs, and the initial learning rate is 0.001. The batch size is set to 128.

2.2. Conditional Autoencoder

The autoencoder (AE) is based on the reconstruction error to realize the detection of anomalous sound. That is, the input feature vector is first mapped to a hidden representation with a lower dimensional space by the encoder component, and then, the decoder component attempts to reconstruct the inverse transformation from the hidden representation to the original input signal. The differ-

Table 1: DCASE 2024 Task 2 experimental results on development dataset (%). The value in the row "Total Score" represents the harmonic mean of the AUC and pAUC scores over all the machine types, sections, and domains.

		Baseline (AE-MSE)	Baseline (AE-MAHALA)	Our system
ToyCar	AUC (source)	**66.98**	63.01	63.35
	AUC (target)	33.75	37.35	**56.93**
	pAUC	48.77	**51.04**	49.91
ToyTrain	AUC (source)	76.63	61.99	**82.76**
	AUC (target)	46.92	39.99	**58.12**
	pAUC	47.95	48.21	**55.14**
bearing	AUC (source)	62.01	54.43	**70.35**
	AUC (target)	61.40	51.58	**71.20**
	pAUC	57.58	58.82	**59.75**
fan	AUC (source)	67.71	**79.37**	77.35
	AUC (target)	**55.24**	42.70	41.98
	pAUC	**57.53**	53.44	57.30
gearbox	AUC (source)	70.40	**81.82**	67.35
	AUC (target)	69.34	**74.35**	66.83
	pAUC	55.65	**55.74**	52.13
slider	AUC (source)	66.51	75.35	**79.32**
	AUC (target)	56.01	**68.11**	65.34
	pAUC	51.77	49.05	**54.91**
valve	AUC (source)	51.07	55.69	**87.25**
	AUC (target)	46.25	53.61	**72.83**
	pAUC	52.42	51.26	**67.37**
Total Score		55.35	55.01	**62.65**

ence between the feature vector of the original input and the output vector of the autoencoder is the reconstruction error. In the training phase, we use the domain information of the machine as the condition, and the domain labels are encoded and input into AE for training along with the audio features. In the test phase, the test data uses the AE model of the corresponding machine, and we calculate the Mahalanobis distance according to different domains, and take the minimum value as the anomaly score. In addition, we perform score normalization by source and target domains. For the nine machines in the evaluation set, since the domain labels of the test set are unknown, we train a domain classifier, and then use the predicted labels to get anomaly scores.

The network architecture we use is similar to the baseline AE [15]. There are two differences between baseline and our system. First, we use the domain information of the machine as the condition for training. Besides, we use convolution layer instead of dense layer in the baseline. 128-dimensional log-Mel spectrogram features are used as input to the network. The batch size of training is set as 256 and Adam optimizer is used to train the model with the learning rate of 0.0005.

2.3. Results

In the development set, there are three machine types, including ToyTrain, Gearbox and Slide rail, having no attribute information. Thus, the detection method of these machine types is conditional autoencoder. For other four machine types, including ToyCar, Fan, Bearing and Valve, which have attribute information, the detection method is attribute classification. Table 1 shows our best results on the development set through the hybrid anomaly detection approach.

3. CONCLUSIONS

In this paper, we propose a hybrid method for anomalous sound detection based on attribute classification and conditional AE. Experimental results show that by seamlessly integrating generative and discriminative approaches, we can achieve a balanced performance and better results than the baseline under attribute-available and unavailable conditions.

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] 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," *In arXiv e-prints: 2305.07828*, 2023.

[3] 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.

[4] 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.

[5] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020)*, November 2020, pp. 81–85.

[6] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on dcase 2021 challenge task 2: Unsupervised anomalous detection for machine condition monitoring under domain shifted conditions," in *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, Barcelona, Spain, November 2021, pp. 186–190.

[7] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, and Y. Kawaguchi, "Description and discussion on dcase 2022

challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," in *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, November 2022, pp. 1–5.

[8] R. Giri, S. V. Tenneti, K. Helwani, F. Cheng, U. Isik, and A. Krishnaswamy, "Unsupervised anomalous sound detection using self-supervised classification and group masked autoencoder for density estimation," DCASE2020 Challenge, Tech. Rep., July 2020.

[9] J. Lopez, G. Stemmer, and P. Lopez-Meyer, "Ensemble of complementary anomaly detectors under domain shifted conditions," DCASE2021 Challenge, Tech. Rep., July 2021.

[10] Y. Zeng, H. Liu, L. Xu, Y. Zhou, and L. Gan, "Robust anomaly sound detection framework for machine condition monitoring," DCASE2022 Challenge, Tech. Rep., July 2022.

[11] J. Jie, "Anomalous sound detection based on self-supervised learning," DCASE2023 Challenge, Tech. Rep., June 2023.

[12] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 687–10 698.

[13] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[14] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[15] 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.