TOYADMOS2: ANOTHER DATASET OF MINIATURE-MACHINE OPERATING SOUNDS FOR ANOMALOUS SOUND DETECTION UNDER DOMAIN SHIFT CONDITIONS

Noboru Harada, Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi
Masahiro Yasuda and Shoichiro Saito

NTT Corporation, Japan

ABSTRACT

This paper proposes a new large-scale dataset called "ToyADMOS" for anomaly detection in machine operating sounds (ADMOS). As with our previous ToyADMOS dataset, we collected a large number of operating sounds of miniature machines (toys) under normal and anomaly conditions by deliberately damaging them, but extended them in this case by providing a controlled depth of damages in the anomaly samples. Since typical application scenarios of ADMOS require robust performance under domain-shift conditions, the ToyADMOS2 dataset is designed for evaluating systems under such conditions. The released dataset consists of two sub-datasets for machine-condition inspection: fault diagnosis of machines with geometrically fixed tasks and fault diagnosis of machines with moving tasks. Domain shifts are represented by introducing several differences in operating conditions, such as the use of the same machine type but with different models and parts configurations, operating speeds, microphone arrangements, etc. Each sub-dataset contains over 27 k samples of normal machine-operating sounds and over 8 k samples of anomalous sounds recorded with five to eight microphones. The dataset is freely available for download at https://github.com/nttcslab/ToyADMOS2-dataset and https://doi.org/10.5281/zenodo.4580270.

Index Terms— Anomaly detection in sounds, machine operating sounds, product inspection, dataset, domain shift conditions

1. INTRODUCTION

Extensive research efforts have recently been focused on anomaly detection. An anomaly detection task is designed to detect anomaly states by learning only normal condition data. Microphones have been used as sensors to detect anomalies, referred to as anomaly sound detection (ASD) or acoustic condition monitoring [1–9]. This task setting is different from other sound event detection tasks such as gunshot detection [10].

In general, it is very difficult or almost impossible to collect massive anomaly data. The ToyADMOS [11] and MIMII [12] datasets were the first to be utilized for evaluating anomaly detection systems using sound. These datasets enable us to compare the performance of different systems. In 2020, a number of systems from various research organizations in academia and industry were submitted to DCASE 2020 Challenge Task 2: Unsupervised detection of anomalous sounds for machine condition monitoring [13]. The submitted systems performed quite well on the task, thus demonstrating the great potential of applying deep learning-based systems for unsupervised anomaly detection tasks [14–19].

However, the ASD scenario given in the DCASE 2020 Challenge was not realistic compared to actual use cases. The task setting was too basic, and the task requirements were much easier than they would be for typical ones in practical applications, where (for

example) the same machine type but different models are used at different operating speeds, and the conditions are not given as training data. Several independent research groups have tackled tasks related to domain shift or domain adaptation [20–25], but few open datasets that could serve this need have been made available. While the previous ToyADMOS dataset has some data variations that can be used for testing domain-shift conditions, we would like to have more variations on the test configuration.

When evaluating the performance of ASD systems, the statistical characteristics of anomalous sound samples should be different from those of normal samples. However, if the difference is too significant, the anomaly detection task might not be difficult enough to properly evaluate the system performance. One way of controlling the difficultly of the test configuration is to adjust the signal-to-noise ratio (SNR) of the added noise level, but noise reduction techniques, such as the ones in [15, 26], can be used to mitigate the difficulty of the task. Therefore, there should be a way to control the difficulty, without relying on the SNR.

The difficulty of the task under domain shifts can be controlled by changing the statistical difference among normal samples across domains and/or the statistical difference between normal and anomaly sound samples within a domain. In designing challenging test conditions, it is preferred to strike an appropriate balance between these two approaches.

To address the application scenarios discussed above, we provide a new ADMOS dataset called ToyADMOS2. The ToyADMOS2 dataset adds more variations on condition arrangements dedicated to domain shifts. As we did for the previous ToyADMOS dataset, we collected normal and anomalous operating sounds of miniature machines by deliberately damaging their components. The ToyADMOS2 dataset has the following characteristics:

- Designed for two ADMOS tasks: product inspection (toy car) and fault diagnosis for a moving machine (toy train).
- Provides controlled domain-shift conditions on machine models, parts configurations, operating speeds, microphone models and arrangements, and environmental noise.
- Enables control of depth of damage in anomaly samples to provide choices on significance levels of statistical differences between normal and anomalous samples.

