Anomalous Machine Sound Detection Based on Time Domain Gammatone Spectrogram Feature and IDNN Model

Primanda Adyatma Hafiz[1], Candy Olivia Mawalim[2], Dessi Puji Lestari[1], Sakriani Sakti[3], Masashi Unoki[2]

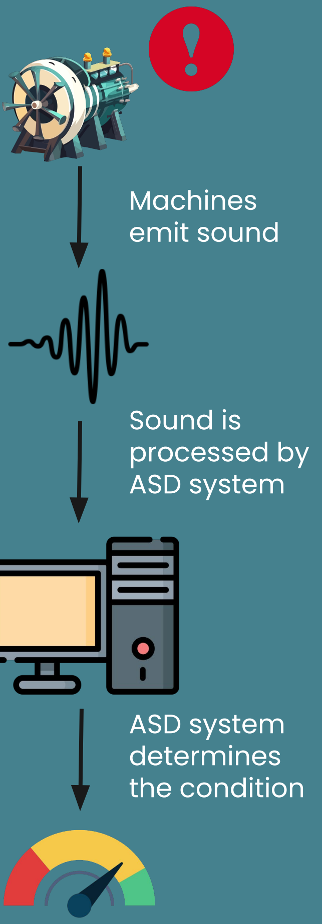
[1] Bandung Institute of Technology, [2] Japan Advanced Institute of Science and Technology, [3] Nara Institute of Technology

Outline

Introduction

Anomalous Machine Sound Detection (ASD)

- Anomalous Machine Detection is a task to **identify** whether there is a **defect in a machine**
- Compared to other parameters, sound is most commonly used due to its low cost and easiness to collect [1]
- **Unsupervised learning is commonly used** because the number of anomalous data is limited



[1] Irfan, M., Saad, N., Ibrahim, R., Asirvadam, V. S., Alwadie, A. S., & Sheikh M. A. (2017). An Assessment on the Non-Invasive Methods for Condition Monitoring of Induction Motors.

Research in ASD Features

- **Most ASD researches used log Mel spectrogram** for representing the machine's sound [1]

- Log Mel spectrogram is a spectrogram **extracted from the frequency domain**
  - Extraction requires short term fourier transform (STFT)
  - However, **STFT has some limitations** including unnecessary frequency approximations [2] and lack of resolution for rapidly changing sound [3]

- Meanwhile, **spectrogram directly calculated in the time domain has potential** to outperform frequency domain spectrogram since it does not utilize STFT

[1] Nunes, E. C. (2021). Anomalous sound detection with machine learning: A systematic review.
[2] Qi, J., Wang, D., Jiang, Y., & Liu, R. (2013). Auditory features based on gammatone filters for robust speech recognition.
[3] Daubechies, I., & Bates, B. J. (1993). Ten lectures on wavelets.

Problems on Modeling Nonstationary Sound

- Baseline ASD model of MIMII dataset **produced bad performance** especially for machine emitting **nonstationary sound** [1]
- Research showed that Interpolation Deep Neural Network **(IDNN) model successfully improved the performance of AE** particularly in machines with **nonstationary sound** characteristic [2]
- Although IDNN is commonly implemented using AE architecture, it also has potentials to be implemented with other architectures as well

[1] Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection.
[2] Suefusa, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., & Kawaguchi, Y. (2020). Anomalous sound detection based on interpolation deep neural network.

Proposed Method

Time Domain Gammatone Spectrogram (Feature)

- Closely **mimics the human auditory system** stimulus response [1]
- Can be directly calculated from the raw waveform (**time domain**)
- ERB has **good frequency resolution** especially in **lower frequency** [2]
- Time domain IIR filter is **able to provide accurate frequency magnitude response** without the trade-off between time and frequency resolution [3]

IDNN (Model)

- **Suitable** for handling spectrogram based input with **unsupervised learning technique**
- Can work well for **nonstationary sound** [4]

