A TWO STAGE FUSION ANOMALY DETECTION APPROACH FOR TASK2

Technical Report

*Jie Yang**

HuaiBei, China
852549193@qq.com

ABSTRACT

This technical report details our approach to addressing Task 2 of the DCASE 2025 Challenge. We propose a two stage fusion adaptive anomaly detection scheme which combine adaptive filtering for denoising and separation and neural network-based classifier. First, traditional signal separation and denoising techniques are employed to preprocess the raw audio signal. This stage focuses on suppressing noise, isolating interfering sound sources, and enhancing the signal-to-noise ratio (SNR) of the target machine sound. For the attribute classification network, we leverage the depth-wise separable convolutions and bottleneck structure of MobileFaceNet to efficiently learn deep discriminative features of anomalous sounds. Finally,an anomaly score is computed based on k-means and cosine distances.

The results demonstrate that the proposed method achieves significant performance improvements and ensures robust adaptability under varying data conditions. Furthermore, the framework's flexibility in handling different types of input data enhances its applicability in real-world industrial machine monitoring scenarios.

Index Terms— Two-stage Fusion, Adaptive Anomaly Detection, Adaptive Filter

1. INTRODUCTION

With the advancements in society and technology, machinery has become increasingly crucial in industrial production. However, various factors often lead to equipment failure during operation, which can impact production efficiency and performance, and even result in severe safety incidents. Anomalous sound detection (ASD) is a vital component of machine condition monitoring tools that involves identifying whether the sound emitted from a target machine is normal or anomalous [1].

The task topic of this year's task2 is First-shot unsupervised anomalous sound detection for machine condition monitoring [2], which is the follow-up task from DCASE 2020 task2 to DCASE 2024 task2 [3].

In view of the above situation, we propose an adaptive anomaly detection framework, which can ensure that the framework can operate efficiently regardless of the existence of attribute information, and is closer to the actual situation in industry. The detection framework is a two stage fusion adaptive anomaly detection scheme which combine adaptive filtering for denoising and separation and neural network-based classifier. Firstly, we consider a preprocessing scheme based on adaptive filtering to perform noise reduction or separation operations on the noisy machine sounds, further improving the signal-to-noise ratio(SNR) of the machine sounds. For machines with attributes, we perform detailed anomaly detection by subdividing by all attributes. In the absence of attribute information, we utilize domain- specific information for effective detection. Finally, we extracted the embedding vectors through MobileFaceNet [4], and used Arcface [5] as the loss function of the model, and the back-end used an anomaly detection algorithm to evaluate the degree of abnormality.

2. METHODOLOGY

2.1. Adaptive filtering Preprocessing

The characteristic of adaptive filtering is that the parameters of its filter (such as weights) are constantly adjusted along with the changes of the input signal. Even when the statistical characteristics of the signal change, the adaptive filter can automatically adjust its parameters to adapt to the new signal characteristics. Thus, adaptive filtering is typically used in scenarios where the statistical characteristics of signals are unpredictable or constantly changing, such as noise cancellation, echo cancellation, channel estimation, etc. Inspired by this, we attempted to filter out the noise and interference signals mixed in the machine sound for the scenario where the machine sound and noise were mixed in Task 2.

The principle block diagram of the adaptive filter is shown in the following figure. The input signal x(n) passes through a parameter-adjustable digital filter to generate an output signal y(n), which is compared with the expected signal d(n) to form an error signal e(n). The filter parameters are adjusted through an adaptive algorithm to ultimately minimize the mean square value of e(n). Adaptive filtering can utilize the results of the filter parameters obtained at the previous moment to automatically adjust the filter parameters at the current moment, so as to adapt to the unknown or time-varying statistical characteristics of the signal and noise, thereby achieving optimal filtering. An adaptive filter is essentially a Wiener filter that can adjust its own transmission characteristics to achieve the optimal result. Adaptive filters do not require prior knowledge about the input signal, have a small amount of calculation, and are particularly suitable for real-time processing. The parameters of the Wiener filter are fixed and suitable for stationary random signals. The parameters of the Kalman filter are time-varying and are suitable for non-stationary random signals.

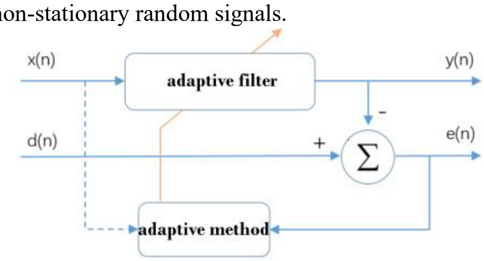

Figure 1: The principle block diagram of the adaptive filter.

For machines that provide noise data, we set: x(n) as the original noisy audio, d(n) as the pure noise audio, and y(n) as the estimated noise audio. In the subsequent inference process, the filtered audio can be obtained by subtracting y(n) from x(n). For machines that don't provide noise data, we attempt to extract pure noise segments from the original noisy audio and perform the above filtering process. If pure noise segments cannot be extracted, no filtering is performed, and the original audio is used.