Note that the ToyADMOS2 dataset can be used alongside the previous ToyADMOS dataset to provide a larger variety of test conditions.

The proposed ToyADMOS2 dataset is freely available for download at https://github.com/nttcslab/ToyADMOS2-dataset and https://doi.org/10.5281/zenodo.4580270. The license and explanation of some of its uses are also available at those links.
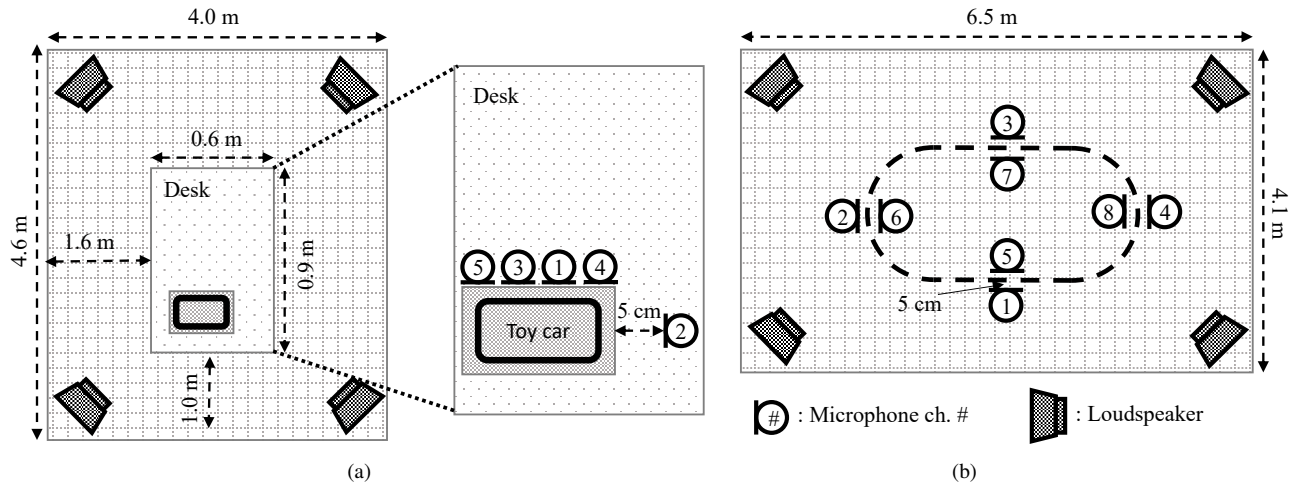
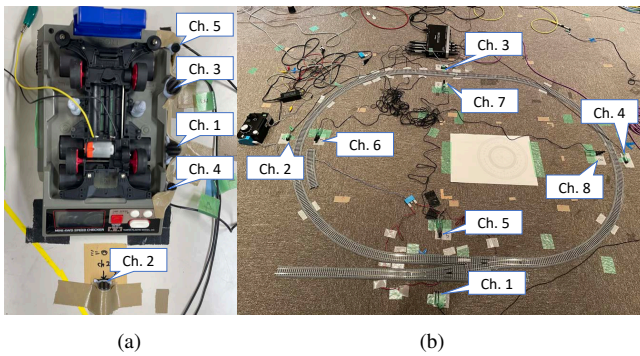Figure 1: Recording-room layouts and microphone arrangements.
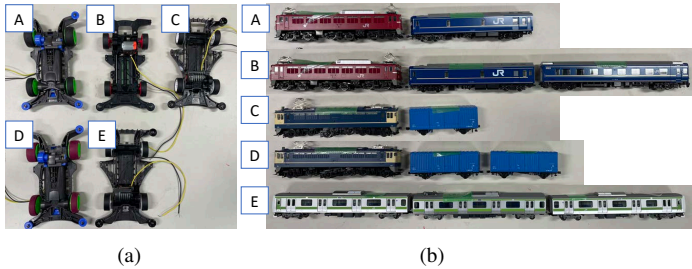

Figure 2: Images of microphone arrangements.

Figure 3: Images of toy-model configurations.

2. DATASET OVERVIEW

The ToyADMOS2 dataset consists of two sub-datasets for two types of ADMOS task. Each sub-dataset has three domain shift conditions. A different type of toy, namely a toy car and a toy train, is used for each task.

