

DESCRIPTION AND DISCUSSION ON DCASE 2022 CHALLENGE TASK 2: UNSUPERVISED ANOMALOUS SOUND DETECTION FOR MACHINE CONDITION MONITORING APPLYING DOMAIN GENERALIZATION TECHNIQUES

Kota Dohi¹, Keisuke Imoto², Noboru Harada³, Daisuke Niizumi³, Yuma Koizumi⁴, Tomoya Nishida¹, Harsh Purohit¹, Takashi Endo¹, Masaaki Yamamoto¹, and Yohei Kawaguchi¹

¹ Hitachi, Ltd., Japan, kota.dohi.gr@hitachi.com

² Doshisha University, Japan, keisuke.imoto@ieee.org

³ NTT Corporation, Japan, noboru.harada.pv@hco.ntt.co.jp

⁴ Google, Japan, koizumiyuma@google.com

ABSTRACT

We present the task description and discussion on the results of the DCASE 2022 Challenge Task 2: “Unsupervised anomalous sound detection (ASD) for machine condition monitoring applying domain generalization techniques”. Domain shifts are a critical problem for the application of ASD systems. Because domain shifts can change the acoustic characteristics of data, a model trained in a source domain performs poorly for a target domain. In DCASE 2021 Challenge Task 2, we organized an ASD task for handling domain shifts. In this task, it was assumed that the occurrences of domain shifts are known. However, in practice, the domain of each sample may not be given, and the domain shifts can occur implicitly. In 2022 Task 2, we focus on domain generalization techniques that detects anomalies regardless of the domain shifts. Specifically, the domain of each sample is not given in the test data and only one threshold is allowed for all domains. Analysis of 81 submissions from 31 teams revealed two remarkable types of domain generalization techniques: 1) domain-mixing-based approach that obtains generalized representations and 2) domain-classification-based approach that explicitly or implicitly classifies different domains to improve detection performance for each domain.

Index Terms— anomaly detection, acoustic condition monitoring, domain shift, domain generalization, DCASE Challenge,

1. INTRODUCTION

Anomalous sound detection (ASD) [1–7] is the task of identifying whether the sound emitted from a target machine is normal or anomalous. Automatic detection of mechanical failure is essential in the fourth industrial revolution, which involves artificial intelligence (AI)-based factory automation. Prompt detection of machine anomalies by observing sounds is useful for machine condition monitoring.

One challenge regarding the application scope of ASD systems is that anomalous samples for training can be insufficient both in number and type. In 2020, we organized the fundamental ASD task in Detection and Classification of Acoustic Scenes and Event (DCASE) Challenge 2020 Task 2 [8]; “*unsupervised ASD*” that was aimed to detect unknown anomalous sounds using only normal sound samples as the training data [1–7]. For the wide spread application of ASD systems, advanced tasks such as handling of domain shifts should be tackled. Domain shifts are differences in acoustic

characteristics between the source and target domain data caused by differences in a machine’s operational conditions or environmental noise. Because these shifts are caused by factors other than anomalies, the detection performance of models trained with the source domain data can degrade for the target domain data. Therefore, in 2021, we organized DCASE Challenge 2021 Task 2 [9], “*unsupervised ASD under domain shifted conditions*” that focused on handling domain shifts using domain adaptation techniques.

The task in 2021 involved the use of domain adaptation techniques under two assumptions. First, all domain shifts have been detected in advance, and the domain of each sample is known. Second, the domain shifts do not occur too frequently for the model to adapt. However, these assumptions may not hold for certain real-world scenarios. For example, a machine’s background sound can be affected by various sound sources surrounding the machine, and it can be difficult to identify the cause of changes and attribute the changes to the domain shift. Also, because the operational conditions of the machine can change within a short period, adapting the model every time can be too costly. Therefore, methods have to be investigated such that the detection of domain shifts is unnecessary and frequent occurrences of domain shifts can be handled.

To solve the problem described above, we designed DCASE challenge 2022 Task 2 “*Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring Applying Domain Generalization Techniques*”. This task is aimed at developing domain generalization techniques to handle domain shifts. The task involves the use of domain generalization techniques so that the developed ASD systems do not require detection of the domain shifts or adaptation of the model. Specifically, to evaluate the generalization performance, the domain of each sample is not provided in the test data. To enhance generalization of the model, attributes that caused domain shifts are also provided in the training data.

