

# DESCRIPTION AND DISCUSSION ON DCASE 2025 CHALLENGE TASK 2: FIRST-SHOT UNSUPERVISED ANOMALOUS SOUND DETECTION FOR MACHINE CONDITION MONITORING

Tomoya Nishida<sup>1</sup>, Noboru Harada<sup>2</sup>, Daisuke Niizumi<sup>2</sup>, Davide Albertini<sup>3</sup>, Roberto Sannino<sup>3</sup>,  
Simone Pradolini<sup>3</sup>, Filippo Augusti<sup>3</sup>, Keisuke Imoto<sup>4</sup>, Kota Dohi<sup>1</sup>, Harsh Purohit<sup>1</sup>,  
Takashi Endo<sup>1</sup>, and Yohei Kawaguchi<sup>1</sup>

<sup>1</sup> Hitachi, Ltd., Japan, tomoya.nishida.ax@hitachi.com

<sup>2</sup> NTT, Inc., Japan, harada.noboru@ntt.com

<sup>3</sup> STMicroelectronics, Switzerland,

<sup>4</sup> Kyoto University, Japan, keisuke.imoto@ieee.org

## ABSTRACT

This paper introduces the task description for the Detection and Classification of Acoustic Scenes and Events (DCASE) 2025 Challenge Task 2, titled “First-shot unsupervised anomalous sound detection (ASD) for machine condition monitoring.” Building on the DCASE 2024 Challenge Task 2, this task is structured as a first-shot problem within a domain generalization framework. The primary objective of the first-shot approach is to facilitate the rapid deployment of ASD systems for new machine types without requiring machine-specific hyperparameter tunings. For DCASE 2025 Challenge Task 2, sounds from previously unseen machine types have been collected and provided as the evaluation dataset. We received 119 submissions from 35 teams, and an analysis of these submissions has been made in this paper. Analysis showed that various approaches can all be competitive, such as fine-tuning pre-trained models, using frozen pre-trained models, and training small models from scratch, when combined with appropriate cost functions, anomaly score normalization, and use of clean machine and noise sounds.

**Index Terms**— anomaly detection, acoustic condition monitoring, domain shift, first-shot problem, DCASE Challenge

## 1. INTRODUCTION

Anomalous sound detection (ASD) [1–7] involves determining whether the sound emitted from a target machine is normal or anomalous. This capability plays a crucial role in automating the detection of mechanical failures, which is essential in the era of the fourth industrial revolution and AI-driven factory automation.

One of the key challenges in developing ASD systems lies in the scarcity and limited diversity of anomalous samples available for training. To address this, the first ASD task was introduced in the DCASE Challenge 2020 Task 2 [8], focusing on “unsupervised ASD (UASD),” which aimed to detect unknown anomalous sounds using only normal sound samples for training. Building on this, subsequent challenges in 2021 and 2022 [9, 10] tackled the issue of domain shifts to enable broader application of ASD systems. Domain shifts refer to discrepancies between data from the source and target domains, arising due to variations in machine operational conditions or environmental noise.

The DCASE 2023/2024 Task 2 (“first-shot” UASD) [11, 12] targeted a realistic setting where systems must detect anomalies

for entirely novel machine types without access to similar-type data for training or hyperparameter tuning. This reflects rapid-deployment scenarios in which collecting diverse training or test data—especially anomalous samples—is infeasible, and therefore manual test-driven tuning is unrealistic. Accordingly, the evaluation data comprise machine types absent from the development set to enforce this constraint.

The DCASE2025 Challenge Task 2 maintains the previous task setting as a first-shot problem under domain generalization conditions, using newly recorded machine sound data as the evaluation dataset. In addition, there are several modifications: We provide additional supplementary data for each machine, including clean machine recordings or noise samples, which can optionally be used to enhance ASD performance in noisy environments. Also, participants are asked to provide the computational complexity of their solutions. Although this score is not used for the official rankings, it helps clarify the balance between model complexity and performance—a key factor for lightweight ASD applications on edge devices. In this paper, we provide explanations on this task and discuss the challenge results.

