TOYADMOS2+: NEW TOYADMOS DATA AND BENCHMARK RESULTS OF THE FIRST-SHOT ANOMALOUS SOUND DETECTION BASELINE

Noboru Harada, Daisuke Niizumi, Yasunori Ohishi, Daiki Takeuchi, Masahiro Yasuda

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

ABSTRACT

This paper introduces the newly recorded ToyADMOS dataset for the DCASE 2023 Challenge Task 2, First-shot anomalous sound detection for machine condition monitoring (DCASE2023T2). New machine types, such as ToyDrone, ToyNscale, Vacuum, and ToyTank, were newly recorded as a part of the Additional training and Evaluation datasets. This paper also shows benchmark results of the First-shot baseline implementation (with simple autoencoder and selective Mahalanobis modes) on the DCASE2023T2 Evaluation dataset and the previous DCASE Challenge Task 2 datasets in 2020, 2021, and 2022, compared with the baselines of those years.

Index Terms— DCASE 2023 Challenge Task 2, First-Shot Anomalous sound detection, ToyADMOS dataset

1. INTRODUCTION

In recent years, exhaustive research has been done on anomalous sound detection (ASD) for machine condition monitoring. Several challenge tasks related to ASD were organized in the Detection and Classification of Acoustic Scenes and Events (DCASE).

The first ASD challenge was DCASE 2020 Challenge Task 2: Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring (DCASE2020T2) [1]. The task setting required systems to have only the normal samples of machine operating sound for training. No anomalous sound sample was available for training because getting enough number of anomaly samples is extremely difficult in real application scenarios. The organizers of DCASE2020T2 had to create specific datasets for the task by intentionally adding damage to machines, e.g., toys. The datasets are called ToyADMOS and MIMII dataset [2, 3]. Since then, the ASD challenge task has been extended to represent more realistic application scenarios, such as DCASE 2021 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring under Domain Shifted Conditions (DCASE2021T2) [4], and DCASE 2022 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques (DCASE2022T2) [5]. ToyADMOS2, MIMII DUE, and MIMII DG datasets had been developed [6, 7, 8].

In DCASE 2023 Challenge Task 2: "First-shot Anomalous Sound Detection for Machine Condition Monitoring" (DCASE2023T2) [9], new datasets and a baseline implementation complying with the First-shot requirements were introduced [10, 6, 8, 11]. First-shot means the system can use only the given training data for the target. The First-shot ASD is characterized as follows:

- No use of data from different machine instances (not given)
- No hyperparameter tuning nor tool ensemble enabled for dedicated machine type (by analyzing ground truth results with both normal and anomaly samples)

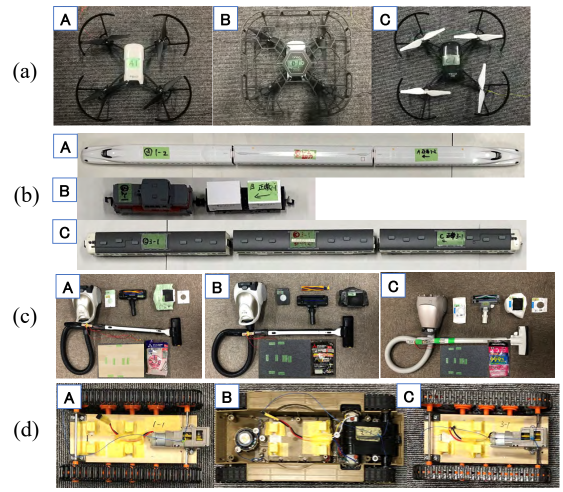

Figure 1: Images of toy-model configurations A, B, and C for (a) ToyDrone, (b) ToyNscale, (c) Vacuum, and (d) ToyTank.

This paper describes the newly added ToyADMOS data for the Additional training and Evaluation datasets of the DCASE2023T2. In addition, this paper also shows benchmark results of the First-shot baseline implementation [12] on the previous DCASE Challenge Task 2 datasets in 2020, 2021, 2022, and 2023, compared with the baselines of those years [13, 14, 15, 16, 17]. The source code and data are available at GitHub [12] and Zenodo [9, 18, 19].