We received 81 submissions from 31 teams. By analyzing these submissions, we found two types of domain generalization techniques: 1) domain-mixing-based approach and 2) domain-classification-based approach. The domain-mixing-based approach aims at obtaining generalized representations across domains by mixing data from different domains. In contrast, the domain-classification-based approach differentiates different domains so that the model can be specialized for each domain.

2. UNSUPERVISED ANOMALOUS SOUND DETECTION APPLYING DOMAIN GENERALIZATION TECHNIQUES

Let the L -sample time-domain observation $\mathbf{x} \in \mathbb{R}^L$ be an audio clip that includes a sound emitted from a machine. The ASD task is a task to determine whether a machine is in a normal or anomalous state using an anomaly score $\mathcal{A}_\theta(\mathbf{x})$ calculated by an anomaly score calculator $\mathcal{A} : \mathbb{R}^L \rightarrow \mathbb{R}$ with parameters θ . The machine is determined to be anomalous when $\mathcal{A}_\theta(\mathbf{x})$ exceeds a pre-defined threshold ϕ as

$$\text{Decision} = \begin{cases} \text{Anomaly} & (\mathcal{A}_\theta(\mathbf{x}) > \phi) \\ \text{Normal} & (\text{otherwise}). \end{cases} \quad (1)$$

The primary difficulty in this task is to train \mathcal{A} with only normal sounds. This is because anomalies are rarely obtained in practice.

Domain-shift is another major issue in real-world applications. Domain shifts mean a difference in conditions between training and testing. The conditions are machine’s operational conditions such as its speed, load, and temperature, or the environmental conditions such as the type of environmental noise, level of the noise, and location of the microphone. Differences in these conditions change the distribution of data and degrades the detection performance. Let us define two domains: **source domain** and **target domain**, where the source domain is the original condition with enough training clips and the target domain is another condition with zero or a few training clips. Also, let \mathcal{D}_S , \mathcal{D}_T , \mathcal{D}_{SA} , and \mathcal{D}_{TA} be the distributions of \mathbf{x} under the normal condition in the source domain, normal condition in the target domain, anomalous condition in the source domain, and anomalous condition in the target domain, respectively.

The task in DCASE 2021 Task 2 involved two tasks. One was to detect anomalies in the source domain: determine whether \mathbf{x}_s is from \mathcal{D}_S or \mathcal{D}_{SA} using an anomaly score calculator $\mathcal{A}_{\theta_s}(\mathbf{x})$ and a threshold ϕ_s . The other was detection in the target domain: whether \mathbf{x}_t is generated from \mathcal{D}_T or \mathcal{D}_{TA} using an anomaly score calculator $\mathcal{A}_{\theta_t}(\mathbf{x})$ and a threshold ϕ_t . The task was set to develop domain adaptation techniques so that the detection performance on the target domain can be improved by adaptation on the model trained with the source domain data. Although this problem setting assumes that the domain (source/target) of each sample is known, in practice, the detection of domain shifts can be difficult and the domain may not be available. Also, the use of domain adaptation techniques can be too costly if the domain shifts occur too frequently.

We show four types of real-world scenarios for these problems.

Domain shifts due to differences in machine’s conditions

Characteristics of a machine sound can change due to changes in the machine’s operational conditions. Although these shifts can be detected, if these conditions change within a short period of time, it can be too costly to adapt the model every time.

Domain shifts due to differences in environmental conditions

Because characteristics of background noise can be affected by various factors, it is difficult to detect these shifts. Therefore, a model that is unaffected by these shifts is desirable.

Domain shifts due to maintenance

Characteristics of a machine sound can change after maintenance or parts replacement. Though these shifts can be detected, adapting the model every time can be costly.

Domain shifts due to differences in recording devices

In real-world scenarios, many microphones are installed at different locations, and these microphones may be from different manufacturers. Although these shifts can be detected, adapting the model

for each location or microphone can be too costly.

As a possible solution to handle these problems, domain generalization techniques should be investigated. Domain generalization techniques for ASD aims at detecting anomalies from different domains with a single threshold. These techniques, unlike domain adaptation techniques, do not require detection of domain shifts or adaptation of the model in the testing phase. Therefore, domain generalization techniques can be used for handling domain shifts that are difficult to detect or too costly to adapt.