## 2. FIRST-SHOT UNSUPERVISED ANOMALOUS SOUND DETECTION UNDER DOMAIN SHIFTED CONDITIONS

Consider an audio clip  $\mathbf{x}$ , which contains sounds produced by a machine. The objective of the ASD task is to classify the machine as either normal or anomalous by calculating an anomaly score  $\mathcal{A}_\theta(\mathbf{x})$  using an anomaly score calculator  $\mathcal{A}$  with parameters  $\theta$ . The input of  $\mathcal{A}$  can be the audio clip  $\mathbf{x}$  with or without additional information such as labels indicating the operation condition of the machine. The machine is then determined to be anomalous when  $\mathcal{A}_\theta(\mathbf{x})$  exceeds a pre-defined threshold  $\phi$  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 the anomaly score calculator with only normal sounds (UASD). The DCASE 2020 Challenge Task 2 [8] was designed to address this issue, and all the following tasks stand on this UASD setting.

Addressing the domain-shift problem is also essential for the practical implementation of ASD systems. Domain shifts refer to variations in conditions between training and testing phases, which

alter the distribution of the observed sound data. These variations can result from differences in operating speed, machine load, heating temperature, microphone arrangement, environmental noise, and other factors. Two domains are defined: the **source domain**, representing the original condition with sufficient training data, and the **target domain**, representing another condition where only limited samples are available. This year’s task follows the 2022 to 2024 Task 2 [10–12] setting, where the domain information is assumed to be unknown in the test phase and anomalies from both domains have to be detected with a single threshold. In this case, domain generalization is required to achieve good performance.

To further pursue the rapid development of ASD systems in real-world scenarios, solving ASD (a) against completely novel machine types (b) with only one section of training data (c) without handcrafted tunings that depend on test data, are highly important. This is because in real-world scenarios, customers may only possess a single novel machine, and collecting test data—especially the anomalous samples—for handcrafted tuning may be infeasible. This problem setting was named as the “first-shot problem”, and the 2023 and 2024 Task 2 [11, 12] was organized based on this problem setting. The first-shot problem was implemented by introducing two key features to the dataset: (i) The development and evaluation datasets consist of entirely different sets of machine types, and (ii) Each machine type in the dataset contains only a single section. Note that until 2022 Task 2, the provided dataset included multiple sections for each machine type, and the development and evaluation datasets sharing the same machine types.

The DCASE2025 Challenge Task 2 retains the previous task setting as a first-shot problem under domain generalization conditions, while introducing several modifications. First, we have provided additional supplementary data, including clean machine recordings and noise samples. These resources may reflect practical scenarios—such as collecting clean machine data when a factory is idle or gathering noise recordings when the machine is not running. Participants are free to incorporate these additional sources to enhance the accuracy of their models. Second, although large-scale models—such as pretrained networks and ensembles—have become increasingly popular in this task, lightweight models capable of running on edge devices also remain an important area of research. To acknowledge this, participants were optionally asked to report the computational complexity of their solutions in terms of Multiply-Accumulate Operations (MACs). Although this metric does not affect the official rankings, it provides valuable insight into the balance between model complexity and performance.

### 3. TASK SETUP

#### 3.1. Dataset

The dataset for this task is divided into three categories: the **development dataset**, the **additional training dataset**, and the **evaluation dataset**. The development dataset contains seven machine types, while the additional training and evaluation datasets include nine machine types, with each machine type consisting of a single section. A **machine type** refers to the category of machines, such as fans or gearboxes, and a **section** represents a subset or the entirety of the data associated with each machine type.

