UNSUPERVISED ANOMALOUS SOUND DETECTION VIA AUTOENCODER APPROACH

Technical Report

Shuo Wang, Zihao Li, Yuxuan Zhang,
Kejian Guo, Shijin Chen, Yan Pang

University of Chinese Academy of Sciences, Beijing, China
zhangyuxuan@mail.ioa.ac.cn

ABSTRACT

In the industrial field, the anomaly detection of mechanical systems has played an important role. This technical report uses four modified autoencoders (AEs) to detect abnormal conditions of different machines in DCASE2021 Task 2. AE has been widely used in image reconstruction due to its excellent generalization ability. The reconstruction error can be used to evaluate the abnormal value of the machine condition when the development set only provide the normal mechanical sound signals. The performance of the anomaly detection system is evaluated by the area under the receiver operating characteristic curve (AUC) and partial-AUC (pAUC) scores. Finally, the experimental results show that the presented models can improve AUC and pAUC compared to the baseline system.

Index Terms— Anomaly detection, Autoencoder, AUC, pAUC

1. INTRODUCTION

Anomaly detection has attracted much attention in signal processing, especially for mechanical condition monitoring [1]. The traditional algorithm of mechanical condition monitoring is to study the statistical characteristics of vibration signal, such as spectral kurtosis [2], frequency center [3], short-time Fourier spectrum [4, 5], etc. However, the generalization ability of these traditional algorithms is poor due to the simplification of the signal model.

Nowadays, the deep neural network has been widely applied to unsupervised anomaly detection, and it can be divided into two categories. One is the autoencoder (AE)-based model. Dcase2020 Task2 [6] and Dcase2021 Task2 [7] both presented a simple AE architecture as the baseline system. For AE-based architectures, the reconstruction error can be viewed as the abnormal value when the training set only contains normal data. Moreover, some improved AE, such as Heteroskedastic Variational AE (HVAE) [8] and Conformer-based AE [9], have also been proposed to improve the performance of anomalous sound detection. The second is the classification-based model, which associates anomaly detection with the recognition of machine ID. The machine may be abnormal when the result predicted by the deep neural network is different from the reference ID [10, 11, 12].

The Unsupervised Anomalous Sound Detection for Machine Condition Monitoring under Domain Shifted Conditions in the DCASE2021 task2 has provided the dataset which contains normal and abnormal sound signals. The task has two main challenges [13, 7, 14]:

- The task only provides normal sound clips as training data, while the trained model needs to detect abnormal sounds.
- The training dataset and evaluation dataset have different acoustic characteristics (i.e., domain shift).

In this report, we utilize four modified AE models to monitor abnormal mechanical conditions.

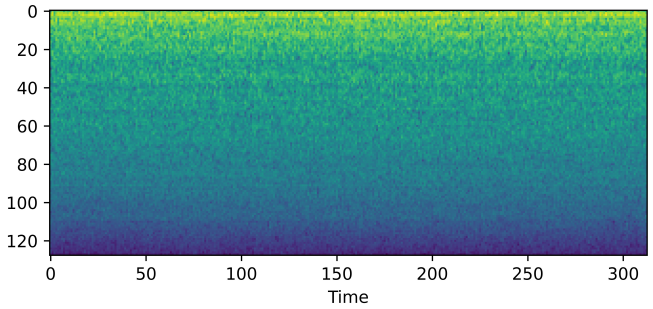

Figure 1: Log-Mel spectrogram of an audio file from the fan

2. PROPOSED METHOD

2.1. Feature extraction

We first extract the Log-Mel spectrogram from the 10-second acoustic signal, where the hop length is 32 ms and the window length is 64 ms. Figure 1 shows the Log-Mel spectrogram of an audio file from the fan. Then the number of frames of the Log-Mel spectrogram is initialized as $P = 5$, that is, $\mathbf{t} = [t_{\mathbf{i+1}}, t_{\mathbf{i+2}}, \ldots, t_{\mathbf{i+5}}]$ can be viewed as a chunk in the spectrogram. Since we set the mel band as $F = 128$, the 5-frames Log-Mel spectrogram is concatenated to generate a vector $x \in \mathbb{R}^{640}$. Finally, the feature vector of $\boldsymbol{t}$ is used for modified AE and VAE, while the vector of $[t_{\mathbf{i+1}}, t_{\mathbf{i+2}}, t_{\mathbf{i+4}}, t_{\mathbf{i+5}}]$ is used for IAE and IVAE.