2. TOYADMOS2+: ADDITIONAL DATA FOR THE DCASE2023 CHALLENGE TASK 2

To provide the First-shot training data for DCASE2023T2, the following four machine types, (a) ToyDrone, (b) ToyNscale, (c) Vacuum, and (d) ToyTank, were newly recorded. Each machine type has three model configurations (A, B, C) shown in Fig. 1. The machine operating sounds were recorded with the room layout and the microphone settings shown in Figs. 2 and 3.

ToyDrone: The ToyDrone flies in a guide frame as shown in Figs. 2(a) and 3(a). There are three flying patterns.
ToyNscale: The N-scale toy train runs on a railway track. Sound data were collected with eight microphones surrounding the track, as shown in Figs. 2(b) and 3(b).
Vacuum: The Vacuum is set on a guide frame in Figs. 2(c) and 3(c). A floor mat or a wooden floor plate was used.
ToyTank: The ToyTank runs in a guide frame as shown in Figs. 2(d) and 3(d).

To generate anomaly samples, some of the parts were intentionally damaged. Anomaly conditions for each machine type are shown in Table 1.

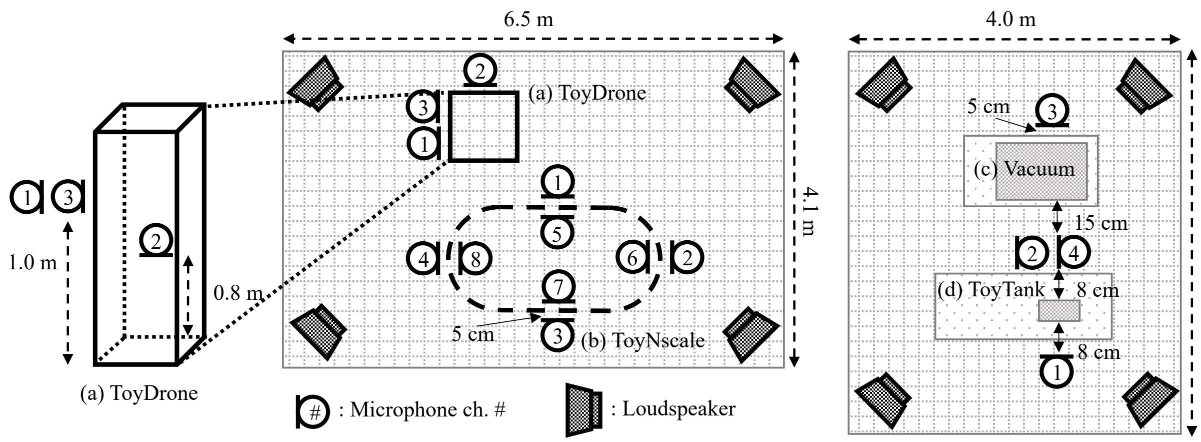

Figure 2: Recording-room layouts and microphone arrangements, (a) ToyDrone, (b) ToyNscale, (c) Vacuum, and (d) ToyTank.

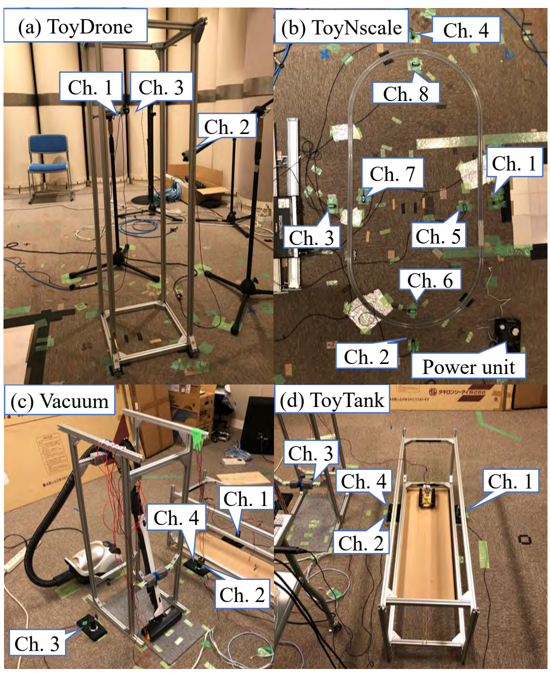

Figure 3: Images of microphone arrangements (a) ToyDrone, (b) ToyNscale, (c) Vacuum, and (d) ToyTank.