Toy car: Designed for the product-inspection task. The toy car runs on an inspection device. Sound data were collected with five microphones arranged close to the inspection device, as shown in Figs. 1(a) and 2(a). The setting is similar to the toy car sub-dataset in ToyADMOS, but the machine models, parts configurations, operating speed settings, and microphone arrangements are different. The toy-car models are shown in Fig. 3(a).

Toy train: Designed for fault diagnosis of a moving machine. The toy train runs on a railway track. Sound data were collected

with eight microphones surrounding the track, as shown in Figs. 1(b) and 2(b). The setting is similar to that of the toy train sub-dataset of the ToyADMOS dataset, but the machine models, cargo arrangements, operating speed settings, and microphone arrangements are different. The toy-train models are shown in Fig. 3(b).

To collect various normal and anomalous sounds depending on individual differences, operating sounds were recorded using several models with different parts configurations.

Normal sound: Operating sound when the target machine operates normally in accordance with its specifications.

Anomalous sound: Operating sound when the target machine was forced to operate anomalously by deliberately damaging its components or adding extraneous objects. The anomaly level was controlled by changing the depth of the damage deliberately made on the parts.

Environmental noise: Environmental noise for simulating a factory or other locations such as a roadside. Noise samples were collected at several locations with multiple microphones. These sounds were emitted from four loudspeakers positioned at the corners of each recording room and recorded with the same microphone configurations used for recording the normal and anomalous sound samples. In addition to newly recorded noise samples, noises from the ToyADMOS dataset were also used.

We utilized omnidirectional dynamic (SURE SM11-CN) and condenser (TOMOCA EM-700) microphones to collect these sounds. All sounds were stored as multiple WAV files.

In ToyADMOS2, several application scenarios representing domain-shift conditions were designed. Examples of possible domain-shift scenarios are as follows:

Model and parts configuration shift: The normal and anomaly data were recorded using several different machine models and parts configurations.

Operating speed shift: Operating speed was varied. Normal and anomaly data were recorded under several different speed levels.

Microphone and environmental noise shift: Different environmental noise was recorded at different positions using different types of microphone.

3. DETAILS OF SUB-DATASETS

3.1. Toy-car sub-dataset

We assumed a product inspection task and took up the task of detecting anomalous sounds from the running sound of a toy car on an inspection device, as shown in Fig. 2(a). The miniature car machine was a toy car called "mini 4WD", which has four wheels and is driven by a small motor through gears and a shaft. A stabilized power supply was connected to the motor, and running sounds on an inspection device were recorded with five microphones. There were five machine configurations: A, B, C, D, and E, as shown in Fig. 3(a). In each configuration, different machine chassis models and parts were used.

To control the operating speed, five voltage levels, 2.8, 3.1, 3.4, 3.7, and 4.0 V, were provided through the stabilized power supply. Each recorded WAV file contains a normal or anomalous sound of a toy car. The duration of the normal and anomalous sound samples is 12 s. This results in over 35 k samples times five channels; in total over 177 k sound samples. Anomalous sounds were generated by deliberately damaging the shaft, gears, or tires. Three damage depth levels were provided for each part, as listed in Table 1. In total, over 8 k samples of anomalous sounds were recorded with various combinations (300 patterns) including three different depths of damage.

For recording the environmental noise, four types of noise files were played through the four loudspeakers and recorded with the same microphone setting shown in Figs. 1(a) and 2(a), and normal and anomalous sounds were recorded. Three dynamic and two condenser microphones were used. Some of the noise files were newly recorded; others were taken from ToyADMOS.

3.2. Toy-train sub-dataset

We assumed the use of fault diagnosis of a moving machine task of anomalous sounds from the running sound of a toy train. We used an HO-scale model train which is a precisely detailed miniature, as shown in Fig. 3(b). Sound data were collected with eight microphones positioned inside and outside the railway track perimeter (four microphones each). The microphones and loudspeakers in the recording room were arranged as shown in Figs. 1(b) and 2(b). Dynamic microphones were used as the four outer microphones (channels 1 to 4), and condenser microphones were used as the four inner microphones (channels 5 to 8). Domain shift conditions in each task configuration are listed in Table 1. There are five machine configurations (A, B, C, D, and E).

