AN EFFICIENCE ANOMALOUS SOUND DETECTION SYSTEM FOR DCASE 2025 TASK 2

Technical Report

Wentao Zhou[1,2]*, Ying Hu*[1,2]*, Xin Fan*[1,2]*, Nannan Teng*[1,2]*, Tianqing Zhou*[1,2]*, Fangxu Chen*[1,2]*, Qingjing Wan*[1,2]*, Qiong Wu*[1,2]*, Qin Yang*[1,2]*,*

[1] XinJiang University, School of Information Science and Engineering, Urumqi, China
{zhouwt}@stu.xju.edu.cn
[2] Key Laboratory of Signal Detection and Processing in Xinjiang, Urumqi, China

ABSTRACT

This technical report describes the system we submitted to DCASE 2025 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring.This year's tasks are fundamentally aligned with those of last year. To reflect practical application scenarios, machine attributes are not always fully known. Building upon this, additional clean machine data or noise-only data have been incorporated into the training set. Our system employs the pre-trained model BEATs, utilizing the LoRA fine-tuning approach for the anomalous sound detection task. Arcface loss is incorporated to constrain machines with unknown attributes. Our best system achieved a harmonic mean of 77.13% in the harmonic mean of AUC in the source domain, 56.07% in AUC in the target domain, and 57.72% in pAUC(p=0.1) on the development set.

***Index Terms*—** First-shot, anomalous sound detection, pre-trained model

1. INTRODUCTION

Anomalous Sound Detection (ASD) task aims to identify the occurrence of abnormalities through machine sound analysis. Due to the scarcity of abnormal sounds, this is typically framed as an unsupervised task [1–4].Furthermore, due to variations in environmental noise during sound recording, physical parameters of equipment, and differences in recording devices, potential domain shifts may occur in sound detection systems. This can result in degraded abnormal sound recognition performance, which constitutes another significant challenge that needs to be addressed [3, 4]. Since machine attributes may not always be known, yet systems must remain operational, building upon previous challenges, DCASE 2024 Challenge Task 2 [5] intentionally configures certain machine attributes as unknown to better simulate real-world scenarios. In DCASE 2025 Challenge Task 2 [6], supplementary training data is provided, including normal machine sounds recorded during factory idle states or pure noise-only recordings when machines are inactive to enhance anomalous sound detection accuracy.

In recent years, classification-based approaches have demonstrated promising performance in anomaly sound detection. This approach typically utilizes machine meta-information, such as attributes and section IDs, with the classification task being utilized as an auxiliary task for anomaly detection. When machine attributes are known, models such as ST-gram [7], SW-WAVENET [8], and Kevin's CNN [9, 10] demonstrate strong performance. However, when partial machine attributes are missing, the fine-grained knowledge beneficial for anomaly detection, learned from audio of machines with known attributes, fails to generalize effectively to machines with missing attributes. The powerful generalization capabilities of pre-trained models help mitigate this limitation. Anbai Jiang et al. [11] first proposed applying pre-trained models BEATs [12] to ASD tasks, achieving top performance on the DCASE23 Task 2 evaluation set by fine-tuning the model [13] using Low-Rank Adaptaion(LoRA) [14]. Subsequently, on the DCASE24 dataset, Subcenter ArcFace [20] was developed, enabling coarse-grained machines (machines with missing attribute labels) to adaptively cluster within the feature space. Here, we

2. SYSTEM DESCRIPTION

2.1. Experimental setup

We conduct experiments on the dataset of DCASE 2025 Challenge Task 2, which comprises a development dataset and an additional dataset [15, 16]. Note that, the attribute information for 3 machine types in the development dataset and 4 machine types in the additional dataset are not provided. Additionally, the machine types in the development dataset are completely different from those in the additional dataset.

The base feature extraction model of the submitted system adopts BEATs. The input features are log-mel spectrograms with a frame length of 25ms and a frame shift of 10ms. The number of mel bins is set to 128. We use the BEATs-iter3 version, which is pre-trained on the full training set of the AudioSet dataset and utilizes 90M parameters. All audio clips are uniformly cropped or padded to 10 seconds. The model is trained for 30 epochs by AdamW [17] with a maximum learning rate of 0.0001 and a batch size of 32.