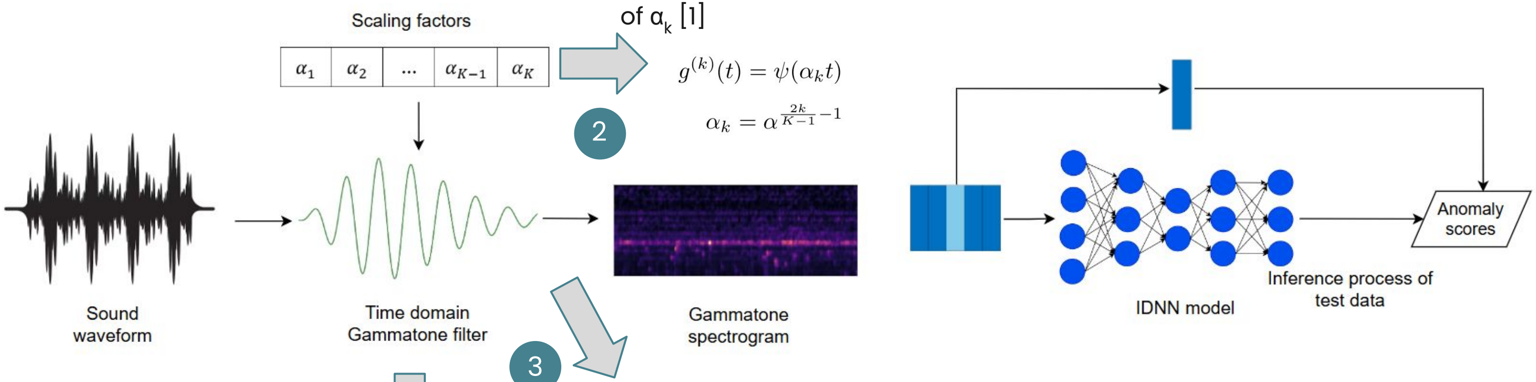
[1] Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform.
[2] Lambamo, W., Srinivasagan, R., & Jifara, W. (2022). Analyzing noise robustness of cochleogram and Mel spectrogram features in deep learning based speaker recognition.
[3] Lyons, R. G. (1997). Understanding digital signal processing, 3/E.
[4] Suefusa, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., & Kawaguchi, Y. (2020, May). Anomalous sound detection based on interpolation deep neural network.

Proposed Method - Feature Extraction

The k-th filter ($g^{(k)}$) can be calculated using **wavelet transform** by scaling the mother wavelet with a factor of α_k [1]

$$g^{(k)}(t) = \psi(\alpha_k t)$$

$$\alpha_k = \alpha^{\frac{2k}{K-1}-1}$$



Impulse response of time domain gammatone filterbank at center frequency f_c is defined as

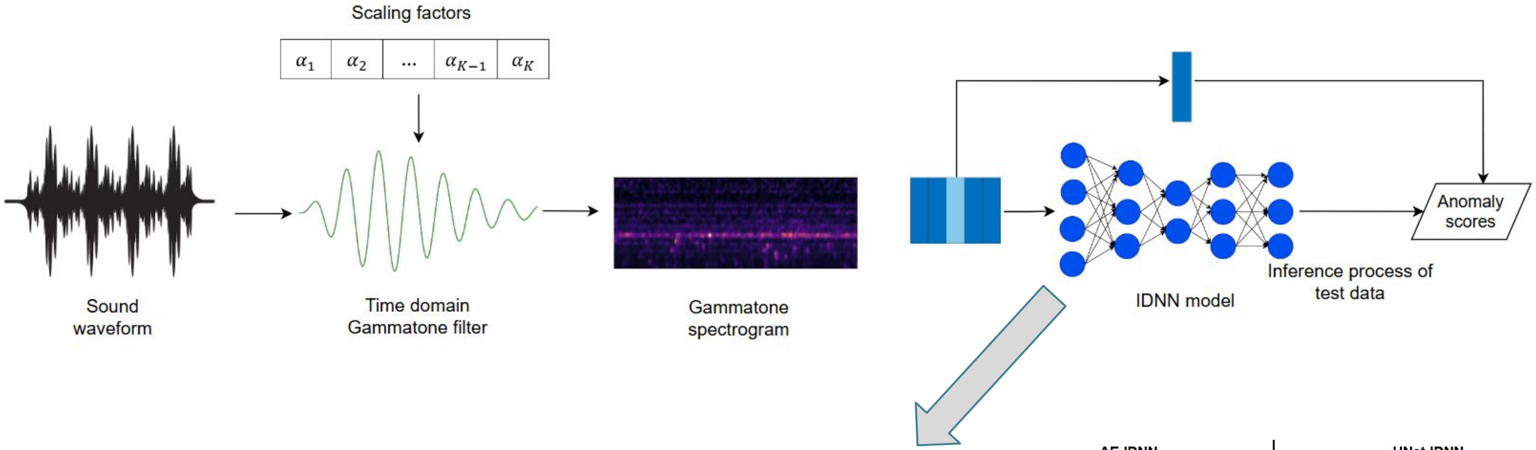
$$g(t) = t^{n-1} e^{-2\pi b \mathrm{ERB}(f_c)t} e^{j2\pi f_c t}$$