2.2. Baseline system

Baseline system utilizes a simple autoencoder, and the network architecture is shown in Figure 2. Each hidden layer of the encoder and decoder uses only 128 units, and the latent

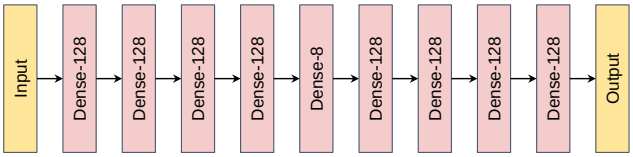

Figure 2: Baseline system

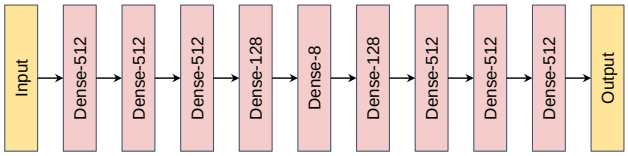

Figure 3: Dense autoencoder architecture

vector of the bottleneck layer utilizes 8 units. In addition, the input and output layers use 640 units. Finally, the reconstruction error of the observed sound signal is viewed as the anomaly score

$$A_\theta(X) = \frac{1}{DT}\sum_{t=1}^{T}\|\psi_t - r_\theta(\psi_t)\|_2^2, \quad (1)$$

where $D = P \times F$, T is the number of time frames, ψ_t is the original Log-Mel spectrogram and r_θ is the reconstructed spectrogram. Because the training set only contains normal sounds of machines, the trained autoencoder cannot perfect reconstruct abnormal mechanical features. Therefore, it is reasonable to choose the reconstruction error as the anomaly score.

2.3. Proposed models

In this challenge, four AE based models and submission systems based on these are summarized as follow:

- Modified AE
  For the AE model, the latent vector is used to represent the feature of the input Log-Mel map. We consider increasing the number of units in hidden layer to improve the performance. Therefore, some hidden layers of Dense128 are substitute by Dense512. Moreover, each dense layer is followed by batch normalization and ReLU activation. The architecture of the modified AE is shown in Figure 3.
- Variational AE
  The Variational AE (VAE) was first proposed by Diederik P.Kingma and Max Welling [15]. Different from AE, VAE suppose that the latent vector follow the Gaussian distribution, and the output map can be decoded from the sampled latent vector. Therefore, the mean and standard deviation hidden layers are added to the network, and the sampled latent vectors are generated through these two layers. The architecture of VAE is shown in Figure 4.
- Interploation AE and Interploation VAE

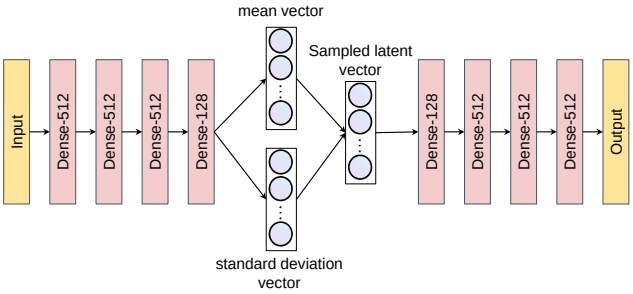

Figure 4: Variational autoencoder architecture

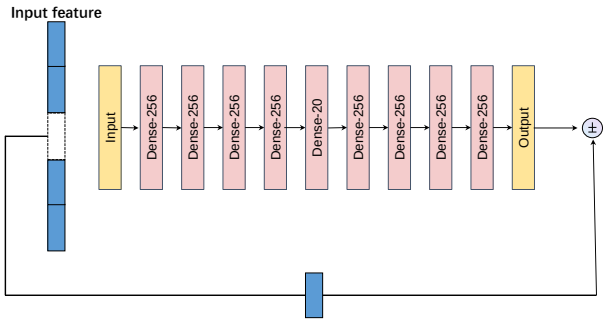

Figure 5: Interploation dense autoencoder architecture

The Interploation AE (IAE) and Interploation DVAE (IVAE) utilize multiple frames of a spectrogram whose center frame is removed as an input, and predicts an interpolation of the removed frame as an output. More details about IAE and IVAE can be found in [16], and the architectures are shown in Figure 5 and Figure 6. Note that the dimensions of input and output feature vector are 640 in VAE and AE. However, the dimension of input feature vector is 512, and the dimension of output feature is 128 in the IAE and IVAE. In addition, the number of units is 20 in the standard deviation layer, mean layer, sampled latent layer for the IVAE and VAE.