All recording are single-channel, lasting 6 to 10 seconds, and have a sampling rate of 16 kHz. The machine sounds recorded at laboratories were mixed with environmental noise recorded at factories and in the suburbs to create each sample in the dataset. For further details of the recording process, please refer to the papers on

ToyADMOS2 [13] and MIMII DG [14].

The **development dataset** provides seven machine types (fan, gearbox, bearing, slide rail, valve, ToyCar, ToyTrain), and each machine type has one section that contains a complete set of the training and test data. Each section contains (i) 990 normal clips from a source domain for training, (ii) 10 normal clips from a target domain for training, (iii) 100 clips of supplementary sound data containing either clean normal machine sounds in the source domain or noise-only sounds, and (iv) 100 normal clips and 100 anomalous clips from both domains for the test. To assist participants, domain information (source/target) was included in the test data. For four machine types (fan, gearbox, valve, and ToyCar) details regarding operational or environmental conditions were provided in the file names and attribute CSV files. However, for the remaining three machine types, these attributes were not disclosed.

The **additional training dataset** provides novel nine machine types (AutoTrash, HomeCamera, ToyPet, ToyRCCar, BandSealer, Polisher, ScrewFeeder, CoffeeGrinder). Each section consists of (i) 990 normal clips in a source domain for training, (ii) 10 normal clips in a target domain for training, and (iii) 100 clips of supplementary sound data containing either clean normal machine sounds in the source domain or noise-only sounds. For five machine types (HomeCamera, ToyRCCar, BandSealer, and CoffeeGrinder), attributes were provided in this dataset. For the other four machine types, attributes were concealed. The **evaluation dataset** provides the test clips that correspond to the additional training dataset, e.g. data of the same machine types as the additional training dataset. Each section consists of 200 test clips, none of which have a condition label (i.e., normal or anomaly), domain information, or attribute information. Participants are required to train a model for each new machine type using only a single section per machine type.

#### 3.2. Evaluation metrics

To assess overall detection performance, we employed the area under the receiver operating characteristic curve (AUC). Additionally, we used the partial AUC (pAUC) to evaluate performance in a low false-positive rate range  $[0, p]$ , where we set  $p = 0.1$ . To evaluate each system under the domain generalization setting, we compute the AUC for each domain and pAUC for each section as

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

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

where  $m$  and  $n$  represent the index of a machine type and a section respectively,  $d \in \{\text{source}, \text{target}\}$  represents a domain,  $\lfloor \cdot \rfloor$  is the flooring function, and  $\mathcal{H}(y)$  returns 1 when  $y > 0$  and 0 otherwise.

Here,  $\{x_i^-\}_{i=1}^{N_d^-}$  are the normal test clips in domain  $d$  in section  $n$  of machine type  $m$  and  $\{x_j^+\}_{j=1}^{N_n^+}$  are all the anomalous test clips in section  $n$  of machine type  $m$ .  $N_d^-$ ,  $N_n^-$ ,  $N_n^+$  represent the number of normal test clips in domain  $d$ , normal test clips in section  $n$ , and anomalous test clips in section  $n$ , respectively.

The official score  $\Omega$  is given by the harmonic mean of the AUC and pAUC scores overall machine types and sections:

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

where  $h\{\cdot\}$  represents the harmonic mean,  $\mathcal{M}$  is the set of given machine types, and  $\mathcal{S}(m)$  represents the set of sections for machine type  $m$ . Specifically,  $\mathcal{S}(m) = \{00\}$  for the dataset in 2024-2025.

Additionally, although not included in the official rankings, participants were optionally asked to provide information on the computational complexity of their models in terms of MAC operations. It was recommended that this be calculated using the open-source implementation available in [15].

### 3.3. Baseline systems and results

The task organizers offer a baseline system using Autoencoders (AEs) with two operating modes, identical to the 2023 Task 2 baseline. While both modes use Autoencoders for training, they differ in anomaly score computation. This paper presents the system and its detection performance; details can be found in [16].