Waveform is transformed using **IIR filter** for each filter to produce

$$X_k(t) = x(t) * g^{(k)}(t)$$

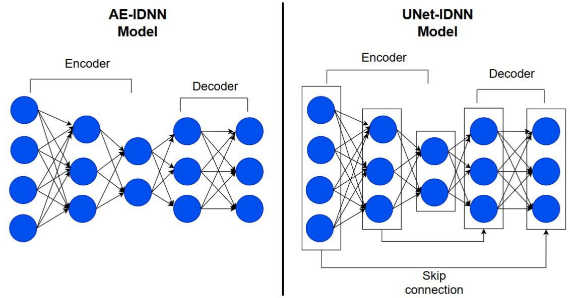
[1] Unoki, M., & Akagi, M. (1999). A method of signal extraction from noisy signal based on auditory scene analysis.

Proposed Method - Model



Feature is **modeled using 2 variation of IDNN**

- IDNN based on AE architecture
- IDNN based on UNet architecture



Experiment

Dataset

- **MIMII dataset** is chosen because it's recorded from **real machinery sound** [1]
    - Consists of 4 machine types = **slider, fan, valve, pump**
- Utilizing MIMII dataset with 6 dB SNR version
- Train and test data allocation
    - Test data = All anomalous data and the same amount of normal data
    - Train data = The remaining normal data

Machine		Normal data	Anomalous data
Slider	ID 00	1068	356
	ID 02	1068	267
	ID 04	534	178
	ID 06	534	89
Fan	ID 00	1011	407
	ID 02	1016	359
	ID 04	1033	348
	ID 06	1015	361
Valve	ID 00	991	119
	ID 02	708	120
	ID 04	1000	120
	ID 06	992	120
Pump	ID 00	1006	143
	ID 02	1005	111
	ID 04	702	100
	ID 06	1036	102
Total		14719	3400

[1] Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection.

Evaluation

- **ROC AUC Score Metrics** [1]
  - Measure the entire 2 dimensional area underneath the ROC curve

$$\mathrm{AUC} = \frac{1}{\mathrm{N}_-\mathrm{N}_+}\sum_{\mathrm{i}=1}^{\mathrm{N}_-}\sum_{\mathrm{j}=1}^{\mathrm{N}_+}\mathcal{H}(\mathcal{A}_\theta(\mathrm{x}_\mathrm{j}^+) - \mathcal{A}_\theta(\mathrm{x}_\mathrm{i}^-))$$

- **Baseline**
  - **Feature**
    - Log Mel spectrogram
  - **Model**
    - Autoencoder (AE) [2]
    - UNet [3]

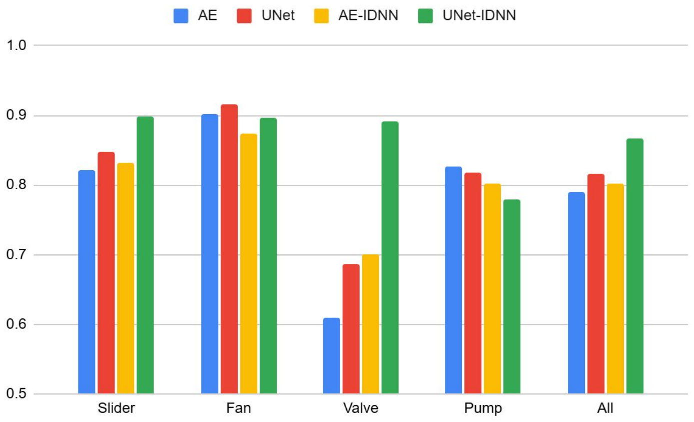
[1] Ling, C. X., Huang, J., & Zhang, H. (2003, August). AUC: a statistically consistent and more discriminating measure than accuracy.
[2] Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection.
[3] Hoang, T. V., Nguyen, H. C., & Pham, G. N. (2020). Unsupervised Detection of Anomalous Sound for Machine Condition Monitoring Using Different Auto-Encoder Methods.