The DCASE 2022 Task 2 is set to develop domain generalization techniques for ASD. Because the domain generalization techniques are expected to work regardless of the domains, the domain of each sample is not given in the test data. The task is to determine if \mathbf{x} is from the normal condition $\mathcal{D}_S \cup \mathcal{D}_T$ or anomalous condition $\mathcal{D}_{SA} \cup \mathcal{D}_{TA}$ using an anomaly score calculator $\mathcal{A}_\theta(\mathbf{x})$ and ϕ . Because the differences in operational or environmental conditions make $\mathcal{D}_S \neq \mathcal{D}_T$, the decision must be executed without being affected by the differences between different domains.

3. TASK SETUP

3.1. Dataset

We used ToyADMOS2 [10] and MIMII DG [11] to generate the dataset. The dataset consists of normal/anomalous operating sounds from seven types of toy/real machines (ToyCar, ToyTrain, fan, gearbox, bearing, slide rail, and valve).

Each recording is a single-channel and 10-sec-long audio with a sampling rate of 16 kHz. We mixed machine sounds recorded at laboratories and the environmental noise recorded at real-world factories to create the training/test data. Details of the recording procedure can be found in [10] and [11].

In this dataset, **Machine type** means the type of machine. **Section** is defined as a subset of the data within a machine type and corresponds to a type of domain shift scenario.

We provide three datasets: **development dataset**, **additional training dataset**, and **evaluation dataset**. The **development dataset** consists of three sections (Sections 00, 01, and 02), which are sets of the training and test data. Each section provides (i) 990 normal clips from a source domain for training, (ii) 10 normal clips from a target domain for training, (iii) 100 normal clips and 100 anomalous clips from both domains for the test. We provided domain information (source/target) in the test data for the convenience of participants. Attributes represent the operational or environmental conditions, e.g. velocity of slide rail and level of noise (SNR) mixed in fan data. The **additional training dataset** provides training clips for three sections (Sections 03, 04, and 05). Each section consists of (i) 990 normal clips in a source domain for training and (ii) 10 normal clips in a target domain for training. Attributes are also provided. The **evaluation dataset** provides test clips for three sections (Sections 03, 04, and 05). Each section consists of 200 test clips, none of which have a condition label (i.e., normal or anomaly) or the domain information. Attributes are not provided. The main difference from our task in 2021 is that the domain information is not given in the evaluation dataset. Thus, the participants have to develop a system that performs well regardless of the domains.

3.2. Evaluation metrics

This task is evaluated with the area under the receiver operating characteristic (ROC) curve (AUC) and the partial AUC (pAUC).

The pAUC is calculated as the AUC over a low false-positive-rate (FPR) range $[0, p]$. In this task, we used $p = 0.1$.

Because the domain generalization task requires detecting anomalies using the same threshold between domains, the pAUC has to be calculated for each section, not for each domain. We calculated the AUC for each domain and pAUC for each section as

$$AUC_{m,n,d} = \frac{1}{N_d^- N_n^+} \sum_{i=1}^{N_d^-} \sum_{j=1}^{N_n^+} \mathcal{H}(\mathcal{A}_\theta(x_j^+) - \mathcal{A}_\theta(x_i^-)), \quad (2)$$

$$pAUC_{m,n} = \frac{1}{P_n^- N_n^+} \sum_{i=1}^{P_n^-} \sum_{j=1}^{N_n^+} \mathcal{H}(\mathcal{A}_\theta(x_j^+) - \mathcal{A}_\theta(x_i^-)), \quad (3)$$

where $P_n^- = \lfloor pN_n^- \rfloor$, m represents the index of a machine type, n represents the index of a section, $d = \{\text{source, target}\}$ represents a domain, $\lfloor \cdot \rfloor$ is the flooring function, and $\mathcal{H}(x)$ returns 1 when $x > 0$ and 0 otherwise. Here, $\{\mathcal{A}_\theta(x_i^-)\}$ and $\{\mathcal{A}_\theta(x_j^+)\}$ are sets of anomaly scores of normal and anomalous test clips, ordered in descending power, respectively. N_d^- is the number of normal test clips in domain d , N_n^- and N_n^+ are the number of normal and anomalous test clips in section n , respectively. We calculated $AUC_{m,n,d}$ to evaluate the contribution of each domain to $AUC_{m,n}$, as it holds that $AUC_{m,n} = \sum_d AUC_{m,n,d}$ if $N_{source}^- = N_{target}^-$.