Table 2 shows the recording setting. All the operating sound and noise samples were recorded with 48 kHz sampling, 24-bit for each channel, and then downsampled to 16 kHz, 24-bit, monaural. Sample duration varies from 6 sec to 18 sec, depending on the machine type, as shown in Table 2. Domain shift conditions were controlled by changing machine instances (ID), operating speed, mic position, and mixed background noise samples. Table 3 shows the domain shift conditions of source and target domains.

For training data (Additional training dataset), there are 1000 normal samples given for training, where 990 samples are from the source domain, and 10 samples are from the target domain. For evaluation data (Evaluation dataset), 50 normal and 50 anomaly samples are from each source and target domain. In total, there are 200 samples for each machine type. In total, 960 min of data was prepared. The data is available at the Zenodo links [18, 19] under the Creative Commons Attribution 4.0 International Public License [20].

Table 1: Anomaly conditions for each machine type.

(a) **ToyDrone**		(b) **ToyNscale**	
Part	Condition	Part	Condition
Propeller	- Cut one side - Cut two sides	Carriage	- Flat tire - Broken shaft
Guard	- Guard missing	Railway	- Disjointed
Unbalance	- Offset weight	track	- Obstructing stone

(c) **Vacuum**		(d) **ToyTank**	
Part	Condition	Part	Condition
Air filter	- No filter	Trackbelt	- Damaged belt
Nozzle	- Thread jam		- Foreign object
Dust bag	- Bag full	Trackbelt wheel	- Wheel lock
	- No dust bag		- Missing wheel

Table 2: Recording conditions for DCASE2023T2 Eval. dataset.

	ToyDrone	ToyNscale	Vacuum	ToyTank
Model variations	A, B, C	A, B, C	A, B, C	A, B, C
Speed levels	Three patterns	Five	Lb, M, Mb, H^{*1}	Five
Mic. config	Ch. 1 - 3	Ch. 1 - 8	Ch. 1 - 4	Ch. 1 - 4
Noise type^{*2}	N1	N1	N2	N2
Sample duration	18 sec	6 sec	15 sec	8 sec

^{*1}*L, M, H* mean Lo, Middle, and High power, and *Lb, Mb* means with a brush on.
^{*2}*N1*: windy outside noise and *N2*: room air conditioning noise.

Table 3: Domain shift settings for DCASE2023T2 Eval. dataset.

Source domain	ToyDrone	ToyNscale	Vacuum	ToyTank
Machine ID	A, B	A, B	A, B	A, B
Speed	2, 3	2, 3	Lb, M, Mb, H^{*1}	3, 4
Mic.	Dynamic 1	Dynamic 1 - 4	Dynamic 1	Dynamic 1
Noise^{*2}	N1	N1	N2	N2
Target domain	**ToyDrone**	**ToyNscale**	**Vacuum**	**ToyTank**
Machine ID	C	C	C	C
Speed	1	1, 5	Mb, H^{*1}	1, 5
Mic.	Dynamic 2	Dynamic 5 - 8	Dynamic 2	Dynamic 1
Noise^{*2}	N1	N1	N2	N2

^{*1}*L, M, H* mean Lo, Middle, and High power, and *Lb, Mb* means with a brush on.
^{*2}*N1*: windy outside noise and *N2*: room air conditioning noise.

3. BENCHMARK RESULTS WITH THE DCASE2023T2 FIRST-SHOT BASELINE

3.1. The DCASE2023T2 First-shot baseline

DCASE2023 Challenge Task 2 baseline has the following two operating modes. For the details, see Sec. 3.2 and [11].