Evaluation

Machine		Log Mel Spectrogram				Time Domain Gammatone Spectrogram			
		AE	UNet	AE-IDNN	UNet-IDNN	AE	UNet	AE-IDNN	UNet-IDNN
Slider	ID 00	0.991	0.989	0.958	0.992	1.000	0.997	0.997	0.997
	ID 02	0.893	0.947	0.916	0.984	0.690	0.715	0.822	0.839
	ID 04	0.744	0.799	0.786	0.875	0.906	0.940	0.984	0.954
	ID 06	0.656	0.653	0.668	0.739	0.976	0.997	1.000	0.999
	Average	0.821	0.847	0.832	0.898	0.893	0.912	0.951	0.947
Fan	ID 00	0.763	0.809	0.716	0.765	0.803	0.844	0.855	0.884
	ID 02	0.937	0.947	0.913	0.903	0.911	0.876	0.939	0.917
	ID 04	0.923	0.920	0.891	0.943	0.996	0.994	0.987	0.991
	ID 06	0.981	0.988	0.975	0.973	0.996	1.000	0.995	1.000
	Average	0.901	0.916	0.874	0.896	0.927	0.929	0.944	0.948
Valve	ID 00	0.550	0.656	0.702	0.847	0.450	0.463	0.922	0.947
	ID 02	0.622	0.682	0.704	0.887	0.929	0.957	1.000	1.000
	ID 04	0.602	0.708	0.665	0.932	0.698	0.845	0.946	0.916
	ID 06	0.666	0.696	0.732	0.900	0.523	0.639	0.854	0.836
	Average	0.610	0.686	0.701	0.892	0.650	0.729	0.931	0.925
Pump	ID 00	0.874	0.813	0.804	0.695	0.839	0.827	0.804	0.813
	ID 02	0.503	0.560	0.498	0.531	0.704	0.681	0.715	0.725
	ID 04	0.997	1.000	0.999	1.000	0.960	0.997	0.997	0.998
	ID 06	0.933	0.894	0.907	0.890	0.974	0.971	0.946	0.979
	Average	0.827	0.817	0.802	0.779	0.869	0.869	0.866	0.879
All	Average	0.790	0.816	0.802	0.866	0.835	0.859	0.923	0.925

Detailed evaluation results for each feature, model, and machine are available in **Table II**

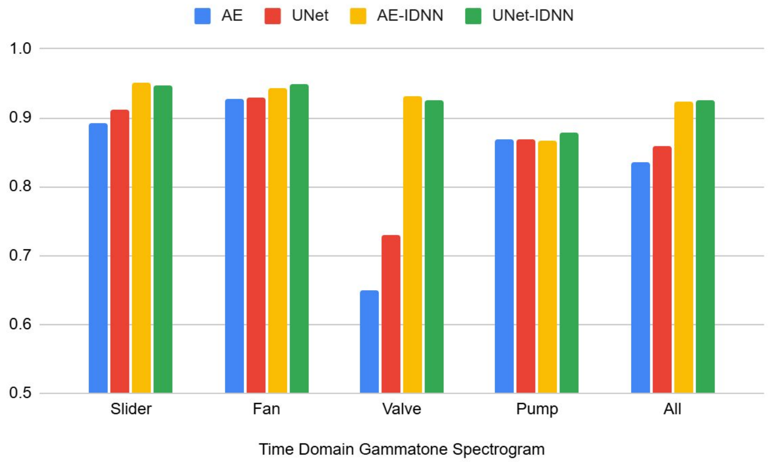
Evaluation – Log Mel Spectrogram



- **Overall statistics**
  - Average AUC (UNet-IDNN) = **0.866**

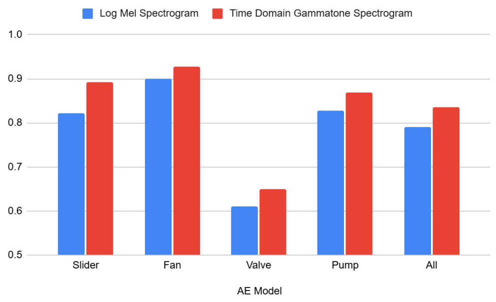
- **Insights**
  - Valve machines gave poor results for some models
  - IDNN based models significantly improved the performance on valve machines