To control the operating speed, five levels of speed control (5, 6, 7, 8, and 9) were used. Each recorded WAV file contains normal or anomalous sounds of a toy train. The duration of the normal and anomalous sound samples are 12 s each. This results in over 35 k samples times two mic types; in total, over 71 k sound samples. Anomalous sounds were generated by deliberately damaging the carriage and railway track. Three depth levels of the damage were provided to each part, as listed in Table 1. In total, over 8 k samples of anomalous sounds were recorded with their various combinations (300 patterns).

For recording the environmental noise, four types of noise files were played through the four loudspeakers and recorded with the same microphone setting (shown in Figs. 1(b) and 2(b)) that the normal and anomalous sounds were recorded with. Four dynamic microphones (channels 1 to 4) and four condenser microphones (channels 5 to 8) were used. Some of the noise files were newly recorded; others were taken from ToyADMOS.

Table 1: Anomaly conditions of sub-datasets.

Toy car		Toy train	
Part	Condition	Part	Condition
Shaft	- Bent	Carriage	- Flat tire
Gears	- Deformed		- Broken shaft
	- Melted	Railway track	- Disjointed
Wheels	- Damaged	(straight, curve)	- Obstructing stone

Table 2: Variation settings of sub-datasets.

	Toy car	Toy train
Model variations	Five	Five
Speed levels	Five	Five
Mic. type and	Dynamic: 1–3	Dynamic: 1–4
channel config.	Condenser: 4, 5	Condenser: 5–8
Noise type	Four recordings	
Normal samples	1,094 samples×5 models×5 speed levels	
(Total hours)	(91 hours × 5 ch)	(91 hours×2 ch-sets)
Anomaly samples	324 samples×5 models×5 speed levels	
(Total hours)	(27 hours×5 ch)	(27 hours×2 ch-sets)
Noise samples	24 hours per a channel	

Table 3: Example domain-shift task configurations.

Domain	Configuration		Training data
Source domain	Car/Train model	B	1,500 normal samples
	Speed level	1, 3	
	Mic. type	Dynamic	
	Noise	N1	
Target domain 1: Model and parts shift	**Car/Train model**	**D**	Four normal samples
	Speed level	1, 3	
	Mic. type	Dynamic	
	Noise	N1	
Target domain 2: Operating speed shift	Car/Train model	B	Four normal samples
	Speed level	**2, 5**	
	Mic. type	Dynamic	
	Noise	N1	
Target domain 3: Mic. and noise shift	Car/Train model	B	Four normal samples
	Speed level	1, 3	
	Mic. type	**Condenser**	
	Noise	**N2**	
Target domain 4: All of above	**Car/Train model**	**D**	Four normal samples
	Speed level	**2, 5**	
	Mic. type	**Condenser**	
	Noise	**N2**	

Test data for each domain consists of 100 normal and 100 anomaly samples.

Further details of the ToyADMOS2 dataset are available at https://github.com/nttcslab/ToyADMOS2-dataset.

4. SAMPLE DOMAIN-SHIFT TASK SETTINGS AND BENCHMARK

To demonstrate how to use the ToyADMOS2 dataset, we designed examples of task configurations for testing ASD systems under various domain-shift conditions. The domain-shift configurations are listed in Table 3.

We assume a typical application scenario with a task setting similar to that in [21]. A relatively large number of normal samples

Table 4: Benchmark AUC results of DCASE 2020 Challenge Task 2 baseline under domain-shift conditions.