Table 4: AUC results of the DCASE 2023 Challenge Task 2 Evaluation dataset.

System	metric	hmean*1	amean*1	**ToyDrone**	**ToyNscale**	**ToyTank**	**Vacuum**	**Bandsaw**	**Grinder**	**Shaker**
First-shot compliant	AUC (source)	0.7421	0.7484	0.8121	0.7650	0.7196	0.8815	0.6710	0.7039	0.6859
simple Autoencoder	AUC (target)	0.5357	0.5436	0.5375	0.4891	0.6655	0.4488	0.5264	0.5590	0.5790
mode (FS-AE) [12]	pAUC (src & tgt)	0.5551	0.5575	0.5316	0.5223	0.5998	0.6040	0.5098	0.5858	0.5489
	TOTAL score	**0.5981**	**0.6165**							
Selective	AUC (source)	0.7877	0.7934	0.8495	0.6684	0.8085	0.8031	0.8328	0.7396	0.8519
Mahalanobis	AUC (target)	0.5377	0.5707	0.4218	0.4377	0.4879	0.8727	0.6014	0.5169	0.6566
AE mode [12]	pAUC (src & tgt)	0.5722	0.5779	0.5140	0.5107	0.5470	0.6795	0.5684	0.6033	0.6227
	TOTAL score	**0.6151**	**0.6474**							
DCASE2022T2	AUC (source)	0.7732	0.7835	0.9018	0.7705	0.8325	0.9047	0.6708	0.7051	0.6993
baseline AE [16]	AUC (target)	0.3644	0.4049	0.2611	0.3940	0.3430	0.2397	0.4554	0.5494	0.5920
	pAUC (src & tgt)	0.5323	0.5346	0.4970	0.5112	0.5807	0.5116	0.5023	0.5789	0.5605
	TOTAL score	**0.5071**	**0.5744**							
cf. DCASE2023T2 Top 1	AUC (source)	0.8313	0.8377	0.8026	0.9042	0.8480	0.9690	0.7655	0.7367	0.8376
Jie_IESEFPT_task2_2	AUC (target)	0.6008	0.6444	0.4544	0.8768	0.4682	0.9548	0.5749	0.6082	0.5734
[21] (operating condition	pAUC (src & tgt)	0.6203	0.6399	0.5158	0.7774	0.6153	0.8532	0.5335	0.6245	0.5597
label classification)	**TOTAL score**	**0.6697**	**0.7073**							

*1 *hmean* denotes harmonic mean, and *amean* denotes arithmetic mean.

First-shot-compliant simple Autoencoder mode (FS-AE): This is a simple autoencoder. For training, the model parameter θ of the AE is trained to minimize the mean square error (MSE) between a normal input sample x^- and its reconstruction $\hat{x}^-$ using

$$Loss \quad = MSE(x^-, \hat{x}^-), \tag{1}$$
$$where \quad \hat{x}^- = Dec_\theta(Enc_\theta(x^-)). \tag{2}$$

For the testing phase, the anomaly score A_θ is calculated with the reconstruction error of the given query sample x using

$$Anomaly\ Score\ \ A_\theta \quad = MSE(x, \hat{x}), \tag{3}$$
$$where\ \hat{x} \quad = Dec_\theta(Enc_\theta(x)). \tag{4}$$

When the anomaly score exceeds the pre-set threshold, the sample is detected as an anomaly sample.

Selective Mahalanobis Autoencoder mode: The anomaly score A_θ is calculated using the covariance matrixes Σ_s^{-1} and Σ_t^{-1} of distance between normal samples x^- and its reconstruction $\hat{x}^-$ for the source and target domains with:

$$Anomaly\ Score\ \ A_\theta \quad = min\{D_s(x, \hat{x}), D_t(x, \hat{x})\}, \tag{5}$$
$$where\ \ D_s(\cdot) \quad = Mahalanobis(x, \hat{x}, \Sigma_s^{-1}), \tag{6}$$
$$D_t(\cdot) \quad = Mahalanobis(x, \hat{x}, \Sigma_t^{-1}). \tag{7}$$

3.2. Experimental setup and evaluation criterion

For the First-shot compliant baseline, the model hyperparameters were set to the values described in [11].