The official score Ω for ranking submitted systems is given by the harmonic mean of the AUC and pAUC scores over all machine types and sections as follows:

$$\Omega = h \{ AUC_{m,n,d}, pAUC_{m,n} \mid m \in \mathcal{M}, n \in \mathcal{S}(m), d \in \{\text{source, target}\} \}, \quad (4)$$

where $h\{\cdot\}$ represents the harmonic mean (over all machine types, sections, and domains), \mathcal{M} represents the set of machine types, and $\mathcal{S}(m)$ represents the set of sections for machine type m .

Participants are required to submit the anomaly score and normal/anomaly decision result of each test clip. Even though the official score can be calculated with only the anomaly scores, decision results are also required because we must determine the threshold in real-world applications.

3.3. Baseline systems and results

The task organizers provide an autoencoder (AE)-based and a MobileNetV2-based baseline systems.

The AE-based system calculates the anomaly score as the reconstruction error of the sound. To determine the threshold, we assume that anomaly scores of normal sound follows a gamma distribution. The parameters of the gamma distribution are estimated from the anomaly scores of normal sound in the training data, and the threshold is calculated by the 90th percentile of the gamma distribution. A test clip is determined to be anomalous if its anomaly score exceeds the threshold.

In the MobileNetV2-based system [12–14], classifiers such as the MobileNetV2 [15] are trained to identify from which section the observed signal was generated. The anomaly score is calculated as the averaged negative logit of the predicted probabilities for the correct section. The threshold is calculated in the same manner as in the AE-based baseline.

Tables 1 and 2 show the AUC and pAUC for the two baselines. Because the results produced with a GPU are generally non-deterministic, the average and standard deviations from five independent trials are also shown in the tables.

Table 1: Results of the AE-based baseline

Section	AUC [%]		pAUC [%]	
	Source	Target		
ToyCar	00	86.42 ± 1.10	41.48 ± 6.11	51.31 ± 1.34
	01	89.85 ± 1.39	41.93 ± 5.36	54.08 ± 1.84
	02	98.84 ± 0.52	26.50 ± 13.52	52.79 ± 1.04
ToyTrain	00	67.54 ± 0.97	33.68 ± 3.12	52.72 ± 1.63
	01	79.32 ± 0.82	29.87 ± 5.62	50.64 ± 2.33
	02	84.08 ± 0.38	15.52 ± 14.90	48.33 ± 2.33
Bearing	00	67.85 ± 19.61	60.17 ± 7.24	54.41 ± 5.72
	01	59.67 ± 12.67	64.65 ± 12.63	55.09 ± 3.36
	02	61.71 ± 33.52	60.55 ± 35.10	64.18 ± 19.79
Fan	00	84.69 ± 1.74	39.35 ± 9.35	59.95 ± 2.00
	01	71.69 ± 0.69	44.74 ± 1.79	51.12 ± 0.55
	02	80.54 ± 1.42	63.49 ± 2.36	62.88 ± 1.55
Gearbox	00	64.63 ± 0.88	64.79 ± 1.06	60.93 ± 2.31
	01	67.66 ± 0.51	58.12 ± 0.38	53.74 ± 0.56
	02	75.38 ± 0.75	65.57 ± 0.82	61.51 ± 0.69
Slide rail	00	81.92 ± 0.81	58.04 ± 1.22	61.65 ± 1.22
	01	67.85 ± 0.53	50.30 ± 1.25	53.06 ± 0.53
	02	86.66 ± 0.39	38.78 ± 5.13	53.44 ± 1.18
Valve	00	54.24 ± 0.68	52.73 ± 1.93	52.15 ± 0.25
	01	50.45 ± 3.67	53.01 ± 1.73	49.78 ± 0.19
	02	51.56 ± 2.89	43.84 ± 1.11	49.24 ± 0.65

