UNSUPERVISED ANOMALY SOUND DETECTION BASED ON GAMMAVAE

Technical Report

Shun Huang

XinJiang University
School of Computer Science and Technology
Urumqi 830017,China
huangswt@stu.xju.edu.cn

Yunxiang Zhang

XinJiang University
School of Computer Science and Technology
Urumqi 830017,China
zhangyunxiang119@foxmail.com

ABSTRACT

This technical report describes the submission to Task 2 of the DCASE 2024 challenge. Assuming the data follows a Gamma distribution, we employed a Gamma Variational Autoencoder (GammaVAE) for modeling, and utilized Mean Squared Error (MSE) scores and Mahalanobis distance for evaluation. Experimental results revealed that our system outperformed the baseline in the target domain on certain machines.

Index Terms— Anomalous sound detection, Gamma Variational Autoencoder, KL loss

1. INTRODUCTION

In this task [1], the purpose of anomaly sound detection is to train a system that can detect whether a machine is operating normally or abnormally based on the sound it produces. However, one issue that arises is that the sounds produced by different machines seem to vary across different seasons. If only sound data from the source domain is used for training, the system may mistakenly classify situations in the target domain as abnormal, leading to false alarms. For autoencoders, their vectors in the latent space are fixed. Considering this, the system may not generalize well to certain target domain situations. Additionally, based on the assumption that the data follows a gamma distribution, we adopt a GammaVAE [2] to train an anomaly sound detection system. The model is optimized by maximizing the evidence lower bound (ELBO), which consists of two components: reconstruction error and KL divergence, to ensure that the latent representation follows a gamma distribution. This enables capturing the complex features of normal sound data, particularly in cases of asymmetric and skewed distributions.

2. PROPOSED METHOD

The GammaVAE model is composed of three main components: an encoder, a latent variable layer, and a decoder,The detailed configuration is shown in the table 1 . The encoder is designed to map high-dimensional input data to the latent space using multiple layers of linear units, batch normalization layers, and ReLU activation functions. The latent variable layer consists of two parallel linear layers that compute the mean and variance of the latent variables. The outputs of these layers are processed through softmax activation functions to ensure non-negative outputs. The decoder mirrors the structure of the encoder and is tasked with reconstructing the variables from the latent space back into the original data space. This architecture allows the model to effectively learn and reconstruct data while ensuring that the latent representations follow a gamma distribution. This is particularly useful for capturing complex features in asymmetric and skewed data distributions.

2.1. Calculation of loss function

To train the GammaVAE model, we designed a loss function that combines reconstruction error and KL divergence. The reconstruction error measures [3, 4, 5] the difference between the decoder's output and the original input, while the KL divergence quantifies the difference between the distribution of latent variables and a prior distribution. The reconstruction error is measured using Mean Squared Error (MSE) and is defined as follows:

$$\mathrm{MSE}(\mathrm{recons}, \mathrm{input}) = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{recons}_i - \mathrm{input}_i)^2 \tag{1}$$

where (recons) denotes the output of the decoder and (input) represents the original input.

2.2. KL divergence

The Kullback-Leibler (KL) divergence is used to measure the difference between the distribution of latent variables Q and a prior distribution P. For Gamma distributions, the KL divergence is calculated using the following formula:

$$\begin{aligned}\mathrm{KL}(Q\|P) = \sum_{i=1}^{n}\Bigg(&\frac{c_i d_i}{a_i} + b_i\log(a_i) - \log(c_i d_i)\\ &- b_i + (b_i - 1)(\psi(d_i) + \log(c_i))\Bigg)\end{aligned} \tag{2}$$

where a_i and b_i are the shape and scale parameters of the prior Gamma distribution, and c_i and d_i are the shape and scale parameters of the latent Gamma distribution.

Table 1: GammaVAE Model Configuration

Component	**Layer Type**	**Output Size**	**Activation**
	Encoder		
Layer 1	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 2	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 3	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 4	Linear BatchNorm1d ReLU	128 128 -	- - -
	Latent Variable Layer		
Mean	Linear (fc_mu)	8	Softmax
Variance	Linear (fc_var)	8	Softmax
	Decoder		
Layer 1	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 2	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 3	Linear BatchNorm1d ReLU	128 128 -	- - -
Layer 4	Linear BatchNorm1d ReLU	128 128 -	- - -
Output Layer	Linear	Input Dim (640)	-

Table 2: Performance Comparison (MAHALA)

	Method	Baseline	Ours
ToyCar	AUC-s	63.64%	58.70%
	AUC-t	37.36%	41.08%
	pAUC	51.00%	50.57%
ToyTrain	AUC-s	64.63%	60.68%
	AUC-t	40.78%	39.28%
	pAUC	48.05%	47.68%
Bearing	AUC-s	55.26%	53.50%
	AUC-t	52.30%	56.14%
	pAUC	58.84%	59.31%
Fan	AUC-s	79.46%	78.90%
	AUC-t	42.64%	44.28%
	pAUC	53.10%	52.26%
Gearbox	AUC-s	80.40%	78.28%
	AUC-t	74.22%	68.30%
	pAUC	55.36%	51.94%
Slider	AUC-s	74.42%	74.76%
	AUC-t	67.66%	65.44%
	pAUC	48.68%	48.21%
Valve	AUC-s	54.56%	54.82%
	AUC-t	51.32%	50.28%
	pAUC	51.36%	51.68%

3. EXPERIMENTS

In the experimental process, the hyperparameters, random seed, and training framework used were handled in a baseline manner [1, 6, 7] .

Performance varies across different mechanical components. Overall, our approach outperforms the Baseline in certain cases, such as ToyCar, Bearing, and Fan, particularly in terms of AUC-t. However, in other cases, like Gearbox and ToyTrain, the Baseline shows slightly better performance. In summary, our method exhibits superior anomaly detection capabilities in specific scenarios, but further optimization is needed to surpass the Baseline across all scenarios.

4. REFERENCES

[1] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, “Description and discussion on DCASE 2024 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring,” *In arXiv e-prints: 2406.07250*, 2024.

[2] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, “Understanding disentangling in beta-vae,” *arXiv preprint arXiv:1804.03599*, 2018.

[3] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[5] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, “First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline,” in *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 191–195.

[6] 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,” in *Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, Barcelona, Spain, November 2021, pp. 1–5.

[7] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, “MIMII DG: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task,” in *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, November 2022.