DETECTION OF ANOMALOUS SOUNDS FOR MACHINE CONDITION MONITORING USING CLASSIFICATION CONFIDENCE

Technical Report

Tadanobu Inoue[1,†] *Phongtharin Vinayavekhin*[1,†], *Shu Morikuni*[1], *Shiqiang Wang*[2], *Tuan Hoang Trong*[2], *David Wood*[2], *Michiaki Tatsubori*[1], *Ryuki Tachibana*[1]

[1]IBM Research – Tokyo, Japan
[2]IBM Research, Yorktown Heights, NY, USA

corresponding author: pvmilk@jp.ibm.com

ABSTRACT

For the DCASE 2020 Task 2 challenge, we propose unsupervised anomalous sound detection methods using an ensemble of two classifiers. Both classifiers are trained with either known or generated properties of normal sounds as labels: (1) one is a model to classify sounds into machine types and IDs; (2) the other is a model to classify transformed sounds into the data augmentation types, where we augment the normal sound by using sound transformation techniques such as pitch shifting and use data augmentation types as labels. For both classifiers, we use the classification confidence as the normality score of the input sample at run-time. We ensemble these approaches using probability aggregation of their anomaly scores. As a result, the experimental results show superior performance to the baseline provided by the DCASE organizer.

Index Terms— Anomaly Detection, Classifier-based Confident Score, Feature Learning

1. INTRODUCTION

Anomaly detection is the task of finding unusual samples in data. In the unsupervised anomaly detection setting, we have a training dataset which consists of only "normal" data, i.e., anomalous samples are not known a priori. Anomaly detection algorithms can be used in a wide range of applications such as quality inspection of products, equipment maintenance, network intrusion detection, fraud detection, etc. In this work, we focus on detecting anomalous sounds to monitor machine condition.

Anomaly detection is an important topic in machine learning. There are many approaches to solve this problem in the literature [1, 2]. Deep learning is used in this work due to the type (i.e., sound) and amount of data. Approaches in deep learning for unsupervised anomaly detection can be broadly categorized into reconstruction based methods and feature learning based methods. In reconstruction based methods, the model is trained to learn the distribution of the normal samples and anomalies can be detected by analyzing the reconstruction errors as an anomaly score [3, 4, 5]. The reconstruction error is usually higher for anomalies as the model is only trained to reconstruct the normal samples. In feature learning based methods, a feature extraction model is trained to map normal data into small region in the feature space. Anomalies can be detected from analyzing the distance in the feature space [6]. As a variation of feature learning based methods, one can utilize the classifier confidence, specifically a maximum softmax probability (MSP). In this method, an anomalous sample is considered to be outside of the distributions that the classifier had learned, and generally has lower MSP [7].

We develop two types of unsupervised anomalous sound detection approaches using classification confidence. The first approach uses normal sounds from all machine types and IDs to train a classifier that predicts the machine type and ID of each normal sound. The other type uses data augmentation techniques to generate pseudo classes from normal sounds, in order to learn a classifier that predicts which data augmentation technique has been used for each sound sample, where the sound sample is generated from normal sounds. In the inference stage, a test sound clip, where it is unknown whether it is normal or abnormal, is sent to each of these two classifiers, to produce an anomaly score that is related to the softmax probability predicted by the classifier. Finally, we ensemble these anomaly scores using probability aggregation.

This report describes our two types of classification-based approaches in Section 2. Section 3 describes techniques to improve the anomaly detection performance including the ensemble methods. The experimental results on the development dataset of DCASE 2020 Challenge Task 2 are shown in Section 4. Finally, our conclusions are provided in Section 5.

2. FEATURE LEARNING FOR ANOMALY DETECTION

Classification confidence-based approaches require classes to train a classifier discriminating a target machine class from others. While external datasets allowed to use could have been leveraged, we explore what we can do with the given dataset of the competition.

2.1. Machine Types and IDs as Class Labels

The dataset for DCASE 2020 Task 2 challenge [8, 9, 10] has six machine types: ToyCar, ToyConveyor, fan, pump, slider and valve. Each machine type has six (ToyConveyor) or seven (other machine types) machine IDs. We use tuples of machine type and ID as class labels. We train a neural network using the training data in the development dataset, which contains normal sounds, to classify a subset of these classes. The last layer of the model is softmax that outputs softmax probabilities. In the inference phase, we classify a test sound using the trained model. Since we know the class of each test sound, i.e. machine type/ID, we calculate the anomaly score

† equal contribution; alphabetical ordering

$s_1(x)$ using the softmax probability of that particular class:

$$s_1(x) = 1 - y_j(x) \tag{1}$$

where $y(x)$ is the trained model's output and j is the target machine type/ID class index. Alternatively, MSP can also be used.