2.2. Classification model

We use the data from DCASE 2025, and take the filtered data as the input to the anomaly sound detection system. The entire anomaly sound detection system comprises a front-end feature extractor and a back-end anomaly detector. The front-end feature extractor serves as a supervised attribute classification model, which is constructed on the MobileFaceNet network structure. This attribute classification model is simultaneously supervised by machine type, attribute, and domain information. To minimize intra-class distance and maximize inter-class distance, ArcFace is employed as the model's loss function.

2.3. Feature extraction

In the step of feature extraction, we deeply analyze four types of feature images, including spectrogram, original waveform graph, Mel-spectrogram and Short-Time Fourier Transform (STFT) spectrogram. In order to ensure the validity and accuracy of the selected features, we performed combined validation of these features from multiple perspectives on the development set. After rigorous experiments and analysis, two feature images, Mel-spectrogram and STFT spectrogram, are finally selected as input. Meanwhile, at the physical level, these two features focus on sound and vibration information, respectively. For both feature representations, we used the magnitude spectrogram with window length 1024 and window shift size 512. These two kinds of features are respectively fed into the embedding vector extraction network based on MobileFaceNet for processing. Through this process, we obtain two kinds of feature vectors and combine them to form the final embedding feature vector.

2.4. Back-end anomaly detector

The back-end anomaly detector of the system consists of three steps [6]. For the source domain, the k-means algorithm is applied to obtain multiple class centers for each machine type, and then all cosine distances between the given test sample and these class centers are calculated. While for the target domain, the cosine distance between the given test sample and all 10 normal samples of the same machine type is calculated. Finally, for a test sample, the minimum value among all cosine distances calculated for this sample is taken as the anomaly score for this sample. Thus, the anomaly score of the sample can indicate whether the sample is abnormal.

3. RESULTS

Our systems include different types of filters and we set different filtering parameters. The proposed system is compared with the benchmark system of DCASE 2025 in Challenge Task 2, namely Baseline MSE and Baseline MAHALA [7]. Our system outperforms the baseline system for most of the machines in the development set, as shown in Table 1.

Table 1: Anomaly detection results for the proposed system

	Method	Baseline MSE	Baseline MAHALA	Proposed system
ToyCar	AUC(source)	71.05%	73.17%	61.58%
	AUC(target)	53.52%	50.91%	67.04%
	pAUC	49.70%	49.05%	51.37%
	score	56.73%	55.87%	59.26%
ToyTrain	AUC(source)	61.76%	50.87%	75.76%
	AUC(target)	56.46%	46.15%	65.74%
	pAUC	50.19%	48.32%	50.37%
	score	55.73%	48.37%	62.16%
bearing	AUC(source)	66.53%	63.63%	63.18%
	AUC(target)	53.15%	59.03%	49.54%
	pAUC	61.12%	61.86%	49.00%
	score	59.75%	61.45%	53.17%
fan	AUC(source)	70.96%	77.99%	57.76%
	AUC(target)	38.75%	38.56%	61.60%
	pAUC	49.46%	50.82%	53.95%
	score	49.90%	51.34%	57.60%
gearbox	AUC(source)	64.80%	73.26%	75.84%
	AUC(target)	50.49%	51.61%	76.92%
	pAUC	52.49%	55.07%	59.63%
	score	55.26%	58.61%	69.84%
slider	AUC(source)	70.10%	73.79%	77.46%
	AUC(target)	48.77%	50.27%	65.04%
	pAUC	52.32%	53.61%	57.21%
	score	55.68%	57.58%	65.55%
valve	AUC(source)	63.53%	56.22%	70.92%
	AUC(target)	67.18%	61.00%	82.76%
	pAUC	57.35%	52.53%	54.05%
	score	62.42%	56.37%	67.14%
All (hmean)	AUC(source)	66.78%	65.51%	68.11%
	AUC(target)	51.39%	50.05%	65.42%
	pAUC	52.94%	52.72%	53.43%
	score	56.26%	55.34%	61.62%

4. CONCLUTION

In this technical report, we introduce our submission systems to DCASE 2025 challenge task 2. We propose an attribute classification scheme based on adaptive filtering preprocessing and MobileFaceNet network, combined with the k-means algorithm of the back-end, to achieve accurate detection of abnormal sounds. When the attribute information is complete, we make a detailed division of each type of attribute, so as to realize the depth monitoring of abnormal conditions. Even in the case of missing attribute information, we divide the domain information as attributes to effectively identify abnormal conditions.

In addition, in the feature extraction stage, we comprehensively consider the characteristics of the machine in the two key dimensions of sound and vibration in the industrial environment. We combine the advantages of Mel-spectrogram and Short-Time Fourier Transform (STFT) spectrogram, and this innovative initiative significantly improves the universality and effectiveness of anomaly detection systems. This integrated feature extraction approach ensures that the system can more accurately identify and process abnormal sound data from different sources and environments.

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

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

[4] C. D. Jones, A. B. Smith, and E. F. Roberts, "A sample paper in conference proceedings," in *Proc. IEEE ICASSP*, 2003, vol. II, pp. 803-806.

[5] A. B. Smith, C. D. Jones, and E. F. Roberts, "A sample paper in journals," *IEEE Trans. Signal Process.*, vol. 62, pp. 291-294, Jan. 2000.

[6] Wilkinghoff K. Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[7] 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.