Table 2: Results of the MobileNetV2-based baseline

Section	AUC [%]		pAUC [%]	
	Source	Target		
ToyCar	00	47.40 ± 7.22	56.40 ± 4.11	49.96 ± 2.56
	01	62.02 ± 11.07	56.38 ± 11.31	50.92 ± 2.52
	02	74.19 ± 7.94	45.64 ± 11.32	56.51 ± 6.07
ToyTrain	00	46.02 ± 12.21	49.41 ± 15.14	50.25 ± 1.49
	01	71.96 ± 5.72	45.14 ± 13.66	52.97 ± 4.61
	02	63.23 ± 25.60	44.34 ± 21.50	51.54 ± 4.34
Bearing	00	67.85 ± 19.61	60.17 ± 7.24	54.41 ± 5.72
	01	59.67 ± 12.67	64.65 ± 12.63	55.09 ± 3.36
	02	61.71 ± 33.52	60.55 ± 35.10	64.18 ± 19.79
Fan	00	71.07 ± 19.84	62.13 ± 12.50	55.40 ± 11.29
	01	76.26 ± 4.95	35.12 ± 13.38	52.14 ± 4.08
	02	67.29 ± 10.34	58.02 ± 7.46	65.14 ± 1.09
Gearbox	00	63.54 ± 9.46	67.02 ± 13.50	62.12 ± 11.66
	01	66.68 ± 12.29	66.96 ± 8.92	56.85 ± 4.47
	02	80.87 ± 7.85	43.15 ± 16.12	50.62 ± 7.73
Slide rail	00	87.15 ± 2.71	80.77 ± 4.53	71.57 ± 5.28
	01	49.66 ± 30.46	32.07 ± 46.84	48.21 ± 2.73
	02	72.70 ± 11.67	32.94 ± 19.77	49.69 ± 1.63
Valve	00	75.26 ± 4.84	43.60 ± 14.38	55.37 ± 5.86
	01	54.78 ± 5.37	60.43 ± 5.08	54.69 ± 3.87
	02	76.26 ± 1.02	78.74 ± 2.64	85.74 ± 0.08

4. CHALLENGE RESULTS

4.1. Results for evaluation dataset

We received 81 submissions from 31 teams, and 22 teams outperformed the MobileNetV2-based baseline in the official score. In Figure 1, the harmonic means of the AUCs are shown for top 10 teams [16–25]. Although the AUCs change drastically between different machine types and teams, these highly ranked teams outperformed the baselines for most of the machine types. It is worth noting that, for these teams, the source-domain AUC did not correlate with the official rank (correlation coefficient was -0.033) while the target-domain AUC did (correlation coefficient was -0.862). This indicates that handling domain shifts and generalizing the model was the key to better ranks among highly ranked teams.

We find that domain generalization approaches adopted by the participants can be categorized into two types: domain-mixing-based approach and domain-classification-based approach. These methods achieved the aim of the task by generalizing the model using data with different attributes. Figure 2 shows the average source-domain and target-domain AUC of the top 20 teams.

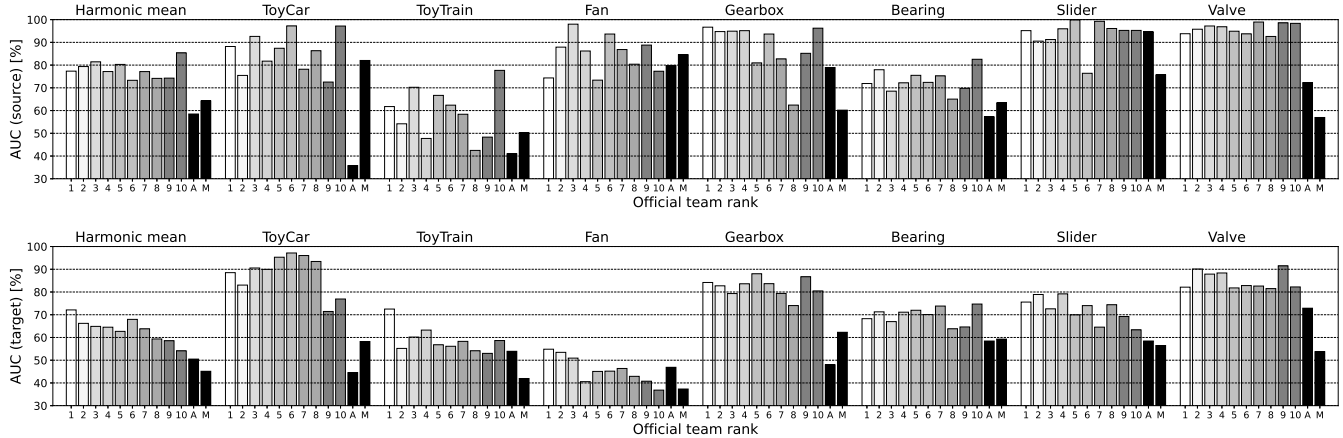


Figure 1: Evaluation results of top 10 teams in the ranking. Average source-domain AUC (Top) and target-domain AUC (bottom) for each machine type. Label “A” and “M” on the x-axis denote AE-based and MobileNetV2-based baselines, respectively.

