ANOMALOUS SOUND DETECTION SYSTEM WITH GAN AND AE FOR DCASE2023 CHALLENGE TASK 2

Technical Report

Tianxin Wu[*]

Suzhou Qimengzhe Technology Co., Ltd
Suzhou, China
tianxin.wu@qdreamer.com

ABSTRACT

This report describes the system for DCASE 2023 Challenge Task 2, which aims to detect anomalous machine states through sound using machine learning methods, where the training dataset itself does not contain any anomalous examples. We constructed a method based on Generative Adversarial Networks (GAN). The system achieved the best score of 86.20% on the development dataset for the machine type "slider," while the corresponding baseline score based on autoencoders was 69.06% and 83.18%.

Index Terms— Dcase, anomalous detection, machine learning, machine health monitoring

1. INTRODUCTION

DCASE Challenge 2023 Task 2[1,2]: First-shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. In the First-shot ASD task, compared to previous years' tasks, systems will not be allowed to perform any hyperparameter tuning based on the performance metrics calculated with the ground truth, especially with anomalous samples provided for evaluating systems. In other words, the performance of the system will be evaluated based on its ability to accurately detect anomalous sounds in a single pass without any opportunity for fine-tuning based on the evaluation dataset. And the set of machine types in the development dataset and evaluation dataset are completely different.

The AEGAN-AD[6] model trains the discriminator and generator using WGAN-GP[7] and feature matching loss techniques, respectively. It detects anomalies from two complementary perspectives: the generator-based reconstruction error and the embeddings extracted from the discriminator. These embeddings are then processed by traditional detection algorithms.

2. BASELINE RESULTS

In order to give a clear picture of the Challenge Task 2, we include the baseline scores on Table 1 and Table 2. To present the results succinctly, the results in all tables in this report present only the harmonic mean of source AUC, target AUC and pAUC for each machine type. Here the harmonic mean is denoted as h-mean. The data used in this challenge is 16 kHz, single-channel audio. For more details, please see [3,4,5].

Table 1: Performance of the AE baseline method(MSE) (%)

	AUC-s	AUC-t	pAUC
ToyCar	69.22	45.44	50.21
ToyTrain	59.30	56.70	48.26
Bearing	66.12	55.00	50.16
Fan	73.34	35.48	60.05
Gearbox	60.46	61.78	53.79
Slider	69.06	48.60	56.53
Valve	55.82	52.04	54.53

Table 2: Performance of the AE baseline method(MAHALA) (%)

	AUC-s	AUC-t	pAUC
ToyCar	73.82	12.68	49.37
ToyTrain	54.60	42.30	47.84
Bearing	65.24	54.04	50.11
Fan	79.26	41.96	59.21
Gearbox	72.42	72.08	55.21
Slider	83.18	73.94	54.37
Valve	55.06	52.20	51.16

3. APPROACHES

We design an autoencoder which reconstructs Mel spectrograms and complement it with a discriminator, resulting in a GAN model. The AEGAN-AD model modifies the generator to function as an autoencoder that reconstructs mel-spectrograms. Meanwhile, the discriminator, is trained to differentiate between real spectrograms and those that have been reconstructed. The structure of G can be seen as an upside-down deep convolutional

GAN (DCGAN), with the encoder of G serving as the discriminator of DCGAN, and the decoder of G serving as the generator of DCGAN. The structure of D is similar to the encoder of G, but with a depth-wise convolutional layer replacing the final layer to promote the extraction of semantic embeddings. The model adopts layer normalization (LN) to prevent the statistics of the source domain from masking those of the target domain, preserving distinguishable features and scalability for both domains.

And the AEGAN-AD model uses the WGANGP approach to train the discriminator, but modifies the generator's loss function with a feature matching loss that focuses on the individual samples rather than the overall distribution. This approach avoids confusion in detecting anomalies by ensuring that both normal and anomalous samples are equally affected by the reconstruction error. To achieve this, the model extracts embeddings from the discriminator and compares the statistics of the real and reconstructed distributions. The generator's loss function combines this statistical difference with the mean squared error to promote a deep understanding of the input spectrogram and improve robustness. The feature matching loss term helps prevent denoising tendencies in the generator.

4. EXPERIMENTS AND RESULTS

The AEGAN-AD framework initially calculates the mel-spectrogram (with a 2048-point FFT and 128 mel filters) on a logarithmic scale. It then scales the result to the range of [-1, 1] by applying an affine transform and divides it into overlapping segments of 128x128 using a sliding window. Both networks use Adam optimizers with a learning rate of 0.0002. For the WGAN-GP parameters, λ is set to 10, and ncritic is 1. The model trains for 60 epochs with a batch size of 512, and a set of parameters is saved for each type of machine.

Each machine includes one sections, and each section is divided into two domains, i.e., source and target. The audio signals are single-channel with around 10 seconds in length, and the sampling rate is 16 kHz. There are 990 training samples from the source domain and 10 training samples from the target domain. 200 samples from both domains are used as the testing set, i.e., 100 from the source domain and 100 from the target domain, whose domain information is not specified for testing. And our system results are shown in the table3.

Table 2: Performance of the our method (%)

	AUC-s	AUC-t	pAUC
ToyCar	79.86	55.12	49.89
ToyTrain	60.42	63.48	53.53
Bearing	81.86	67.40	52.11
Fan	64.44	81.94	63.21
Gearbox	77.46	74.12	65.58
Slider	86.20	81.54	62.53
Valve	71.20	63.60	57.68

5. CONCLUSION

In this technical report, we have presented our system submitted for DCASE2023 challenge Task 2, which is an system based on GAN and AE models. The systems are evaluated on the official dataset of DCASE2023 challenge Task 2. The results show that all our systems can outperform the baseline systems.

6. REFERENCES

[1] http://dcase.community/challenge2023/.

[2] K. Dohi, K. Imoto, N. Harada, D. Niizumi, and T. Nishida, “Description and discussion on dcase 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring,” 2023, 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 miniaturemachine operating sounds for anomalous sound detection under domain shift conditions,” in Proc. 6th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE), 1–5. Barcelona, Spain, November 2021..

[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] H. Harada, Y. Musashijima, and D. Niizumi, “First-shot anomaly detection for machine condition monitoring: a domain generalization baseline,” In arXiv e-prints: 2303.00455, 2023.

[6] Jiang A, Zhang W Q, Deng Y, et al. Unsupervised Anomaly Detection and Localization of Machine Audio: A Gan-Based Approach[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[7] Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in neural information processing systems 30 (2017).