UNSUPERVISED DETECTION OF ANOMALOUS SOUNDS TECHNICAL REPORT

Technical Report

Yao Xiao

Tsinghua University
School of Software, Beijing, China
xiao-y18@mails.tsinghua.edu.cn

ABSTRACT

This report describes the solution to Task 2 of the DCASE 2020 challenge. Besides the autoencoder-based unsupervised anomaly detector used in the baseline, the classifier-based unsupervised anomaly detector is used and the classification error of the normal or anomalous machine sounds is used as anomaly score.

Index Terms— Autoencoder, convolutional neural, classification error

1. INTRODUCTION

The autoencoder-based unsupervised anomaly detector is based on an basic assumption that the trained model will give high reconstruction error for anomalous machine sounds. We train a model to classify the machine sounds into machine types(ToyCar ToyConveyor, fan, pump, slider, valve). We just take normal machine sounds as train dataset and we give the assumption that the trained classifier will give high classification error for anomalous machine sounds.

2. ARCHITECTURES

Two architectures of classifiers are used for this submission. One is simple convolutional neural networks. Dropout is applied. Convolutional layers used ReLU activation BatchNormalization. The model take raw waveform of machine sounds as input. The other is simple fully connected networks and it take log-Mel spectrogram as the input feature. The feature vector is the same as the baseline.

Table 1: Architecture of the convolutional networks.

layer	output	kernel	stride
Conv1D+ReLU+BN	16	9	3
Conv1D+ReLU+BN	16	9	3
MaxPooling+dropout(p=0.1)		16	
Conv1D+ReLU+BN	32	5	2
Conv1D+ReLU+BN	32	5	2
MaxPooling+dropout(p=0.1)		4	
Conv1D+ReLU+BN	64	3	1
Conv1D+ReLU+BN	64	3	1
MaxPooling+dropout(p=0.1)		4	
Conv1D+ReLU+BN	128	3	1
Conv1D+ReLU+BN	128	3	1
AvgPooling+dropout(p=0.1)	128		
Dense	256		
Dense	64		
Dense	6		

Table 2: Architecture of the fully connected networks.

layer	output	kernel	stride
Dense+ReLU+BN	128		
Dense+ReLU+BN	128		
Dense+ReLU+BN	128		
Dense+ReLU+BN	128		
Dense+ReLU	16		
Dense+Softmax	6		

3. TRAIN

SGD is used for optimization. There are 20 epochs for each model. We use a batch size of 64 for convolutional networks and a batch size of 512 for fully connected networks. We use the cross entropy as a loss function.

4. RESULTS

The results of development test data are showed blow.

Table 3: Result of the convolutional networks

machine type	id	AUC	pAUC
ToyCar	01	0.582911255	0.499088631
	02	0.629509434	0.526996737
	03	0.486684636	0.497148532
	04	0.51909434	0.518257909
	Avg.	0.554549916	0.510372952
ToyConveyor	01	0.493482813	0.494660684
	02	0.491802817	0.492253521
	03	0.497891731	0.49237325
	Avg.	0.494392453	0.493095818
fan	00	0.31544226	0.488081383
	02	0.670877437	0.570590822
	04	0.45433908	0.507713249
	06	0.757936288	0.630995772
	Avg.	0.549648767	0.549345306
pump	00	0.84506993	0.704821494
	02	0.485405405	0.489331437
	04	0.87175	0.793999422
	06	0.753284314	0.538183695
	Avg.	0.738877412	0.631584012
slider	00	0.155294944	0.476324238
	02	0.839007491	0.580327223
	04	0.933483146	0.826387376
	06	0.49	0.490870032

	Avg.	0.604446395	0.593477217
valve	00	1	1
	02	0.497375	0.497171731
	04	0.6295	0.626933405
	06	0.558333333	0.558333333
	Avg.	0.671302083	0.670609617

Table 4: Result of the fully connected networks

machine type	id	AUC	pAUC
ToyCar	01	0.567570346	0.529785045
	02	0.56702965	0.52757838
	03	0.475309973	0.49374734
	04	0.49954717	0.49521918
	Avg.	0.527364285	0.511582486
ToyConveyor	01	0.506390625	0.502161891
	02	0.499396127	0.500878779
	03	0.49542562	0.49605135
	Avg.	0.500404124	0.49969734
fan	00	0.487248157	0.538730118
	02	0.960557103	0.853980355
	04	0.619224138	0.604204477
	06	0.831135734	0.847062254
	Avg.	0.724541283	0.710994301
pump	00	0.90006993	0.73500184
	02	0.528378378	0.550497866
	04	0.96405	0.843684211
	06	0.872156863	0.611971104
	Avg.	0.816163793	0.685288755
slider	00	0.704410112	0.631726789
	02	0.684831461	0.509166174
	04	0.936011236	0.8710822
	06	0.868202247	0.651685393
	Avg.	0.798363764	0.665915139
valve	00	0.596554622	0.564794339
	02	0.569	0.488596491
	04	0.693166667	0.54122807
	06	0.558	0.520175439
	Avg.	0.604180322	0.528698585

5. SUBMISSIONS

We will combine the results of our models and the result of baseline to generate the final results that we submit.

6. REFERENCES

[1] Yuma Koizumi, Shoichiro Saito, Hisashi Uematsu, Noboru Harada, and Keisuke Imoto. ToyADMOS: a dataset of miniature-machine operating sounds for anomalous sound detection. In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 308–312. November 2019

[2] Harsh Purohit, Ryo Tanabe, Takeshi Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. MIMII Dataset: sound dataset for malfunctioning industrial machine investigation and inspection. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), 209–213. November 2019.

[3] Yuma Koizumi, Yohei Kawaguchi, Keisuke Imoto, Toshiki Nakamura, Yuki Nikaido, Ryo Tanabe, Harsh Purohit, Kaori Suefusa, Takashi Endo, Masahiro Yasuda, and Noboru Harada. Description and discussion on DCASE2020 challenge task2: unsupervised anomalous sound detection for machine condition monitoring. In arXiv e-prints: 2006.05822, 1–4. June 2020.