UNSUPERVISED ABNORMAL SOUND DETECTION BASED ON MACHINE CONDITION MIXUP

Technical Report

Yafei Jia[1], Jisheng Bai[1,2], Siwei Huang[1], Jianfeng Chen[1,2]

1 Joint Laboratory of Environmental Sound Sensing,
School of Marine Science and Technology,
Northwestern Polytechnical University, Xi′an, China

2 LianFeng Acoustic Technologies Co., Ltd. Xi′an, China

{ j.yafei, baijs, hsw838866721}@mail.nwpu.edu.cn, chenjf@nwpu.edu.cn

ABSTRACT

Anomaly detection has a wide range of applications such as finding fraud cases in industry or indicating network intrusion in network security. Anomalous sound detection (ASD) for machine condition monitoring can detect anomalies in advance and prevent causing damage. However, the operational conditions of machines often change, leading to the different acoustic characteristics between training and test data. Domain generalization techniques are required to adapt the model to different conditions. In this paper, we present an unsupervised method for ASD, which uses MSE, KLD, and BCE as joint loss and Condition-Mixup data augmentation strategies for the GAN-VAE model. The proposed Condition-Mixup strategy mixes data from the target domain of the unified condition in the time domain to balance the difference in data volume between the source domain and the target domain. In addition, we adopted a GAN-VAE model to learn common potential information between the source and target domains. Finally, we use acoustic representation to train anomaly detectors to detect abnormal sounds. The experimental results on the DCASE2023 taks2 development dataset show that our method outperforms the baseline system.

Index Terms— Anomalous sound detection, domain generalization, condition mixup, GAN-VAE

1. INTRODUCTION

Anomaly detection has been valued by researchers in recent years, and has been widely used in surveillance and mechanical equipment monitoring. A damaged bearing may not be found visually, but it can be detected acoustically through different acoustic manners. In addition, the acoustic monitoring system is cheap and easy to develop. Developing a reliable anomalous sound detection (ASD) system for the early detection of abnormal events can optimize and save a lot of resources in industrial production.

In DCASE2023 task2 (“First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring”) [1], it is necessary to detect the abnor-mal sound of the machine using unsupervised methods, because we can only get the normal sound of the machine for training. In the task, the test data is mixed with samples not only affected by data in source domain, but also affected by data in target domain, and whether each sample belongs to the source domain or the target domain is not specific. In addition, 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.

2. PROPOSED METHOD

2.1. Condition mixup

In this task, the source domain and the target domain of the training data are extremely unbalanced. When loading training data, the batch may not contain the data from target domain. This leads to the overfitting on the source domain of the final ASD model. Therefore, the proposed condition mixing (C-Mixup) strategy mixes data from the target domain under the same working conditions with weights in the time domain. This method can balance the difference in data volume between the source domain and the target domain, and better model different domains.

In the development dataset of DCASE2023 task2, each audio label provides the type of machine and specific operating conditions, such as the placement of microphones and background noise. The C-Mixup method is to select two audio files from the target domain that have one or more identical tasks and perform a weighted mixing operation.

$$\mathbf{X} = \delta \cdot \mathbf{X}_{(\mu_1,\mu_2,\mu_3)} + (1-\delta) \cdot \mathbf{X}_{(\nu_1,\nu_2,\nu_3)} \quad (1)$$

Where $\mathbf{X}$ is the time-domain representation of the target domain audio, μ and ν are the operating conditions of the machine respectively, and $(\mu_1, \mu_2, \mu_3) \cap (\nu_1, \nu_2, \nu_3) \notin \varnothing$. δ is the weight of the mixture, where y is greater than 0 but less than 1.

2.2. Model

Variational Autoencoder (VAE) [2]: VAE is a generative model that can learn the underlying distribution of the input data and generate new samples. It consists of an encoder and a decoder, which are trained together to maximize the likelihood of reconstructing the original data and minimizing the distance between the latent variables and a prior distribution.

Generative Adversarial Networks-VAE(GAN-VAE) [3]: GAN-VAE is a hybrid model that combines the generative adversarial network (GAN) with the variational autoencoder (VAE) to generate high-quality and diverse samples. The GAN component learns the distribution of real data and generates realistic samples, while the VAE component learns the distribution of latent variables and generates diverse samples.

2.3. Loss function

In the experiment, we employed different loss functions for different models.

