Machine Condition Monitoring from Acoustic Signatures using Auto-Encoders

Technical Report

Nitesh Kumar Chaudhary nitesh.lnmiit@gmail.com
Josey Mathew josey.mathew@gmail.com
Sunil Sivadas sunil.sivadas@gmail.com

Corp Lab, NCS Pte. Ltd.
5 Ang Mo Kio Street 62, NCS Hub, Singapore, 569141

ABSTRACT

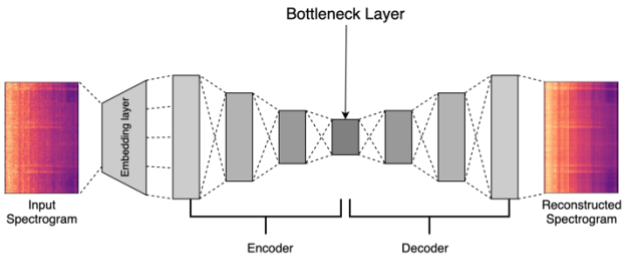
Condition monitoring of machinery is critical for early detection and prevention of failures in factories. Recent advancements in machine learning is driving the development of data driven tools like monitoring acoustic signatures from microphones. This report presents a modified dense connected autoencoder (AE) and convolutional autoencoder (CAE) trained to minimize the acoustic spectrogram reconstruction error during normal operation of the machine. The model is pre-trained on machines of similar type and then fine-tuned on a specific machine. The reconstruction error is used as the anomaly score for an unseen acoustic sample. The proposed model improves performance compared to baseline system for the DCASE2020 challenge task2 [2].

Index Terms— Anomaly detection, acoustic condition monitoring, autoencoders, convolution autoencoders, DCASE Challenge

1. INTRODUCTION

In this technical report we present our system solution for DCASE challenge task 2 [1, 2] problem defined by unsupervised detection of anomalous sounds for machine condition monitoring. The goal of the challenge Task 2 is to evaluate and identify whether the sound emitted from a target machine is normal or anomalous. System of anomaly detection varies from normal classification problem where we have labelled set of datasets for each of the type of event to be classified. On contrary in context of anomaly detection outlier cannot form a dense cluster as available estimators assume that the anomalies are located in low density regions because in real-world factories, actual anomalous sounds rarely occur and are highly diverse. Therefore, exhaustive patterns of anomalous sounds are impossible to deliberately gather from all possible set of data by which anomaly can occur from different type of machines. Thus, challenge come under category of unsupervised learning where we have to detect unknown anomalies that were not observed in training set. Only normal dataset has been provided for development of system and validation data consist of normal and anomaly for testing.

Baseline system provided for this DCASE challenge is based on dense connected Autoencoder system where bottleneck layer has far fewer neurons than typical deep learning models, so the autoencoder model has to find a way to represent data by letting go of all its noise and capture most relevant representation of original inputs. It is expected that autoencoder will do a really good job at reconstructing normal sound generated from different category of machines in development set, as that is exactly what the autoencoder was trained to do — and if we were to look at the reconstruction error between the input spectrogram and the reconstructed spectrogram, we would find that it's quite low for normal category of sound and high for anomaly as it has never seen anomaly samples in training set. On top of baseline system, we have proposed convolutional autoencoder (CAE) trained to minimize the reconstruction error during normal operation of machine. We have used embedding layers for customization CAE for different machine Id.



Operation	Parameters
Convolution2D and eLU activation	Kernel size 3x3
Maxpooling2D	Pool size 2x2
UpSampling2D	Size 2x2, Interpolation: Bilinear

Figure 1. Id specific autoencoder model architecture

2. DATASET

There are six categories of dataset used for this task and were collected from two different source ToyADMOS [3] and the MIMII Dataset [4]. Each recording is a single-channel (approximately) 10-sec length audio that includes both a target machine's operating sound and environmental noise. Originally these datasets were collected using multi-channel microphone, but for this challenge first channel of multichannel recording downsampled to 16kHz have been provided. Each category of machine data gathered from 6-7 different machine Ids.
Data is available in three sets, first development set consist of train data (labelled as normal), test (labelled as normal and anomaly) second additional data again for training (labelled as normal) and evaluation set (unlabeled). Table 1 shows an overview of dataset provided DCASE Task 2, which consists of a development dataset, an additional training dataset, and an evaluation dataset.