- Based on the above models and their combinations, we submitted the following four systems:
  - system 1: Modified AE.
  - system 2: Variational AE.
  - system 3: AE+VAE: Normalize the anomalous scores of the AE and VAE models, respectively, and take the average with equal weights as the output of the system.
  - system 4: IAE+IVAE: Normalize the anomalous scores of the IAE and IVAE models respectively, and take the average with equal weights as the output of the system.

The results of the two individual models and the two ensemble models are summarized in Table 1.

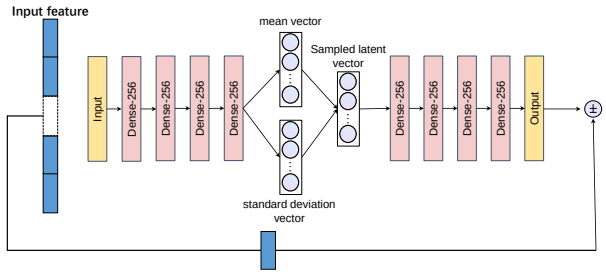

Figure 6: Interploation dense autoencoder architecture

2.4. Evaluation metrics

To evaluate the performance of anomaly detection system, the AUC and the pAUC are utilized as the evaluation metrics, which can be expressed as

$$\begin{aligned} \mathrm{AUC}_{m,n,d} &= \frac{1}{N_- N_+} \sum_{i=1}^{N_-} \sum_{j=1}^{N_+} \mathcal{H}\left(\mathcal{A}_\theta\left(x_j^+\right) - \mathcal{A}_\theta\left(x_i^-\right)\right) \\ \mathrm{pAUC}_{m,n,d} &= \frac{1}{\lfloor pN_- \rfloor N_+} \sum_{i=1}^{\lfloor pN_- \rfloor} \sum_{j=1}^{N_+} \mathcal{H}\left(\mathcal{A}_\theta\left(x_j^+\right) - \mathcal{A}_\theta\left(x_i^-\right)\right) \end{aligned} \tag{2}$$

where m represents the index of a machine type, n represents the index of a section, $d = \{$ source, target $\}$ represents a domain, $\lfloor \cdot \rfloor$ is the flooring function, and $\mathcal{H}(x)$ returns 1 when $x > 0$ and 0 otherwise. Furthermore, $\{x_i^-\}_{i=1}^{N_-}$ and $\{x_j^+\}_{j=1}^{N_+}$ are normal and anomalous test clips. N_- and N_+ are the number of normal and anomalous test clips in the domain d in the section n in the machine type m, respectively.

3. EXPERIMENT

3.1. Dataset description

The dataset of DCASE2021 Task 2 consists of normal/abnormal data of seven types of machinery, namely fan, gearbox, pump, slide rail, toycar, toytrain and valve. In addition, the training set only contains normal sounds, which belong to the source domain or target domain. However, the amount of data in the source domain is much larger than that in the target domain in the training set, while the data in evaluation set is balanced.

3.2. Training hyperparameter

In this challenge, we train models separately for different machines, and Adam is selected as the model optimizer. To calculate the Mel spectrogram, the frame size is set to 1024, the hop size is set to 512, and the number of Mel filter banks is set to 128. In addition, we train the network for 80 epochs of every mechaine, and the learning rate is initialized as 0.001.

3.3. Experimental results

The performance comparison of these modified AE models is shown in Table 1. It is obvious that the performance of modified AE is the best among these models for most mechanical types, while other models outperform the baseline model only on specific machines. Therefore, we make the predictions of the modified AE model on the evaluation as a submission. The other three submissions are the predictions of VAE, AE+VAE and IAE+IVAE, which has been mentioned in section 2.3.

4. CONCLUSION

The purpose of this report is to analyze the performance of AE model on mechanical anomalous sound detection. To this end, we have utilized four AE models. The results of the experiment show that modified AE models can outperform the baseline system.