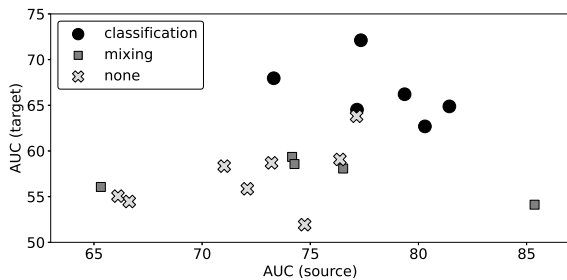


Figure 2: Average source-domain AUC and target-domain AUC of the top 20 teams. “classification” denotes teams that used domain-classification-based approaches, “mix-up” denotes teams that used domain-mixing-based approaches, and “none” denotes teams that did not use particular domain generalization techniques.

Domain-classification-based approaches outperformed other approaches especially for the target domain. However, these approaches may be specialized for the types of target domain data provided in both the training and test data, and thus may not perform well for those not included in the train data. We describe the details in the following.

4.2. Domain-mixing-based approach

Domain-mixing-based approach extracts common representations between domains. These include batch mixing that use data from both domains in a batch to train a model [17, 23], Mixup [26] that synthesizes data from both domains to obtain intermediate representations [17, 23, 25, 27], and data augmentation techniques to obtain robust representations [24]. These techniques use the target domain data to expand the normal conditions for the model so that the model can be generalized to better handle domain shifts. However, as shown in Figure 2, they have been outperformed by domain-classification-based approaches. This can be that, for the Mixup and data augmentation, synthesized data was not useful for representing the target domain data. One future direction can be on obtaining meaningful synthetic representations with the aid of external information such as the attribute information.

4.3. Domain-classification-based approach

Domain-classification-based approach distinguishes the source and target domain data to obtain better detection performance for each domain. The 1st and 6th place teams [16, 21] used distances between the embedding of a domain and that of the test data to calculate anomaly scores. Because the domain of each sample can be estimated by the domain with shorter distance, this approach can be regarded as implicitly classifying the domain of each sample. The 2nd, 3rd, 4th, and 5th place teams [17–20] explicitly trained a classifier to distinguish the attributes or the domains. The 5th place teams trained an attribute classifier and a section classifier so that both the domain-wise information from the attribute classifier and the domain-independent information from the section classifier can be obtained.

As shown in Figure 2, the domain-classification-based approach outperformed the domain-mixing-based approach. This can be because the normal conditions are defined for each specific domain, unlike the domain-mixing-based approach that defines normal conditions over all domains. However, this approach assumes that the target domain data in the train data includes all types of the target domain data in the test data. If the target domain data in the test data contains too many types of data not included in the train data, the classifier may fail to distinguish domains, which can degrade the detection performance. Therefore, further investigation is needed to examine the ability of this approach to handle completely unseen target domain data.

5. CONCLUSION

This paper presented an overview of the task and analysis of the solutions submitted to DCASE 2022 Challenge Task 2. To handle domain shifts that occur implicitly, the task was dedicated to developing domain generalization techniques. The organization of the task revealed two approaches that can be useful for domain generalization task: domain-mixing-based approach and domain-classification-based approach. For the former approach, obtaining more meaningful synthetic representations from multiple domains is left for future works. For the latter approach, future works can focus on analyzing the effect of this approach on completely unseen types of target domain data.