We found that during the model training process, due to the scarcity of samples in the target domain, the VAE model cannot fit the distribution of the target domain data well, which affects the overall performance of the model. Therefore, different weights are adopted for the reconstruction loss of the source domain and the target domain.

$$L_{VAE} = MSE_{s\_loss} + \alpha MSE_{t\_loss} + \beta KL_{loss} + \gamma BCE_{loss} \quad (2)$$

Where s and t represent the source and target domain data, α, β and γ represent the weights of corresponding losses respectively, with values ranging from 0 to 1. When neither of them is 0, the network is GAN-VAE. When α is 0, the network degenerates into VAE. When α and β are both equal to 0, the network degenerates into AE.

3. EXPERIMENTAL SETTINGS

3.1. Dataset

The dataset of task2 consists of seven types of machines: toyCar, toyTrain, bearing, fan, gearbox, slider and valve [4].

The development dataset consists of only one part provided by each machine, which includes approximately 990 normal recordings in the source domain and 10 normal recordings in the target domain for training, as well as approximately 100 segments of normal and abnormal recordings in the source and target domains for testing. Each recording is a 10 second audio recording that records the running sound of the machine and its environmental noise.

3.2. Feature

The sample rate used in the experiments is 16KHz, and we applied log-Mel spectrograms, we use 128 Mel-bins and n frames of 64. In addition, the 5 consecutive frames are concatenated and 640 dimensions are input to model.

4. RESULTS

We conducted experiments using the development and additional training dataset of DCASE2023 task2, and compared it with the baseline system[5]. The AUC scores of the experiment are shown in Table 1.

5. CONCLUSIONS

In this paper, we present an unsupervised method for ASD, which uses MSE, KLD, and BCE as joint loss and Condition-Mixup data augmentation strategies for the GAN-VAE model. The proposed Condition-Mixup strategy mixes data from the target domain of the unified condition in the time domain to balance the difference in data volume between the source domain and the target domain. In addition, we adopted a GAN-VAE model to learn common potential information between the source and target domains. Finally, we use acoustic representation to train anomaly detectors to detect abnormal sounds. The experimental results on the DCASE2023 taks2 development dataset show that our method outperforms the baseline system.

6. REFERENCES

[1] Dohi K, Imoto K, Harada N, et al. Description and Discussion on DCASE 2023 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring[J]. arXiv preprint arXiv:2305.07828, 2023.

[2] Burgess C P, Higgins I, Pal A, et al. Understanding disentangling in β-VAE[J].arXiv preprint arXiv:1804.03599, 2018.

[3] Genevay A, Peyré G, Cuturi M. GAN and VAE from an optimal transport point of view[J]. arXiv preprint arXiv:1706.01807, 2017.

[4] Dohi K, Nishida T, Purohit H, et al. MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task[J]. arXiv preprint arXiv:2205.13879, 2022.

[5] Noboru Harada, Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, and Masahiro Yasuda. First-shot anomaly detection for machine condition monitoring: a domain generalization baseline. In arXiv e-prints: 2303.00455, 2023.

Table 1: AUC scores of experiments

	Model (AD)	Score	ToyCar	ToyTrain	Bearing	Fan	Gearbox	Slider	Valve	mean
--	Baseline (Mahala)	s_AUC	74.5%	56.0%	87.1%	71.9%	65.2%	84.0%	56.3%	70.7%
		t_AUC	43.4%	42.5%	46.0%	70.8%	55.3%	73.3%	51.4%	54.7%
		pAUC	49.2%	48.1%	59.3%	54.3%	51.4%	54.7%	51.1%	52.6%
Submission 3,4	VAE (Mahala)	s_AUC	71.6%	56.7%	92.3%	75.2%	63.5%	85.5%	56.5%	**71.6%**
		t_AUC	52.9%	46.3%	76.9%	73.3%	60.2%	77.3%	51.4%	**62.6%**
		pAUC	48.8%	48.1%	68.4%	57.2%	51.9%	54.4%	51.0%	**54.2%**
Submission 1,2	GAN-VAE (Mahala)	s_AUC	72.2%	61.9%	92.8%	77.0%	64.4%	87.8%	55.9%	**73.2%**
		t_AUC	52.7%	45.8%	74.3%	73.6%	61.5%	78.8%	50.4%	**62.5%**
		pAUC	50.0%	48.2%	66.2%	56.2%	51.8%	55.1%	50.8%	**54.0%**