5. REFERENCES

[1] Y. Wang, J. Xiang, R. Markert, and M. Liang, "Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications," Mechanical Systems and Signal Processing, vol. 66, pp. 679–698, 2016.

[2] M. A. Haile and B. Dykas, "Blind source separation for vibration-based diagnostics of rotorcraft bearings," Journal of Vibration and Control, vol. 22, no. 18, pp. 3807–3820, 2016.

[3] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," Expert Systems with Applications, vol. 37, no. 2, pp. 1419–1430, 2010.

[4] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," Mechanical systems and signal processing, vol. 70, pp. 1–35, 2016.

[5] J. Zhang, H. Gao, Q. Liu, F. Farzadpour, C. Grebe, and Y. Tian, "Adaptive parameter blind source separation technique for wheel condition monitoring," Mechanical Systems and Signal Processing, vol. 90, pp. 208–221, 2017.

[6] 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 Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), November 2020, pp. 81–85. [Online]. Available: http://dcase.community/documents/workshop2020/proceedings/DCASE2020Workshop_Koizumi_3.pdf

[7] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on DCASE 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions," In arXiv e-prints: 2106.04492, 1–5, 2021.

[8] P. Daniluk, M. Gozdziewski, S. Kapka, and M. Kosmider, "Ensemble of auto-encoder based and wavenet

Table 1: The results of different AE models on the development set

Model	ToyCar		ToyTrain		Fan		Gearbox		Pump		Slider		Valve	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
Baseline	0.6249	0.5236	0.6171	0.5381	0.6324	0.5338	0.6597	0.5276	0.6192	0.5441	0.6674	0.5594	0.5341	0.5054
AE	**0.6594**	**0.5343**	**0.6726**	**0.5519**	0.6260	**0.5342**	**0.6661**	0.5283	**0.6218**	**0.5476**	**0.6678**	**0.5618**	**0.5446**	0.5051
VAE	0.6109	0.5188	0.6165	0.5369	0.6159	0.5169	0.6440	**0.5337**	0.6015	0.5340	0.6423	0.5136	0.5243	**0.5066**
AE+VAE	0.6320	0.5231	0.6379	0.5426	**0.6271**	0.5261	0.6536	0.5314	0.6114	0.5397	0.6545	0.5521	0.5295	**0.5066**
IAE	0.6029	0.5167	0.5860	0.5221	0.5944	0.5321	0.6366	0.5228	0.5657	0.5407	0.6580	0.5530	0.5317	0.5026
IVAE	0.5804	0.5220	0.6568	0.5395	0.5974	0.5135	0.6513	**0.5337**	0.5880	0.5308	0.6073	0.5276	0.5338	0.5014
IAE+IVAE	0.5949	0.5283	0.6201	0.5308	0.6041	0.5216	0.6575	0.5306	0.5884	0.5364	0.6253	0.5331	0.4938	0.5012

like systems for unsupervised anomaly detection," Tech. report in DCASE2020 Challenge Task, Tech. Rep., 2020.

[9] T. Hayashi, T. Yoshimura, and Y. Adachi, "Conformer-based id-aware autoencoder for unsupervised anomalous sound detection," Tech. Rep., DCASE2020 Challenge, Tech. Rep., 2020.

[10] R. Giri, S. V. Tenneti, F. Cheng, K. Helwani, U. Isik, and A. Krishnaswamy, "Self-supervised classification for detecting anomalous sounds," in Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), 2020, pp. 46–50.

[11] P. Primus, "Reframing unsupervised machine condition monitoring as a supervised classification task with outlier-exposed classifiers," DCASE2020 Challenge, Tech. Rep, Tech. Rep., 2020.

[12] T. Inoue, P. Vinayavekhin, S. Morikuni, S. Wang, T. H. Trong, D. Wood, M. Tatsubori, and R. Tachibana, "Detection of anomalous sounds for machine condition monitoring using classification confidence," Tech. report in DCASE2020 Challenge Task, Tech. Rep., 2020.

[13] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, "Mimii due: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions," In arXiv e-prints: 2006.05822, 1–4, 2021.

[14] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions," arXiv preprint arXiv:2106.02369, 2021.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.

[16] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Anomalous sound detection based on interpolation deep neural network," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 271–275.