6. REFERENCES

- [1] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, “Optimizing acoustic feature extractor for anomalous sound detection based on Neyman-Pearson lemma,” in *Proc. 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 698–702.
- [2] Y. Kawaguchi and T. Endo, “How can we detect anomalies from subsampled audio signals?” in *Proc. 27th IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2017.
- [3] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, “Unsupervised detection of anomalous sound based on deep learning and the Neyman-Pearson lemma,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212–224, Jan. 2019.
- [4] Y. Kawaguchi, R. Tanabe, T. Endo, K. Ichige, and K. Hamada, “Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction,” in *Proc. 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 865–869.
- [5] Y. Koizumi, S. Saito, M. Yamaguchi, S. Murata, and N. Harada, “Batch uniformization for minimizing maximum anomaly score of DNN-based anomaly detection in sounds,” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 6–10.
- [6] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, “Anomalous sound detection based on interpolation deep neural network,” in *Proc. 45th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 271–275.
- [7] H. Purohit, R. Tanabe, T. Endo, K. Suefusa, Y. Nikaido, and Y. Kawaguchi, “Deep autoencoding GMM-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization,” in *Proc. 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2020, pp. 175–179.
- [8] 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 *Proc. 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2020, pp. 81–85.
- [9] 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 detection for machine condition monitoring under domain shifted conditions,” in *Proc. 6th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2021, pp. 186–190.
- [10] 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 *Proc. 6th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2021, pp. 1–5.
- [11] 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,” *arXiv preprint arXiv:2205.13879*, 2022.
- [12] R. Giri, S. V. Tenneti, F. Cheng, K. Helwani, U. Isik, and A. Krishnaswamy, “Self-supervised classification for detecting anomalous sounds,” in *Proc. 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2020, pp. 46–50.
- [13] P. Primus, V. Haunschmid, P. Praher, and G. Widmer, “Anomalous sound detection as a simple binary classification problem with careful selection of proxy outlier examples,” in *Proc. 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2020, pp. 170–174.
- [14] T. Inoue, P. Vinayavekhin, S. Morikuni, S. Wang, T. H. Trong, D. Wood, M. Tsubori, and R. Tachibana, “Detection of anomalous sounds for machine condition monitoring using classification confidence,” in *Proc. 5th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2020, pp. 66–70.
- [15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proc. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [16] Y. Zeng, H. Liu, L. Xu, Y. Zhou, and L. Gan, “Robust anomaly sound detection framework for machine condition monitoring,” DCASE2022 Challenge, Tech. Rep., 2022.
- [17] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Two-stage anomalous sound detection systems using domain generalization and specialization techniques,” DCASE2022 Challenge, Tech. Rep., 2022.
- [18] F. Xiao, Y. Liu, Y. Wei, J. Guan, Q. Zhu, T. Zheng, and J. Han, “The dcase2022 challenge task 2 system: Anomalous sound detection with self-supervised attribute classification and gmm-based clustering,” DCASE2022 Challenge, Tech. Rep., 2022.
- [19] Y. Deng, J. Liu, and W.-Q. Zhang, “Aithu system for unsupervised anomalous detection of machine working status via sounding,” DCASE2022 Challenge, Tech. Rep., 2022.
- [20] S. Venkatesh, G. Wichern, A. Subramanian, and J. Le Roux, “Disentangled surrogate task learning for improved domain generalization in unsupervised anomalous sound detection,” DCASE2022 Challenge, Tech. Rep., 2022.
- [21] Y. Wei, J. Guan, H. Lan, and W. Wang, “Anomalous sound detection system with self-challenge and metric evaluation for dcase2022 challenge task 2,” DCASE2022 Challenge, Tech. Rep., 2022.
- [22] K. Morita, T. Yano, and K. Tran, “Comparative experiments on spectrogram representation for anomalous sound detection,” DCASE2022 Challenge, Tech. Rep., 2022.
- [23] J. Bai, Y. Jia, and S. Huang, “Jless submission to dcase2022 task2: Batch mixing strategy based method with anomaly detector for anomalous sound detection,” DCASE2022 Challenge, Tech. Rep., 2022.
- [24] S. Verbitskiy, M. Shkhanukova, and V. Vyshegorodtsev, “Unsupervised anomalous sound detection using multiple time-frequency representations,” DCASE2022 Challenge, Tech. Rep., 2022.
- [25] K. Wilkinghoff, “An outlier exposed anomalous sound detection system for domain generalization in machine condition monitoring,” DCASE2022 Challenge, Tech. Rep., 2022.
- [26] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” in *International Conference on Learning Representations*, 2018. [Online]. Available: <https://openreview.net/forum?id=r1Ddp1-Rb>
- [27] I. Nejjar, J. P. J. Meunier-Pion, G. M. Frusque, and O. Fink, “Dcase challenge 2022: Self-supervised learning pre-training, training for unsupervised anomalous sound detection,” DCASE2022 Challenge, Tech. Rep., 2022.