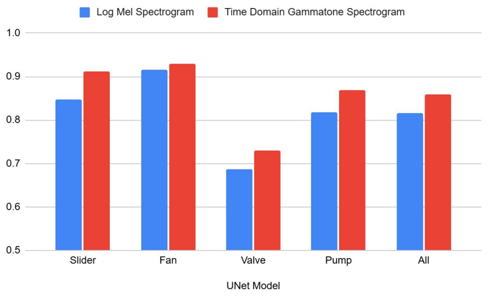
Evaluation - Time Domain Gammatone Spectrogram

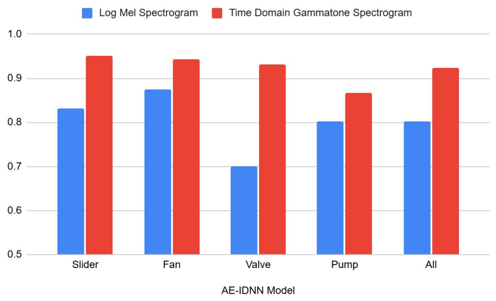


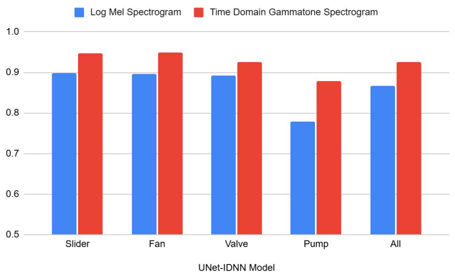
- **Overall statistics**
  - Average AUC (UNet-IDNN) = **0.925**
- **Insights**
  - Valve machines gave poor results for baseline models
  - Generally, IDNN based models gave improvement for all machine types
  - IDNN based models gave significant improvement in valve machines

Evaluation - Time Domain Gammatone Spectrogram vs Log Mel Spectrogram



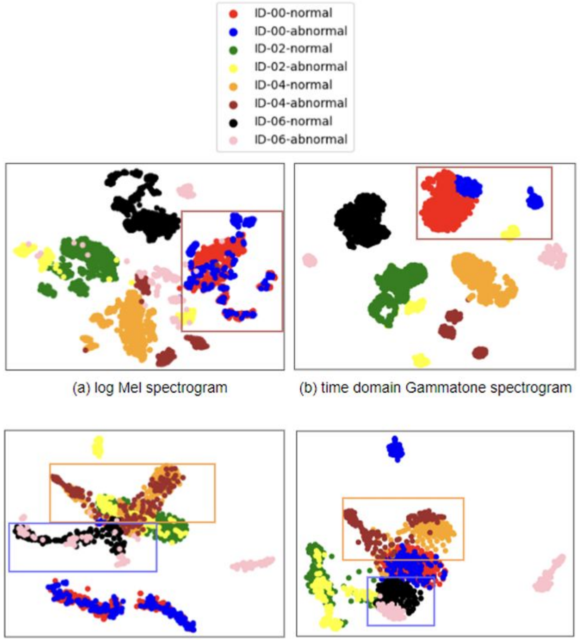






- Time Domain Gammatone spectrogram outperformed log Mel spectrogram in every scenario
- In each feature's best model, **time domain Gammatone spectrogram outperforms** log Mel spectrogram **by 5.9 pp**

t-SNE Distribution Analysis



(a) log Mel spectrogram

(b) time domain Gammatone spectrogram

(c) bottleneck features of log Mel spectrogram

(d) bottleneck features of time domain Gammatone spectrogram

Time domain Gammatone spectrogram exhibited better separability compared to log Mel spectrogram in both raw features and bottleneck features

Conclusion

Conclusion

- **Time domain Gammatone spectrogram demonstrated a significant improvement of 5.9 pp** over the baseline feature, the log Mel spectrogram

- Time domain Gammatone spectrogram **performed better** overall than the log Mel spectrogram **in all comparative models**

- **Utilizing the IDNN model further enhanced the effectiveness** of the time domain Gammatone spectrogram feature, particularly IDNN with UNet architecture (UNet-IDNN)

Thank You