UNSUPERVISED ABNORMAL SOUND DETECTION SYSTEM BASED ON MULTI – ATTRIBUTE

Technical Report

Chengliang Jiang

PFU SHANGHAI Co.,LTD
46 Building 4~5 Floors, 555 GuiPing Road
XuHui District, Shanghai 200233, CHINA
jiang_chengliang.pfu@fujitsu.com

Yan Wang

PFU SHANGHAI Co.,LTD
46 Building 4~5 Floors, 555 GuiPing Road
XuHui District, Shanghai 200233, CHINA
Yan.wang.alex@gmail.com

ABSTRACT

This technical report describes submission to DCASE 2023 Task 2. In this report, we propose a multi-attribute training method for anomalous sound detection, which includes feature preprocessing, model training, center loss, triplet loss, and anomaly score selection. The experimental results show that our anomalous sound detection model is superior to the official model.

Index Terms— Anomalous sound detection, feature preprocessing, triplet loss, center loss

1. INTRODUCTION

In this task, the purpose of anomalous sound detection is to detect whether the sound emitted from the target machine type is abnormal. But we don't have anomalous sound sample. We usually train the classifier with different attributes of the machine as an auxiliary task. This is based on the realistic assumption that the same kind of machine using different power or different speeds usually produces somewhat different sounds. A classifier can be built to distinguish machines with different attributes. In the test phase, we use cosine similarity to represent outlier scores. We used a variety of noise and masks to prevent overfitting of the model. For each machine type, we trained a dedicated classifier to distinguish its attributes. We use exactly the same model parameters, and we want to not use specialized operations on any machine type.

2. PROPOSED METHOD

2.1. Model

During experiments, we found efficientnet and resnet did not actually improve detection performance. In this work, we train our anomalous sound detection model based on mobilefacenet. The model structure is provided in Table 1.

Table 1: Model architecture, where k is the number of attributes, c is the output channels, n denotes the number of Inverted residuals blocks, and s is stride.

Operator	c	n	s
Conv2d 3x3	64	-	2
Conv2d 3x3	64	-	1
Residual	128	-	2
Blockneck	128	2	1
Residual	128	-	2
Blockneck	128	2	1
Residual	128	-	2
Blockneck	128	2	1
Conv2d 3x3	512	-	1
Linear Conv2d 7x7	512	-	1
Flatten	-	-	-
Linear	128	-	-
Dropout	-	-	-
Linear	k	-	-

2.2. Anomaly score

The calculation of anomaly scores is very important. We tested Mahalanobis distance and cosine similarity as outliers and found that better AUC performance was obtained using corotational similarity. At the same time, we separate source domain data and target domain data as the basis for calculating cosine similarity to obtain better exception detection performance.

2.3. Triplet loss

We use triplet loss to push the vectors generated by source domain data and target domain data far apart to improve their separation, so that the corresponding domain classification can be known through cosine similarity during testing, and the corresponding domain data can be used as the benchmark of cosine similarity.

2.4. Center loss

We used center loss to reduce the class spacing during attribute classification, and found that this improved the final exception detection performance.

2.5. Separation of domains

We found that when only the source or target domain data is used as the base feature, the exception detection performance of the corresponding domain will be greatly improved, and the exception detection performance of the other domain will be greatly decreased, so it is critical to find the corresponding domain of audio. We finally determine the ownership of the domain by the score of target.

	Method	Baseline AE	Ours
ToyCar	AUC-s	74.53%	45.72%
	AUC-t	43.42%	62.54%
	pAUC	49.18%	52.52%
ToyTrain	AUC-s	55.98%	49.43%
	AUC-t	42.45%	58.26%
	pAUC	48.13%	49.57%
bearing	AUC-s	65.16%	62.02%
	AUC-t	55.28%	53.34%
	pAUC	51.37%	51.21%
fan	AUC-s	87.1%	85%
	AUC-t	45.98%	88.68%
	pAUC	59.33%	68.84%
gearbox	AUC-s	71.88%	65.08%
	AUC-t	70.78%	53.58%
	pAUC	54.34%	53.78%
slider	AUC-s	84.02%	39.65%
	AUC-t	73.29%	33.84%
	pAUC	54.72%	48.31%
valve	AUC-s	56.31%	41.5%
	AUC-t	51.4%	48.5%
	pAUC	51.08%	49.94%

3. EXPERIMENTS

3.1. Dataset and audio processing

All experiments are based on the DCASE 2023 task 2 dataset [1, 2, 3, 4], which includes 14 machine types. The kaldi.fank is used as the input feature. Three seconds of audio are randomly selected from ten seconds of audio files for conversion, and mask and noise are added after conversion to kaldi.fank.

3.2. Experimental settings

We use pytorch for experiments. In training, the model is trained for 2000 epochs with Adam as the optimizer, where the batch size is 48 and the learning rate is 0.0003. The alpha hyperparameter for center loss is set to 0.1.

4. RESULTS AND DISCUSSIONS

We evaluate the detection performance using the area under the receiver operating characteristic curve (AUC) and the partial AUC(pAUC) with p = 0.1. Table 2 shows our experimental results. Because performance increases or decreases slightly depending on the machine type, and because the resulting blind test set and the public test set have completely different machine types, the performance of the public test set is not entirely indicative of the blind test set's performance.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Ryo Tanabe, Takashi Endo, and Yohei 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.