Class	Development set		Additional set	Evaluation set
	#Train	#Test	#Train	#Eval
ToyCar	4000	2459	3000	1545
ToyConveyor	3000	3509	3000	1665
Fan	3675	1875	2846	1342
Pump	3349	856	2417	777
Slider	2804	1290	2370	834
Valve	3291	879	2531	940

Table 1. DCASE Task2 machine data for anomaly detection

3. SYSTEM DETAILS

For classification and regression task in general CNN combines the convolution and pooling layers. A convolution layer extracts local spatial patterns, and a pooling layer reduces the amount of data while retaining useful information. CNN's ability to extract complex hidden features from high dimensional data with complex structure has enabled its use as feature extractors in outlier detection for both sequential and image dataset [6]. Inspired by CNN's widely used techniques of extracting the patterns of spatial arrangement of features for data compression, outlier detection in computer vision, fraudulent detection in textual data and reconstruction task, we have extended the baseline system of dense connected AE network to Convolution autoencoder (CAE) for machine anomaly detection.
CAE is an autoencoder (AE) neural network that uses convolution layers and pooling layers to extract the hidden patterns of input features (i.e., encoding), and convolution layers and upsampling layers with bilinear interpolation to reconstruct the features from the hidden patterns (i.e., decoding). By integrating convolutional and deconvolutional operation (convolution + upsampling) in an AE structure, CAE is capable of learning the spatial structure of input features and reconstructing these features while taking into account their spatial structural patterns [5]. CAE provides a promising result to reconstruct the spectrogram by learning the spatial structural pattern of samples and can thus identify those anomalous samples by their relatively larger differences from the reconstructed spectrogram.

3.1. Features extraction

Time frequency representation of audio has been extracted by log mel spectrogram as presented in baseline system of task2. So feature extractor computes the log Mel spectrogram of input signal in frame of window size 64ms with sampling rate = 16000, No. FFT points = 1024, hop size = 512 and number of mel bands = 128, and after computing the log Mel spectrogram of each clip of signal, acoustic feature is obtained by concatenating before/after several frames (5 frames) of log-mel-filterbank outputs then fed into the training AE/CAE model.

3.2. Proposed network

We started the development of network architecture starting with baseline system which gives reasonable performance on DCASE Task2. Further we have trained our modified AE with bigger bottleneck representation and proposed CAE network architecture with dense connection at decoder part for reconstruction of output spectrogram as input spectrogram (Architecture shown in Figure 2).

Based on our experimentations on development dataset we have observed that CAE is well suited for machine classes ToyCar, ToyConveyor, Fan and pump whereas modified dense connected AE performs well for machine class Slider and Valve. In modified dense connected AE, we have changed the number of dense unit and bigger bottleneck representation shown below (Figure. 2-Left)

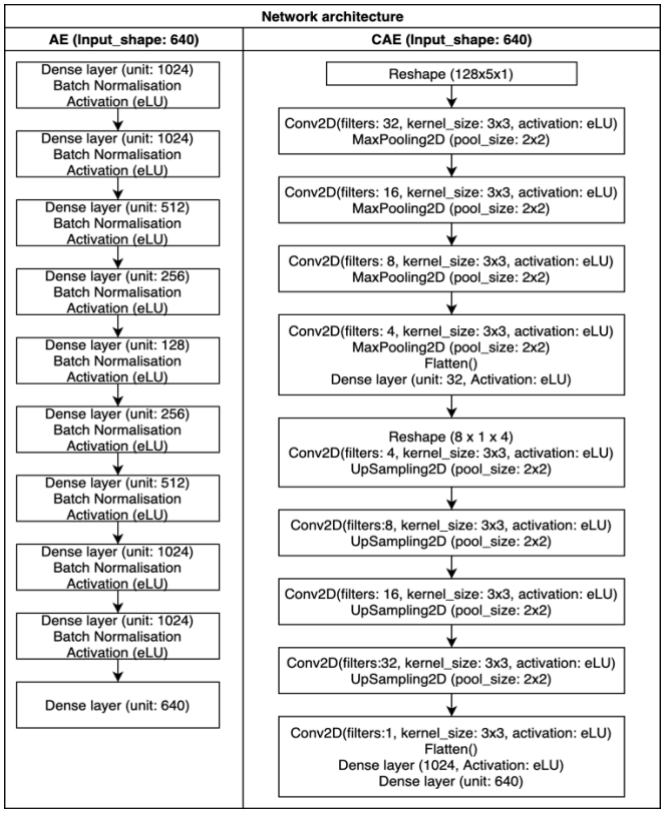


Figure 2. Detailed Network topology (AE and CAE)

In both of the network architecture we have changed non-linearity as exponential linear unit (eLU). We found that eLU lead not only to faster learning, but also to better generalization performance. It is defined as below

$$f(x) = \begin{cases} x & if\ x \geq 0 \\ \propto (\exp(x) - 1) & if\ x < 0 \end{cases} \quad (1)$$

$$f'(x) = \begin{cases} 1 & if\ x \geq 0 \\ f(x) + \propto & if\ x < 0 \end{cases} \quad (2)$$

The eLU hyper parameter $\propto$ (in our case $\propto = 1$) controls the value to which an eLU saturate for negative net. eLUs avoid the vanishing gradient problem as rectified linear units (ReLUs) and leaky ReLUs (LReLUs) do. The gradient does not vanish because the positive part of these functions is the identity, therefore their derivative is one. Thus, these activation functions are well suited for deep neural networks with many layers where a vanishing gradient impedes learning.