#### 3.3.1. Autoencoder training

The AE is trained for both operating modes using log-mel-spectrograms of training sound clips  $X = [X_1, \dots, X_T]$ , where  $X_t \in \mathbb{R}^F$  for  $t = 1, \dots, T$  represents frame-wise feature vectors at frame  $t$ , where  $F = 128$  and  $T$  is the number of mel-filters and time-frames, respectively. For input,  $P = 5$  consecutive frames are concatenated as  $\psi_t = [X_t^T, \dots, X_{t+P-1}^T]^T \in \mathbb{R}^D$  for each  $t$ , with  $D = P \times F = 640$ . Model parameters are trained by minimizing the mean squared error (MSE) between the input  $\psi_t$  and the reconstructed output  $r_\theta(\psi_t)$  for all inputs from the training data.

#### 3.3.2. Simple Autoencoder mode

This mode uses the mean MSE of all features derived from the given sound clip as its anomaly score, e.g.,

$$A_\theta(X) = \frac{1}{DK} \sum_{k=1}^K \|\psi_k - r_\theta(\psi_k)\|_2^2, \quad (5)$$

where  $K = T - P + 1$ , and  $\|\cdot\|_2$  represents  $\ell_2$  norm.

#### 3.3.3. Selective Mahalanobis mode

In this mode, the Mahalanobis distance between the system input and reconstructed feature is used to compute the anomaly score. The anomaly score is defined as

$$A_\theta(X) = \frac{1}{DK} \sum_{k=1}^K \min\{D_s(\psi_k, r_\theta(\psi_k)), D_t(\psi_k, r_\theta(\psi_k))\}, \quad (6)$$

$$D_s(\cdot) = \text{Mahalanobis}(\psi_k, r_\theta(\psi_k), \Sigma_s^{-1}), \quad (7)$$

$$D_t(\cdot) = \text{Mahalanobis}(\psi_k, r_\theta(\psi_k), \Sigma_t^{-1}), \quad (8)$$

where  $\Sigma_s^{-1}$  and  $\Sigma_t^{-1}$  are the covariance matrices of  $r_\theta(\psi_k) - \psi_k$  for the source and target domain data of each machine type, respectively.

#### 3.3.4. Results

Tables 1 present the AUC and pAUC results for the two baseline systems on the development dataset, with the averages and standard deviations computed from five independent trials.

Table 1: Baseline results for development dataset.

Machine type	Mode	AUC [%]		pAUC [%]
		Source	Target	
ToyCar	MSE	71.05 $\pm$ 0.50	53.32 $\pm$ 0.56	49.79 $\pm$ 0.49
	MAHALA	73.17 $\pm$ 0.39	50.91 $\pm$ 0.85	49.05 $\pm$ 0.05
ToyTrain	MSE	61.76 $\pm$ 0.74	56.46 $\pm$ 0.47	50.19 $\pm$ 0.25
	MAHALA	50.87 $\pm$ 2.88	46.15 $\pm$ 1.77	48.32 $\pm$ 0.05
bearing	MSE	66.53 $\pm$ 2.63	53.15 $\pm$ 1.99	61.12 $\pm$ 0.59
	MAHALA	63.63 $\pm$ 1.15	59.03 $\pm$ 1.79	61.86 $\pm$ 0.36
fan	MSE	70.96 $\pm$ 0.94	38.75 $\pm$ 0.74	49.46 $\pm$ 0.53
	MAHALA	77.99 $\pm$ 0.23	38.56 $\pm$ 0.58	50.82 $\pm$ 0.06
gearbox	MSE	64.80 $\pm$ 1.48	50.49 $\pm$ 1.22	52.49 $\pm$ 0.37
	MAHALA	73.26 $\pm$ 0.78	51.61 $\pm$ 0.52	55.07 $\pm$ 0.47
slider	MSE	70.10 $\pm$ 1.01	48.77 $\pm$ 1.07	52.32 $\pm$ 0.36
	MAHALA	73.79 $\pm$ 1.95	50.27 $\pm$ 1.15	53.61 $\pm$ 0.26
valve	MSE	63.53 $\pm$ 2.90	67.18 $\pm$ 1.75	57.35 $\pm$ 1.96
	MAHALA	56.22 $\pm$ 2.22	61.00 $\pm$ 2.98	52.53 $\pm$ 1.32

