ANOMALOUS SOUND DETECTION WITH PANNS MOBILENETV1 EMBEDDINGS

Technical Report

Ilya Kodua, Sophia Kazakova, Andrey Semenov

ITMO University
Department of Speech Information Systems
Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia
sophie.a.kaz@gmail.com

ABSTRACT

This technical report describes the PANNs MobileNetv1-based approach for DCASE 2022 Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques task [1]. The objective of this task is to determine whether the sound emitted from the target machine class is normal or anomalous while having only the normal data for training purposes.

We extract embeddings using external PANNs MobileNetV1 pre-trained model [2]. For anomaly score assignment we concatenate the embeddings obtained and then calculate the cosine distance to the first nearest neighbor in the embedding space for this sample's class and section.

For this report the GitHub code is available. PANNs embedding extraction is [3], and anomaly score calculation is [4].

Index Terms— PANNs, MobileNetv1, anomaly detection

1. INTRODUCTION

The task of DCASE 2022 Challenge 2 concerns solving the problem of searching for abnormal audio recordings using exclusively normal ones when shifting the domain, which consists of different recording conditions. In real conditions, the background noise when recording audio is constantly changing, and the details of the machines are constantly wearing out, which affects the quality of anomaly detection. Therefore, anomaly detection should be invariant to background noise and changes in various parameters of mechanical parts that are directly related to equipment wearing out.

The authors of the task propose a dataset consisting of 10-second recordings of the operation of 7 types of devices: fan, gearbox, bearing, slider, toy car, toy train, and valve [5], [6]. Participants are offered to train a neural network exclusively on sound fragments of a 'normal' class, i.e. recordings of the proper operation of equipment. 'Anomalies' appear only in the test dataset. The abnormal data was obtained by purposefully damaging the machines. The characteristics of breakdowns are not given, there is no additional data except for the wav audio file class (normal/anomaly).

The challenge of the dataset lies in the domain shift, i.e. changing conditions for recording data. This means that audio samples for each type of equipment were recorded at different load levels, environmental characteristics (including background noises), etc. The test sample contains records in conditions that were not presented in the training sample.

Moreover, an additional study of the data revealed that most of the data for human hearing is not obvious. Without additional preparation, it is difficult and sometimes impossible to determine the boundary between the target sound and background noise, the difference between normal recording and anomaly.

For further comparison of the results with the baseline, metrics submitted by the authors are presented in table 1. Every value is the corresponding harmonic mean.

	AE baseline			MNet baseline		
target class	AUC source	AUC target	mean pAUC	AUC source	AUC target	mean pAUC
bearing	54.42	58.38	51.98	60.58	59.94	57.14
fan	78.59	47.18	57.52	70.75	48.22	56.9
gearbox	68.93	62.64	58.49	69.21	56.19	56.03
slider	77.95	47.67	55.78	65.15	38.23	54.67
ToyCar	90.41	34.81	52.74	59.12	51.96	52.27
ToyTrain	76.32	23.35	50.48	57.26	45.90	51.52
valve	52.01	49.46	50.36	67.09	57.22	62.42

Table 1: PANNS MobileNetV1 system results.

2. PROPOSED APPROACH

In this submission, the MobileNet neural network from the PANNS library was used for embedding extraction. MobileNet architecture was chosen as it was widely used in previous years and showed strong performance. Although, we used MobileNet of the first generation, as it showed greater accuracy on AudioSet data than MobileNetV2 in the PANNS library.

The anomaly score is calculated as the cosine distance to the nearest neighbour among the embeddings. This distance calculates separately for each class and each section. This approach was introduced in one of the top teams from DCASE 2021 [1], [7].

3. EXPERIMENTS

Results are presented in table 2. For most of the machine classes target AUC score exceeds the baseline. For source AUC and pAUC, the relative result is not so clear.

Every value is the corresponding harmonic mean.

4. REFERENCES

[1] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Takashi Endo,

target class	AUC source	AUC target	pAUC
bearing	56.42019	54.54773	51.46632
fan	61.21901	57.40769	52.21617
gearbox	60.43659	61.72283	53.33228
slider	74.51367	63.18836	56.82731
ToyCar	62.42814	50.07193	51.74736
ToyTrain	56.97004	53.97974	52.12997
valve	61.97372	56.18598	50.82024

Table 2: PANNS MobileNetV1 system results.

Masaaki Yamamoto, and Yohei Kawaguchi. Description and discussion on DCASE 2022 challenge task 2: unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques. In arXiv e-prints: 2206.05876, 2022.

[2] Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Mark Plumbley. PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. 2019.

[3] Ilya Kodua. PANNS embedding extractor. [Online]. Available: `https://github.com/IlyaKodua/panns_embeddings_extractor`

[4] Ilya Kodua. Anomaly Score Calulation. [Online]. Available: `https://github.com/IlyaKodua/calculate_anomaly_score`

[5] 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 arXiv e-prints: 2205.13879, 2022.

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

[7] Morita, Kazuki et al. ANOMALOUS SOUND DETECTION USING CNN-BASED FEATURES BY SELF SUPERVISED LEARNING. DCASE2021 Challenge, Invalid Date 2021