		Toy car				Toy train			
	Damage level	Clean	6 dB	0 dB	–6 dB	Clean	6 dB	0 dB	–6 dB
Source domain	high	**0.99**	0.99	0.96	0.85	**1.00**	0.86	0.66	0.61
	mid	**0.99**	0.95	0.94	0.80	**0.98**	0.66	0.57	0.56
	low	**1.00**	0.96	0.95	0.80	**0.83**	0.57	0.59	0.49
	avg.	**0.99**	0.97	0.95	0.82	**0.94**	0.70	0.61	0.55
Target domain 1: Model and parts shift	high	0.37	0.29	0.34	0.38	0.82	0.75	0.71	0.62
	mid	0.92	0.55	0.58	0.59	0.71	0.61	0.61	0.53
	low	0.80	0.69	0.68	0.69	0.57	0.56	0.58	0.58
	avg.	0.70	0.51	0.53	0.55	0.70	0.64	0.63	0.57
Target domain 2: Operating speed shift	high	0.77	0.78	0.83	0.76	0.71	0.65	0.62	0.57
	mid	0.61	0.63	0.65	0.66	0.65	0.61	0.51	0.52
	low	0.64	0.72	0.67	0.69	0.52	0.52	0.48	0.44
	avg.	0.68	0.71	0.71	0.70	0.63	0.59	0.54	0.51
Target domain 3: Mic. type and noise shift	high	0.55	0.53	0.45	0.52	0.72	0.62	0.57	0.57
	mid	0.39	0.55	0.47	0.48	0.66	0.55	0.54	0.52
	low	0.45	0.57	0.56	0.52	0.65	0.52	0.52	0.52
	avg.	0.46	0.55	0.49	0.50	0.68	0.56	0.54	0.54
Target domain 4: All above shifts	high	0.55	0.71	0.59	0.65	0.60	0.57	0.53	0.49
	mid	0.44	0.64	0.54	0.56	0.51	0.45	0.44	0.45
	low	0.52	0.71	0.55	0.65	0.51	0.45	0.42	0.44
	avg.	0.50	0.69	0.56	0.62	0.54	0.49	0.46	0.46

recorded in a source domain (e.g., 1,500 samples) are available as a training dataset; however, a very limited number of normal samples recorded in the target domains (four samples each) are given for training. As shown in Table 3, there are four example domain-shift configurations: **Target domain 1**, Model and parts shift; **Target domain 2**, Operating speed shift; **Target domain 3**, Mic. type and noise shift; and **Target domain 4**, A mixture of shifts 1 to 3.

For the source domain training dataset, 1,500 normal samples of Toy-car model B and Toy-train model B were used. For Toy car, normal sample data recorded with one dynamic microphone (ch 1) were used. For Toy train, normal sample data recorded by dynamic microphones (ch 1 to 4) were used. In these normal samples, operating speeds were set to levels 1 and 3. For target domains 1 to 4, only four normal samples were given as training data for each of the target domains. Environmental noises N1 and N2 recorded using microphones at the same positions as for the normal samples were used. SNRs calculated by

$$SNR = 20\log_{10}\left\{\frac{1}{J}\sum rms(S_j)\right\}/rms(N_k), \quad (1)$$

where S_j is a recorded machine-operating sound sample and N_k is a recorded environmental noise sample with the same duration as S_j, were set to $+\infty$ dB (clean), 6 dB, 0 dB, and –6 dB. After being mixed with the noise, all sound samples were down-sampled at a sampling rate of 16 kHz. Detailed conditions of each target domain are shown in Table 3.

To simplify the evaluation system example, a baseline system of DCASE 2020 Challenge Task 2 [13]—a simple unsupervised-ASD one with an auto encoder as a normal model—was tested under the task setting described above. All the training data—1,500 normal samples of the source domain, and other normal samples from the target domains (four samples each)— were merged to formulate a training dataset that contains 1,516 normal samples. The baseline system was trained for each SNR condition of Toy car and Toy train.

The trained baseline systems were tested using 200 unknown test samples that were not used for training. We calculated anomaly scores on each time frame for all the test WAV files as described

in [13]. Calculated area under curve (AUC) results are shown in Table 4. The results in bold indicate that when the damage level was higher, anomaly detection was easier for the source domain/clean for Toy train. For Toy car, the AUC results were around 0.99 for all the levels. Distribution of the anomaly scores for the source domain/clean samples are shown in Fig. 4. The target domain results show that the DCASE baseline failed to distinguish normal and anomaly data under domain-shifted conditions even with a high damage level.
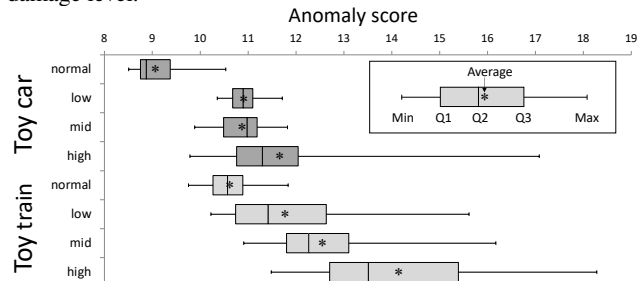

Figure 4: Anomaly scores of source domain/clean samples.