## 4. CHALLENGE RESULTS

### 4.1. Overall results

We received 119 submissions from 35 teams. 20 teams outperformed both baselines, which slightly increased compared to last year's task (11 out of 27 teams). Looking at the results for each domain separately, six teams surpassed the baselines on the source-domain AUC, while 25 did so on the target-domain AUC. Four teams achieved higher AUCs than the baselines in both domains. This shows the difficulty of improving the performance on both the source and target domain at the same time. Figure 1 shows the AUC values for the top 10 teams. In the source domain, whether each team could beat the baseline was highly machine-dependent, and many teams struggled to outperform the baseline on average. Specifically, machines for which attribute information was available tended to perform poorly, although it is unclear whether this factor is actually relevant. In contrast, all of the top 10 teams outperformed the baselines in the target domain in the harmonic mean.

Figure 2 compares the AUC values of the top 20 teams between the development and evaluation datasets. As can be seen, achieving high AUC values in the development dataset does not indicate high AUC in the evaluation dataset. This is a typical trend in the first-shot problem setting which started from 2023, and shows the difficulty in finding an approach robust to machine types in the absence of test data. Thus, building an ASD system that works well for unknown machine types remains both a difficult and important challenge. In Figure 3, we compare the harmonic mean AUC values in the source and target domain of the evaluation dataset among the top 20 teams. The figure shows a negative correlation between the AUC in the source and target domain. While the top 3 teams achieved very close official scores (within 1.0% difference), their balance between the source and target AUC varied. Achieving well balanced performance between the source and target domain may be important to achieve high ranks.

### 4.2. Trends in model size

We analyzed the computational-complexity trends of the submitted systems. MACs were reported by 21 teams. Figure 4 plots each submission's MAC count against its official score; if a team submitted multiple systems with identical MAC values, only the highest-ranked system was retained.

The figure shows a broad range of MAC counts across submissions. It also highlights that larger computational budgets did not necessarily yield higher scores. Notably, two submissions from different teams [17, 18] achieved scores exceeding the baselines while

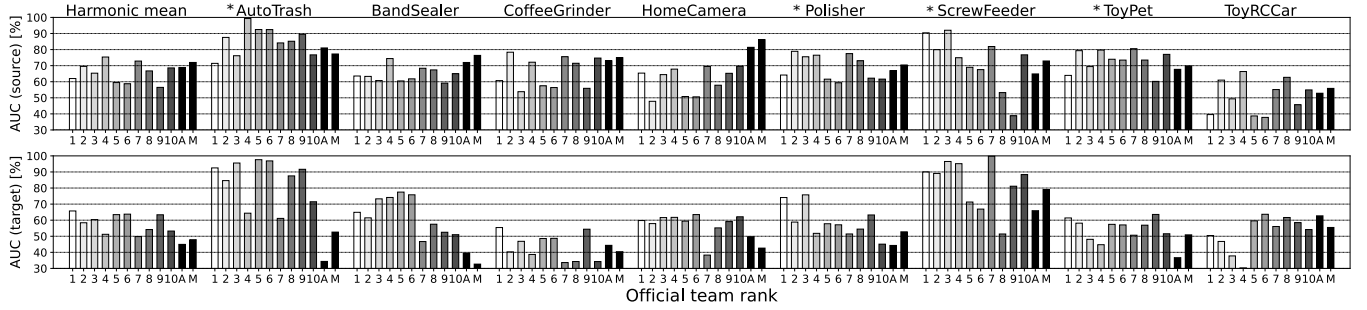


Figure 1: Evaluation results of top 10 teams in ranking. Average source (top) and target-domain AUC (bottom) for each machine type (“\*” indicates that attributes are hidden.). Labels “A” and “M” on denote simple Autoencoder mode and selective Mahalanobis mode, respectively.

