Unsupervised Multi-View Reconstruction Autoencoder and Liquid Time Constant Model-Based First-Shot Anomaly Detection For Machine Condition Monitoring

Technical Report

JeongSik Kim[1], YoungHoon Jo[1], DongMin Lee[1] , JiHeon Kim[1] , GeonYong Jeong[1]
Jeongil Seo[1] , SukHwan Lee[1]

[1] Dong-A University Computer Engineering Dept,
Busan, Republic of Korea,
{2343783, pool1224, dlehdals0327, 2243790, 2344066}@donga.ac.kr
{jeongilseo, skylee}@dau.ac.kr

ABSTRACT

This technical report presents our approach for DCASE 2024 Task 2: first-shot unsupervised anomalous sound detection for machine condition monitoring. In this year task 2 focuses on first-shot challenge and also introduces real-world application that provides no additional information for some machine type. In order to tackle this challenges we have developed multi-view reconstruction based system to detect first-shot anomalous sounds more accurately and also without any additional attribute information. Our proposed system has been evaluated using DCASE 2024 Task 2 Development Dataset, and from the result we have achieved an average area under the curve detection result of 74.59%.

Index Terms— Anomalous sound detection, Multi-View Reconstruction, Liquid Time Constant

1. INTRODUCTION

Anomaly detection in industrial settings is crucial for predicting equipment failures, minimizing downtime, and ensuring worker safety. However, the scarcity and diversity of anomalous sound patterns, along with the occurrence of domain shifts and limited access to machines, present unique challenges in developing effective anomaly detection systems.

This technical report focuses on addressing these challenges through the development of an anomaly detection system designed for industrial environments. Our system is trained on normal sound data for an unsupervised learning scenario, enabling it to detect unknown anomalies not present in the training data. It also employs domain generalization techniques to handle domain shifts effectively.

A key feature of this system is implementation of multi-view method and Liquid Time Constant model. This system can be trained for novel machine types without hyperparameter tuning and can operate effectively using sound data from a limited number of machines within a machine type. Furthermore, we did not used any additional attribute while training for more general adaptability.

From the DCASE2023 Task2 [1] baseline [5] we were inspired to test 2D autoencoder. As a result our “Basic autoencoder” has achieved 63% AUC in average of both development dataset’s source and target domain. And knowing this we wanted to implant Liquid Time Constant(LTC)[2] to improve the performance of the model. And to open the further possibility of multi-modality of the future systems we have implemented Multi-View as an input method of the model.

2. DATASET

While examining the development dataset[3] we have found out first and the last part of the dose not include or very minor information of both normal and anomalous sounds. From this information we have cropped out both train / evaluation dataset’s from 1 second from start of the clip and also for the end of the clip. In order to train out system with multi-view method we have cropped our audio into 3 sections.

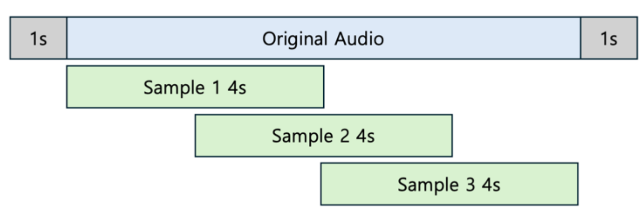

Figure 1: Data pre-processing

And we’ve also used 16k sampling rate STFT (50% hop size) and band size of 256 as feature extraction method. For more general and real-world scenario we did not trained model with any attribute.

3. MODEL ARCHITECTURE

As multi-view suggests our model uses 3 sample as an input and 2D autoencoder that consists of convolution, batch normalization, followed by relu activation. We didn't normalized data because from our test normalization had either no effect on result or worse result.

After autoencoder feature extraction has been done we concatenate reconstructed embeddings followed by flattening in to 1D. Our LTC module consists of 98304 neuron and 128 motor neuron.

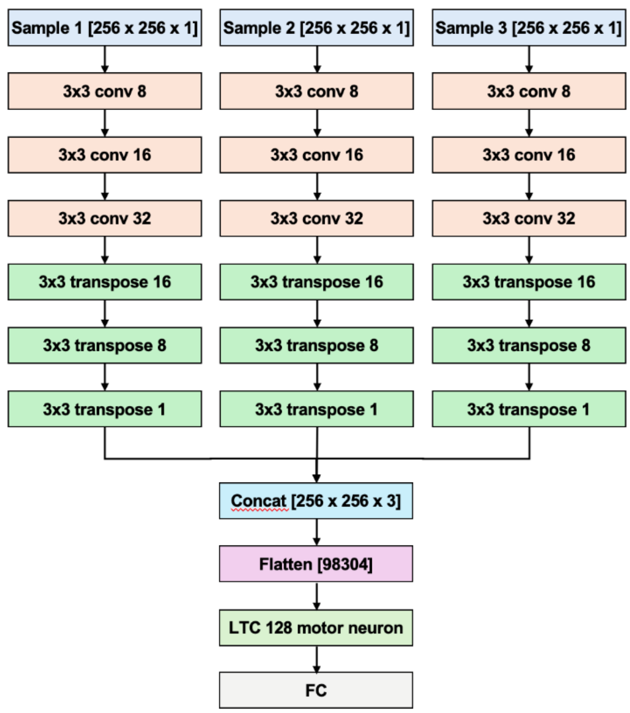

Figure 2. Model architecture

For Loss calculation and originally for binary class probability estimation we have added fully-connected layer but we don't rely on the full-connected layer's output as an result.