5. CONCLUSION

This paper proposed a new large-scale dataset called "ToyAD-MOS2" for ADMOS. The ToyADMOS2 dataset is designed for evaluating systems under domain-shift conditions and is available for free download. It consists of two sub-datasets for machine-condition inspection: fault diagnosis of machines with geometrically fixed tasks and fault diagnosis of machines with moving tasks. Domain shifts are represented by introducing several differences in operating conditions, such as the use of the same machine type but with different machine models and parts configurations, operating speeds, microphone arrangements, etc. Each sub-dataset contains over 27 k samples of normal machine-operating sounds and over 8 k samples of anomalous sounds recorded at a 48-kHz sampling rate. A subset of the ToyADMOS2 dataset was used in the DCASE 2021 Challenge Task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain-shifted conditions.

6. REFERENCES

[1] 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, 2019.

[2] E. Rushe and B. M. Namee, "Anomaly detection in raw audio using deep autoregressive networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3597–3601.

[3] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, "Optimizing acoustic feature extractor for anomalous sound detection based on neyman-pearson lemma," in *25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 698–702.

[4] Y. Kawachi, Y. Koizumi, and N. Harada, "Complementary set variational autoencoder for supervised anomaly detection," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2366–2370.

[5] Y. Koizumi, S. Murata, N. Harada, S. Saito, and H. Uematsu, "Sniper: Few-shot learning for anomaly detection to minimize false-negative rate with ensured true-positive rate," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 915–919.

[6] Y. Kawachi, Y. Koizumi, S. Murata, and N. Harada, "A two-class hyper-spherical autoencoder for supervised anomaly detection," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3047–3051.

[7] 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 *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019, pp. 6–10.

[8] Y. Koizumi, M. Yasuda, S. Murata, S. Saito, H. Uematsu, and N. Harada, "Spidernet: Attention network for one-shot anomaly detection in sounds," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 281–285.

[9] S. Heincke, A. K. Kalan, O. J. Wagner, R. Mundry, H. Lukashevich, and H. S. Kühl, "Assessing the performance of a semi-automated acoustic monitoring system for primates," *Methods in Ecology and Evolution*, pp. 753–763, 2015.

[10] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, 2007, pp. 21–26.

[11] 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.

[12] 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.

[13] 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)*, 2020, pp. 81–85.

[14] R. Giri, S. V. Tenneti, K. Helwani, F. Cheng, U. Isik, and A. Krishnaswamy, "Unsupervised anomalous sound detection using self-supervised classification and group masked autoencoder for density estimation," DCASE2020 Challenge, Tech. Rep., July 2020.

[15] P. Daniluk, M. Gozdziewski, S. Kapka, and M. Kosmider, "Ensemble of auto-encoder based systems for anomaly detection," DCASE2020 Challenge, Tech. Rep., July 2020.

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

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

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

[19] Q. Zhou, "Arcface based sound mobilenets for dcase 2020 task 2," DCASE2020 Challenge, Tech. Rep., July 2020.

[20] M. Yamaguchi, Y. Koizumi, and N. Harada, "Adaflow: Domain-adaptive density estimator with application to anomaly detection and unpaired cross-domain translation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2019*, 2019, pp. 3647–3651.

[21] Q. Wang, G. Michau, and O. Fink, "Domain adaptive transfer learning for fault diagnosis," in *2019 Prognostics and System Health Management Conference (PHM-Paris)*, 2019, pp. 279–285.

[22] G. Michau and O. Fink, "Domain adaptation for one-class classification: Monitoring the health of critical systems under limited information," *International journal of prognostics and health management.*, vol. 10, 2019.

[23] A. Kumagai, T. Iwata, and Y. Fujiwara, "Transfer anomaly detection by inferring latent domain representations," in *NeurIPS*, 2019.

[24] Z. Yang, I. S. Bozchalooi, and E. F. Darve, "Anomaly detection with domain adaptation," *ArXiv*, vol. abs/2006.03689, 2020.

[25] V. Vincent, M. Wannes, and D. Jesse, "Transfer learning for anomaly detection through localized and unsupervised instance selection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 6054–6061, Apr. 2020.

[26] Y. Kawaguchi, R. Tanabe, T. Endo, K. Ichige, and K. Hamada, "Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 865–869.