2.2. Sound Data Augmentation Types as Pseudo Labels

Golan *et al.* [11] proposed an anomaly detection approach using geometric transformations on image data, where a classifier is trained to infer geometric transformations of images. The geometric transformations consist of combinations of flip, xy-translations, and rotation. During inference phase, anomalies are detected by combining the softmax values of each geometric transformation.

When we naively apply these geometric transformations to the spectrogram of sound data as images, the result of anomaly detection was not good. Instead, we apply k types of sound data augmentation, which includes combinations of pitch shift and time stretch, to create pseudo labels to build a classifier.

Then we train a model to classify sound segments into k classes. During inference phase, we apply k types of sound data augmentation to the target sound clip. We divide the augmented k sound clips into multiple sound segments and then infer them using the trained model. We get the average of the softmax probabilities over the sound segments for the target sound clip to get the clip-wise values. Then we calculate the anomaly score $s_2(x)$ using the clip-wise softmax probabilities corresponding to the actual data augmentation type:

$$s_2(x) = 1 - \frac{1}{k}\sum_{j=0}^{k-1}[y(T_j(x))]_j \tag{2}$$

where $y(x)$ is the clip-wise softmax probabilities, $T_j(x)$ is the jth data augmentation type, and k is the number of data augmentation types. Anomalies can be detected based on this anomaly score.

3. TECHNIQUES FOR PERFORMANCE IMPROVEMENT

We apply the following techniques to improve the overall performance of anomaly detection.

3.1. Sound Segments

Our proposed method can take either the whole sound clip or segments of each sound clip as classifier inputs. However, we found that segmenting a sound clip into multiple sound segments improves the anomaly detection performance. This could be because the anomaly usually occurs only within a small part of the sound clip. Another explanation could be that dividing a sound clip into multiple segments increases the amount of training data and makes the classifier more sensitive to the anomaly.

Fig. 1 shows how we divide a sound clip into multiple sound segments with a specific hop size. We also segment each test sound into small segments, individually calculate their anomaly score, and aggregate them using averaging.

3.2. Center Loss

Perera and Patel [12] suggested that to learn a good feature for one-class classification, two types of losses are required: a descriptiveness loss and a compactness loss. The classifier represents the descriptive part of the feature. We explain how we compress the features in this subsection.

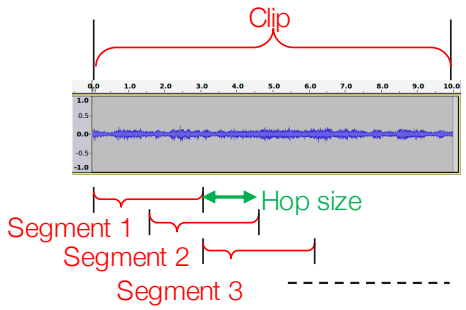

Figure 1: Sound clip is divided into multiple sound segments.

Ruff *et al.* [6] proposed deep support vector data description (Deep SVDD) approach for anomaly detection. They train a deep learning model to map normal input data to a minimized volume hypersphere in the feature space, in order to maximize the difference between normal and anomaly inputs in the feature space. Wen *et al.* [13] proposed center loss for deep face recognition application, in order to enhance the discriminative power of the deeply learned features:

$$L = L_s + \lambda L_c \tag{3}$$

$$= \sum_{i=1}^{m} log\frac{e^{{W_{y_i}}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{{W_j}^T x_i + b_j}} + \frac{\lambda}{2}\sum_{i=1}^{m} \parallel x_i - c_{y_i} \parallel_2^2 \tag{4}$$

where L_s is the softmax loss and L_c is the proposed center loss.

We use the center loss for the classifier training to map normal input data to a minimized volume hypersphere in the feature space.

3.3. Ensemble Methods

Ensemble methods focus on the idea of combining different results of dissimilar sub-models to enhance the overall performance. In ensemble methods, there exist challenges such as interpretability and compatibility of scores across different types of sub-models. There are two methods that can be applied to tackle such challenges: score unification and score aggregation. We use the statistical scaling method described in [14] and the probability aggregation method described in [15], respectively.