4. ANOMALY SCORE

For threshold calculation we first sum all the MSE loss from three autoencoder module and finally add the Cross entropy loss from LNN and set it to anomaly threshold θ in training phase. And for inferencing and evaluating we set the sum of total loss as L_{total} and use training total loss as threshold θ.

$$anomaly\ score = \begin{cases} \dfrac{\theta - L_{total}}{\theta}, & if\ \theta > L_{total} \\ \dfrac{L_{total} - \theta}{L_{total}}, & otherwise \end{cases}$$

And in the case of $\theta > L_t$ we classify the result as anomaly in evaluation phase.

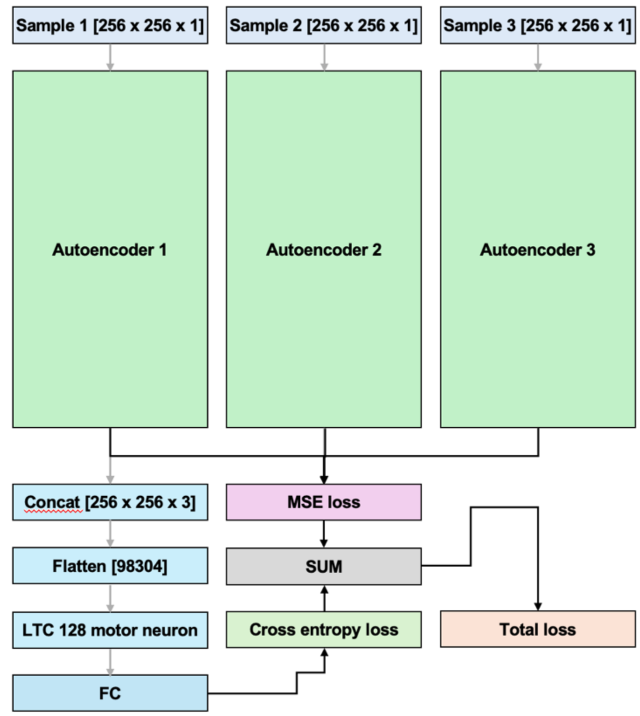

Figure 3. Anomaly score calculation of total loss

5. EXPEREMENT AND RESULTS DISCUSSION

We have trained our system on 2x Nvidia H100 at batch size of 32 and fixed learning rate of 1e-4 at 100 epoch.

Machine	Method	Baseline	Our system
ToyCar	AUC (Souce)	66.98 %	**74.29 %**
	AUC (Target)	33.75 %	**59.37 %**
	pAUC	48.77 %	**66.65 %**
ToyTrain	AUC (Souce)	**76.63 %**	56.46 %
	AUC (Target)	46.92 %	**68.32 %**
	pAUC	47.95 %	**53.43 %**
bearing	AUC (Souce)	62.01 %	**74.43 %**
	AUC (Target)	61.40 %	**85.86 %**
	pAUC	57.58 %	**70.95 %**
fan	AUC (Souce)	67.71 %	**83.31 %**
	AUC (Target)	55.24 %	**68.59 %**
	pAUC	57.53 %	**75.95 %**
gearbox	AUC (Souce)	70.40 %	**78.23 %**
	AUC (Target)	69.34 %	**68.79 %**
	pAUC	55.65 %	**72.32 %**
slider	AUC (Souce)	66.51 %	**73.77 %**
	AUC (Target)	56.01 %	**85.66 %**
	pAUC	51.77 %	**71.36 %**
valve	AUC (Souce)	51.07 %	**81.65 %**
	AUC (Target)	46.25 %	**74.40 %**
	pAUC	52.42 %	**72.48 %**

Table 1. Experiment result

Other than ToyTrain Development dataset we have achieved better result in pAUC and AUC. Some of the result shows that our model perform better AUC from target domain rather than source domain. Our Multi-view autoencoder LTC approach demonstrates significant overall improvement over the baseline, showing better performance on generalization and robustness. Test results that presented in table 1. was average result on three different random seed.

6. CONCLUSION

This tech report presents a multi-view reconstruction-based system with Liquid Time Constant (LTC) modules to address the challenges of first-shot unsupervised anomalous sound detection in machine condition monitoring. By focusing on real-world scenarios with limited data and domain shifts, the system effectively detects unknown anomalies without relying on additional attribute information or extensive hyperparameter tuning.

The evaluation results on the DCASE 2024 Task 2 Development Dataset, achieving an average AUC of 74.59%, demonstrate the effectiveness of the approach. This research contributes to advancements in anomaly detection for industrial settings, improved predictive maintenance and enhanced machine safety.

Future research directions will include exploring the integration of additional modalities, refining the model architecture for increased accuracy, and validating the system on a wider range of real-world industrial datasets.

7. REFERENCES

[1] Nishida, Tomoya and Harada, Noboru and Niizumi, Daisuke and Albertini, Davide and Sannino, Roberto and Pradolini, Simone and Augusti, Filippo and Imoto, Keisuke and Dohi, Kota and Purohit, Harsh and Endo, Takashi and Kawaguchi, Yohei, "Description and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring", arXiv e-prints: 2406.07250, 2024.

[2] Hasani, Ramin and Lechner, Mathias and Amini, Alexander and Rus, Daniela and Grosu, Radu, “Liquid Time-constant Networks”, Proceedings of the AAAI Conference on Artificial Intelligence, vol.35,pp. 7657–7666, doi:10.1609/aaai.v35i9.16936, May 2021

[3] 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," Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), pp. 1-5, November 2021.

[4] 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," Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022). November 2022.

[5] Harada, Noboru and Niizumi, Daisuke and Ohishi, Yasunori and Takeuchi, Daiki and Yasuda, Masahiro. "First-Shot Anomaly Sound Detection for Machine Condition Monitoring: A Domain Generalization Baseline," *The 31st European Signal Processing Conference (EUSIPCO)*. pp. 191-195. 2023.