The frame size for STFT was 64 ms with 50 % hop size translated into 128 frequency bands Log-mel energies. Five consecutive frames were concatenated to formulate 640 dimensions (128×5) as input to the system. In the autoencoder model, there were three layers of 128 dimensions linear, Batch normalization, and Activation with ReLU each, in encoder and decoder. The bottleneck layer had eight dimensions. The number of epochs for training was 100. The batch size was 256, and the Adam optimizer used a 0.001 learning ratio.

The performances of the two operating modes of the DCASE2023T2 baseline [11, 12] were compared with the previous baselines, such as DCASE2020T2 baseline AE [13], DCASE2021T2 baseline AE [14], DCASE2021T2 baseline MobileNetV2AE [15], DCASE2022T2 baseline AE [16], and DCASE2022T2 MobileNetV2 [17]. The Evaluation datasets and baseline systems dedicated to the DCASE Challenges were used. For those baseline systems, hyperparameters were set to the ones used in the corresponding previous DCASE Challenges [1, 4, 5, 10, 9, 22, 23, 24]. The total scores Ω for evaluating the systems are calculated based on Area Under the Receiver Operating Characteristic (ROC) curve (AUC) and partial AUC (pAUC) with a harmonic mean ($hmean$) and an arithmetic mean ($amean$) of AUC and pAUC. All the results were the averaged score of systems trained with three different random seeds except for Jie_IESEFPT_task2_2 [21] that were copied from the official score.

3.3. Experimental results

Experimental results of the DCASE2023T2 First-shot baseline on the DCASE2023T2 Evaluation datasets are shown in Table 4. Some other results of the DCASE2023T2 baseline compared with the previous baseline systems on the Evaluation datasets of DCASE 2020, 2021, and 2022 Challenge tasks are shown in Tables 5, 6, and 7, respectively. The DCASE2023T2 baseline with the selective Mahalanobis AE mode performed better than others in the DCASE2021T2, DCASE2022T2, and DCASE2023T2. For DCASE2020T2, the DCASE2023T2 baseline with the FS-AE mode performs better because of no domain shift. The DCASE2023T2 baseline can be a performance benchmark for the tasks.

4. CONCLUSION

This paper introduces the newly recorded ToyADMOS dataset for the DCASE 2023 Challenge Task 2, First-shot anomalous sound detection for machine condition monitoring. New machine types, such as ToyDrone, ToyNscale, Vacuum, and ToyTank, are newly recorded as a part of the Additional training and Evaluation datasets. This paper also shows benchmark results of the First-shot baseline implementation (with simple autoencoder and selective Mahalanobis modes) on the previous DCASE Challenge Task 2 datasets in 2020, 2021, 2022, and 2023, compared with the baselines of those years. ToyADMOS2+ dataset (DCASE 2023 Challenge Task 2 Additional Training Dataset and Evaluation Dataset) is available at [9, 18, 19] with the Creative Commons Attribution 4.0 International Public License [20]. The updated source code of the DCASE2023T2 baseline supporting all the previous task settings is available at [12].

5. ACKNOWLEDGMENT

The authors thank Mr. Hideaki Sanada, Mr. Akito Karasawa, Mr. Keisuke Hasegawa, and Mr. Yuri Musashijima for their hard work implementing the systems and conducting experiments.

Table 5: AUC results of the DCASE 2020 Challenge Task 2 Evaluation dataset.