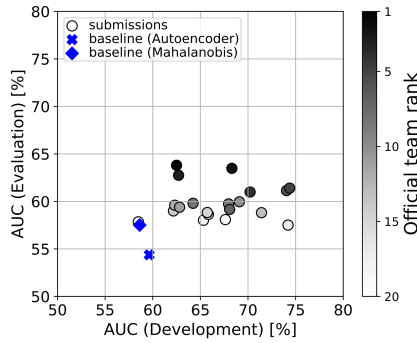


Figure 2: Comparison of harmonic mean of AUC for development and evaluation dataset across teams.

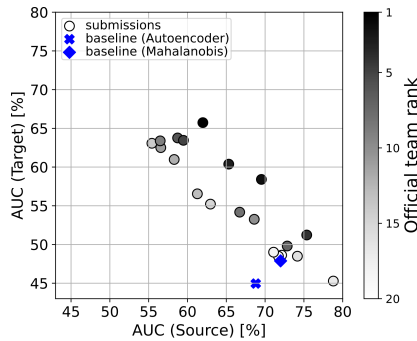


Figure 3: Comparison of harmonic mean of AUC for source and target domain in evaluation dataset across teams.

using fewer MACs. This shows the feasibility of computationally efficient solutions for first-shot UASD, and could be one of the future directions for research.

#### 4.3. New approaches seen in the top-ranked teams

##### a. Use of pretrained models

This year, following the trend from previous years, many participants adopted pretrained models in their anomaly-detection pipelines. Many of those teams fine-tuned them with an attribute or domain classification-based auxiliary task, such as the 1, 4, 5th ranked teams [19–21]. On the other hand, interestingly, several of the high-ranking teams achieved strong performance using frozen pretrained networks, exploiting intermediate-layer features along with anomaly-score normalization [22]. The 2nd [23] and 8th [24]

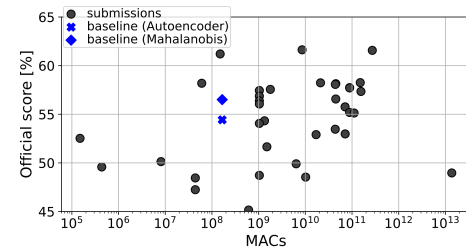


Figure 4: Comparison of official scores of submissions against the MACs values.

place teams solely used frozen models within their submissions, while the 4th-place team [20] ensemble frozen networks to the fine-tuned ones. Nevertheless, teams that trained lightweight models from scratch with classification-based tasks also achieved high ranks, including the 3rd ranked team [17, 18], showing that pre-trained models are not an absolute prerequisite for competitive performance. Overall, diverse approaches were all competitive this year and each approach may still have room for further research.

##### b. Use of supplemental data

Participants tried utilizing the newly released supplemental clean-machine and noise recordings, applying them in two distinct ways. In the first way, they were used for data augmentation. The 1st [19] and several other top-10 teams [24–26] injected the supplemental clips as an extra class in auxiliary classifiers, blended the noise signals with training samples, or leveraged them in contrastive learning [18] to enrich feature space diversity. Conversely, the 3rd and 4th-place teams [17, 20], used the clean/noise data to build enhancement modules that extracted or denoised target machine sounds from the noisy training data, and supplied these extracted signals to their main anomaly-detection networks.

## 5. CONCLUSION

We presented an overview of the DCASE 2025 Challenge Task 2. The task’s aim was to develop ASD systems that work for a novel machine type with a single section for each machine type, where supplemental data such as clean machine sounds or noise-only sounds were also provided. We discussed several new approaches seen in the challenge, such as how pretrained models were (or were not) used and the use of the newly provided supplemental data. While we were not able to discuss all new approaches, we hope that all the technical reports will contribute to the advancements in the field of anomalous sound detection.