2.2. Train

SpecAugment [18] is employed to input log-mel spectrograms with a maximum mask length of 80 for time axis and 24 for frequency axis. We incorporated LoRA [13] fine-tuning specifically on the query and value projection matrices within the attention layers of the first four transformer blocks in BEATs, utilizing the fourth layer's output as the model's feature representation, and the rank of LoRA is set to 64. We employed attentive statistics pooling [19] layer followed by linear layers to project the output features from BEATs to a 128-dimensional space. We employ ArcFace [20] for machine attribute classification. Machines without attribute labels are categorized into two classes: the source domain and the target domain.

Table 1: Result on DCASE 2025 task 2 development dataset

	ToyCar	ToyTrain	bearing	fan	gearbox	slider	valve	All(hmean)
AUC Source	83.74	80.74	76.82	72.70	80.16	74.30	72.90	77.13
AUC Target	69.08	60.80	46.46	41.86	60.92	50.26	82.04	56.07
pAUC	57.84	55.21	53.37	55.58	60.79	52.47	73.37	57.72
hmean	68.64	63.90	56.31	53.92	66.17	57.24	75.88	62.34

2.3. Test

We leverage the KNN detector from AnoPatch [11] to compute anomaly scores based on pairwise cosine distances. Furthermore, we augment each machine's feature memory bank by incorporating clean machine sound features from supplementary datasets, while explicitly excluding pure noise segments.

2.4. Submission

The system we submit is one that utilizes the same model to output anomaly scores at different epochs.

3. RESULTS

Performance evaluation metrics are computed based on the area under the Receiver Operating Characteristic(ROC) curve (AUC). These include the AUC Source, AUC Target, partial AUC (pAUC, with p=0.1), and the Harmonic Mean. This aligns with the official evaluation criteria.

Table 1 presents the results of our optimal system on the development set of DCASE 2025 Challenge Task 2 for the 7 machine types. The Source AUC, Target AUC, and pAUC (p=0.1) scores for all machine types are 77.13% , 56.07%, and 57.72%, respectively. The final Harmonic Mean score is 62.34%.

4. CONCLUSION

In this technical report, we described our submission systems for the DCASE 2025 Challenge Task 2. Our submitted system is based on the pre-trained model BEATs and fine-tuned with LoRA which achieves promising performance under the constraint of the ArcFace loss even when some machine attribute labels are missing.

5. REFERENCES

[1] 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 DCASE 2020 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] ——, "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.

[7] Y. Liu, J. Guan, Q. Zhu, and W. Wang, "Anomalous sound detection using spectral-temporal information fusion," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 816–820.

[8] H. Chen, L. Ran, X. Sun, and C. Cai, "Sw-wavenet: Learning representation from spectrogram and wavegram using wavenet for anomalous sound detection," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[9] K. Wilkinghoff, "Design choices for learning embeddings from auxiliary tasks for domain generalization in anomalous sound detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[10] K. Wilkinghoff, "Self-supervised learning for anomalous sound detection," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 276–280.

[11] A. Jiang, B. Han, Z. Lv, Y. Deng, W. Zhang, X. Chen, Y. Qian, J. Liu, and P. Fan, "Anopatch: Towards better consistency in machine anomalous sound detection," in *25th Annual Conference of the International Speech Communication Association,*

Interspeech 2024, Kos, Greece, September 1-5, 2024, I. Lapidot and S. Gannot, Eds. ISCA, 2024.

[12] S. Chen, Y. Wu, C. Wang, S. Liu, D. Tompkins, Z. Chen, W. Che, X. Yu, and F. Wei, "Beats: Audio pre-training with acoustic tokenizers," in *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, ser. Proceedings of Machine Learning Research, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., vol. 202. PMLR, 2023, pp. 5178–5193.

[13] X. Zheng, A. Jiang, B. Han, Y. Qian, P. Fan, J. Liu, and W. Zhang, "Improving anomalous sound detection via low-rank adaptation fine-tuning of pre-trained audio models," in *IEEE Spoken Language Technology Workshop, SLT 2024, Macao, December 2-5, 2024*. IEEE, 2024, pp. 969–974.

[14] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," in *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[15] 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, pp. 1–5.

[16] 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 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, Barcelona, Spain, November 2021, pp. 1–5.

[17] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] 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," in *Proc. Interspeech 2019*, 2019, pp. 2613–2617.

[19] N. Dawalatabad, M. Ravanelli, F. Grondin, J. Thienpondt, B. Desplanques, and H. Na, "Ecapa-tdnn embeddings for speaker diarization," in *Proc. Interspeech 2021*, 2021, pp. 3560–3564.

[20] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.