First, based on the observation of how the normal samples' scores in the training dataset resemble, we choose the Gamma distribution assuming $s_i(x) \sim \Gamma(\alpha_i, \beta_i)$, where $s_i(x)$ is the anomaly score of sub-model i. The scaled sub-models' scores $\hat{s}_i(x)$ can be obtained by taking the cumulative distribution function (CDF) $F_i(x; \alpha_i, \beta_i)$ over the fitted distribution from normal samples in the training dataset, i.e.:

$$\hat{s}_i(x) = F_i(x; \alpha_i, \beta_i) \tag{5}$$

$$= \frac{\gamma_i(\alpha_i, \beta_i s_i(x))}{\Gamma(\alpha_i)} \tag{6}$$

where α_i is a shape parameter and β_i is a rate parameter.

Next we obtain the ensemble score by probability aggregation using either $s_i(x)$ or $\hat{s}_i(x)$:

$$s_e(x) = 1 - \prod_{i=1}^{n}(1 - s_i(x)) \tag{7}$$

$$\hat{s}_e(x) = 1 - \prod_{i=1}^{n}(1 - \hat{s}_i(x)) \quad (8)$$

where n is the number of sub-models, s_e is the ensemble anomaly score by direct probability aggregation, and $\hat{s}_e(x)$ is the ensemble anomaly score scaled with CDF.

4. EXPERIMENTAL RESULTS

We implemented anomaly detection on the DCASE 2020 Task 2 dataset with the following four types of approaches:

(1) Ensemble method [Eq. (7)]

(2) Ensemble method scaled with CDF [Eq. (8)]

(3) Machine types and IDs as class labels [Eq. (1)]

(4) Sound data augmentation types as pseudo labels [Eq. (2)]

On the test data in the development dataset, we achieved the following AUC and pAUC results as shown in Tables 1 and 2.

Table 1: AUC performance

Machine Type	Baseline	(1)	(2)	(3)	(4)
ToyCar	78.77	95.66	95.74	92.48	91.37
ToyConveyor	72.53	81.71	81.60	76.90	79.45
fan	65.83	89.05	88.73	89.13	81.27
pump	72.89	93.32	93.20	91.60	90.44
slider	84.76	99.50	99.47	99.31	98.08
valve	66.28	99.77	99.77	99.53	98.81

Table 2: pAUC performance

Machine Type	Baseline	(1)	(2)	(3)	(4)
ToyCar	67.58	88.13	88.15	77.68	87.77
ToyConveyor	60.43	67.68	67.71	63.48	66.81
fan	52.45	80.98	79.82	81.49	71.94
pump	59.99	82.95	82.52	80.83	79.24
slider	66.53	97.35	97.20	96.48	90.76
valve	50.98	98.77	98.79	98.65	94.98

5. CONCLUSIONS

In this technical report, we have described how we apply a classifier confidence based approach to the anomaly detection problem in DCASE 2020 Task 2. We train a classsifer using only normal sound in the development set to learn the distribution of the normal data and use the model confidence to calculate anomaly score. The proposed system does not use any external dataset nor any pre-trained model. The experimental results show superior performance to the baseline on the development dataset.

6. REFERENCES

[1] C. C. Aggarwal, *Outlier Analysis*, 2nd ed. Springer Publishing Company, Incorporated, 2016.

[2] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," in *arXiv:1901.03407*, 2019.

[3] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '17, New York, NY, USA, 2017, pp. 665—-674.

[4] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 146–157.

[5] D. Kimura, S. Chaudhury, M. Narita, A. Munawar, and R. Tachibana, "Adversarial discriminative attention for robust anomaly detection," in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 2161–2170.

[6] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proceedings of the 35th International Conference on Machine Learning*, 10–15 Jul 2018, pp. 4393–4402.

[7] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proceedings of International Conference on Learning Representations*, 2017.

[8] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. 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)*, November 2019, pp. 308–312.

[9] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. 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)*, November 2019, pp. 209–213.

[10] 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*, June 2020, pp. 1–4.

[11] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS'18, 2018, p. 9781–9791.

[12] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450–5463, 2019.

[13] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer Vision – ECCV 2016*, 2016, pp. 499–515.

[14] H. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Interpreting and unifying outlier scores," in *Proceedings of the 11th SIAM International Conference on Data Mining, SDM 2011*, Dec. 2011, pp. 13–24.

[15] J. Gao and P.-N. Tan, "Converting output scores from outlier detection algorithms into probability estimates," in *Sixth International Conference on Data Mining (ICDM'06)*, 2006, pp. 212–221.