Further the whole system is fined tuned for Id specific of each machine class. At first the model has been trained for each machine class using AE/CAE (depending on machine specific AUC/pAUC performance) then for each machine id model has been fine-tuned using trained class weight as an embedding layer (Figure. 1) which results in improvement of AUC and pAUC of each machine types. Comparison of different system with respect to baseline shown in Table 2

3.3. Evaluation system

As described in task description system has to be evaluated with the area under the receiver operating characteristic (ROC) curve (AUC) and the partial-AUC (pAUC). The pAUC is an AUC calculated from a portion of the ROC curve over the pre-specified range of interest. At first anomaly score calculated as reconstruction error of observed sound, then fed into AUC and pAUC function. For system-A, Anomaly score is calculated as

$$A_\theta(x) = \frac{1}{DT}\sum_{t=1}^{T} \|\psi_t - F(\psi_t)\|_2^2 \qquad (3)$$

Where F function returns prediction from either AE or CAE based on class model, $\| \|_2$ is ℓ_2 norm, T is time index, ψ_t is log-mel-filterbank output and D is dimension of input to AE/CAE (640).

Whereas for system-B we have applied geometric mean of four different anomaly score calculated over ℓ_2 norm of reconstruction error. For system-B, fusion anomaly score is defined as

$$A_\theta(x) = GM\left(A_\theta^1(x), A_\theta^2(x), A_\theta^3(x), A_\theta^4(x)\right) \qquad (4)$$

Where GM is the geometric mean, $A_\theta^1(x)$ is same as $A_\theta(x)$ of system-A, $A_\theta^2(x)$ is anomaly score calculated by mean over Time index T followed by median over input dimension D to the ℓ_2 norm of reconstruction error, $A_\theta^3(x)$ is anomaly score calculated by median over Time frame T followed by median over input dimension D to the ℓ_2 norm of reconstruction error and $A_\theta^4(x)$ is anomaly score calculated by median over Time frame T followed by mean over input dimension D to the ℓ_2 norm of reconstruction error

3.4. Experiment results and conclusion

We conducted experiment on development data (including additional data) and compared our Id specific AE/CAE method for system-A and fusion system-B with baseline dense connected AE system using anomaly score calculator A_θ described in section 3.3

There is no such thing as one method works best for all classes, methods lead to different results for different classes, which differs greatly sometimes. But in general, based on our experimentations on development dataset we have observed that CAE is well suited for machine classes ToyCar, ToyConveyor, Fan and pump whereas modified dense connected AE performs well for machine class Slider and Valve. Further Id specific fine-tuning on respective class models boost the AUC and pAUC performance. Best two models system-A and system-B submitted to DCASE challenge, results shown below in system-A represent the average AUC and pAUC evaluated from Id-specific models and in system-B we combine multiple ways of calculating anomaly score together to reduce reconstruction loss for normal sounds (Eq 4.). So conclusively proposed AE and CAE provides a promising improvement in result for anomaly detection.

Class	Baseline (Average)		System A (Average)		System B (Average)	
	AUC	pAUC	AUC	pAUC	AUC	pAUC
ToyCar	78.77	67.58	84.53	70.16	85.70	72.12
ToyConveyor	72.53	60.43	74.53	60.87	74.19	59.87
Fan	65.83	52.45	69.96	56.22	70.18	56.25
Pump	72.89	59.99	75.08	63.28	74.34	63.75
Slider	84.76	66.53	87.68	67.29	85.05	66.54
Valve	66.28	50.98	80.72	54.07	76.66	52.83

Table 2. Comparison of proposed systems with baseline system

4. REFERENCES

[1] http://dcase.community/challenge2020/task-unsupervised-detection-of-anomalous-sounds

[2] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. 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. URL: https://arxiv.org/abs/2006.05822.

[3] 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.

[4] 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.

[5] Chen Lirong "A Multi-Convolutional Autoencoder Approach to Multivariate Geochemical Anomaly Recognition" Minerals (2075-163X). May2019, Vol. 9

[6] Oleg Gorokhov, Mikhail Petrovskiy, and Igor Mashechkin. Convolutional neural networks for unsupervised anomaly detection in text data. In International Conference on Intelligent Data Engineering and Automated Learning, pages 500–507. Springer, 2017.

[7] Raghavendra Chalapathy, Sanjay Chawla: Deep Learning for Anomaly Detection: A Survey. journals/corr/abs-1901-03407