## 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. EUSIPCO*, 2017, pp. 698–702.
- [2] Y. Kawaguchi and T. Endo, "How can we detect anomalies from subsampled audio signals?" in *Proc. IEEE 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 TASLP*, 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. IEEE 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 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. IEEE 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 hyperparameter optimization," in *Proc. DCASE Workshop*, 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. DCASE Workshop*, 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. DCASE Workshop*, 2021, pp. 186–190.
- [10] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, and Y. Kawaguchi, "Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," in *Proc. DCASE Workshop*, 2022, pp. 26–30.
- [11] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," in *Proc. DCASE Workshop*, 2023, pp. 31–35.
- [12] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. San-nino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, R. Tanabe, 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 *Proc. DCASE Workshop*, 2024, pp. 111–115.
- [13] 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. DCASE Workshop*, 2021, pp. 1–5.
- [14] 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 *Proc. DCASE Workshop*, 2022.
- [15] L. Zhu, "Thop: Pytorch-opcounter," <https://github.com/Lyken17/pytorch-OpCounter>, 2019.
- [16] N. Harada, N. Daisuke, T. Daiki, O. Yasunori, and Y. Masahiro, "First-shot anomaly detection for machine condition monitoring: A domain generalization baseline," in *Proc. EUSIPCO*, 2023, pp. 191–195.
- [17] J. Yang, "A two stage fusion anomaly detection approach for Task2," DCASE2025 Challenge, Tech. Rep., June 2025.
- [18] Q. Zhou and S. Wu, "Machine anomalous sound detection combining convolutional auto-encoder and contrastive learning," DCASE2025 Challenge, Tech. Rep., June 2025.
- [19] L. Wang, "Pre-trained model enhanced anomalous sound detection system for DCASE2025 Task2," DCASE2025 Challenge, Tech. Rep., June 2025.
- [20] F. Takuya, I. Kuroyanagi, and T. Toda, "The NU systems for DCASE 2025 Challenge Task 2," DCASE2025 Challenge, Tech. Rep., June 2025.
- [21] A. Jiang, W. Liang, S. Feng, Y. Qiu, Y. Zhao, J. Li, P. Fan, W.-Q. Zhang, C. Lu, X. Chen, Y. Qian, and J. Liu, "THUEE system for DCASE 2025 anomalous sound detection challenge," DCASE2025 Challenge, Tech. Rep., June 2025.
- [22] K. Wilkinghoff, H. Yang, J. Ebberts, F. G. Germain, G. Wichern, and J. Le Roux, "Keeping the balance: Anomaly score calculation for domain generalization," in *Proc. IEEE ICASSP*. IEEE, 2025, pp. 1–5.
- [23] P. Saengthong and T. Shinozaki, "GenRep for first-shot unsupervised anomalous sound detection of DCASE2025 Challenge," DCASE2025 Challenge, Tech. Rep., June 2025.
- [24] T. Shiraga, K. Ozeki, T. Masuzaki, N. Tanaka, and T. Kuriyama, "Anomalous sound detection method using contrastive learning," DCASE2025 Challenge, Tech. Rep., June 2025.
- [25] X. Zheng, A. Jiang, B. Han, S. Zhang, W.-Q. Zhang, X. Chen, C. Lu, P. Fan, J. Liu, and Y. Qian, "SJTU-AITHU system for DCASE 2025 anomalous sound detection challenge," DCASE2025 Challenge, Tech. Rep., June 2025.
- [26] S. Zhang, F. Xiao, S. Fan, Q. Zhu, W. Wang, and J. Guan, "Anomalous sound detection using pre-trained model with statistical feature difference representation," DCASE2025 Challenge, Tech. Rep., June 2025.