System	metric	hmean^{*1}	amean^{*1}	**ToyCar**	**ToyConvayor**	**fan**	**pump**	**slider**	**valve**
First-shot compliant	AUC	0.7817	0.8003	0.8399	0.7859	0.8601	0.8618	0.7975	0.6567
simple Autoencoder	pAUC	0.6187	0.6342	0.6936	0.6651	0.6903	0.6499	0.5872	0.5191
mode (FS-AE) [12]	**TOTAL score**	**0.6907**	**0.7173**						
Selective	AUC	0.7434	0.7845	0.8101	0.7727	0.9478	0.9175	0.7280	0.5305
Mahalanobis	pAUC	0.6115	0.6366	0.5892	0.6700	0.7886	0.7333	0.5292	0.5095
AE mode [12]	**TOTAL score**	**0.6710**	**0.7105**						
DCASE2020T2	AUC	0.7774	0.7959	0.7973	0.8822	0.8593	0.8408	0.8148	0.5810
baseline AE [13]	pAUC	0.6189	0.6312	0.6630	0.7075	0.6768	0.6443	0.5869	0.5085
	TOTAL score	**0.6892**	**0.7135**						

^{*1}*hmean* denotes harmonic mean, and *amean* denotes arithmetic mean.

Table 6: AUC results of the DCASE 2021 Challenge Task 2 Evaluation dataset.

System	metric	hmean^{*1}	amean^{*1}	**ToyCar**	**ToyTrain**	**fan**	**gearbox**	**pump**	**slider**	**valve**
First-shot compliant	AUC (source)	0.6421	0.6506	0.7049	0.7009	0.6818	0.6605	0.6336	0.6483	0.5245
simple Autoencoder	AUC (target)	0.5716	0.5817	0.6211	0.5808	0.5850	0.6278	0.5699	0.5592	0.5279
mode (FS-AE) [12]	pAUC (source)	0.5224	0.5244	0.5232	0.5265	0.5225	0.5541	0.5175	0.5166	0.5104
	pAUC (target)	0.5206	0.5253	0.5680	0.5082	0.5081	0.5712	0.5098	0.5189	0.4926
	TOTAL score	**0.5601**	**0.5705**							
Selective	AUC (source)	0.6488	0.6696	0.8438	0.5677	0.7773	0.6581	0.6774	0.6593	0.5033
Mahalanobis	AUC (target)	0.5664	0.5885	0.6560	0.4706	0.6060	0.6333	0.6284	0.5849	0.5401
AE mode [12]	pAUC (source)	0.5418	0.5507	0.6896	0.4937	0.5601	0.5485	0.5408	0.5063	0.5159
	pAUC (target)	0.5279	0.5328	0.6009	0.5093	0.5308	0.5717	0.5067	0.5101	0.5001
	TOTAL score	0.5676	**0.5854**							
DCASE2021T2	AUC (source)	0.6468	0.6556	0.7490	0.7141	0.6624	0.6736	0.6312	0.6444	0.5146
baseline AE [14]	AUC (target)	0.5693	0.5834	0.6232	0.6451	0.5568	0.6330	0.5612	0.5408	0.5236
	pAUC (source)	0.5272	0.5305	0.5409	0.5660	0.5176	0.5516	0.5143	0.5149	0.5082
	pAUC (target)	0.5318	0.5386	0.5692	0.5959	0.4982	0.5807	0.5101	0.5219	0.4945
	TOTAL score	**0.5650**	**0.5770**							
DCASE2021T2	AUC (source)	0.5351	0.6034	0.4279	0.5215	0.7505	0.5620	0.7013	0.7246	0.5360
baseline	AUC (target)	0.5236	0.5892	0.5800	0.3852	0.6396	0.4889	0.7107	0.7517	0.5681
MobileNetV2 [15]	pAUC (source)	0.5569	0.5736	0.5299	0.5312	0.6467	0.5522	0.6342	0.6060	0.5149
	pAUC (target)	0.5617	0.5780	0.6505	0.4921	0.6288	0.4962	0.6236	0.6241	0.5304
	TOTAL score	**0.5431**	0.5860							

^{*1}*hmean* denotes harmonic mean, and *amean* denotes arithmetic mean.

Table 7: AUC results of the DCASE 2022 Challenge Task 2 Evaluation dataset.

System	metric	hmean^{*1}	amean^{*1}	**ToyCar**	**ToyTrain**	**fan**	**gearbox**	**bearing**	**slider**	**valve**
First-shot compliant	AUC (source)	0.6515	0.6803	0.8331	0.4998	0.6562	0.7116	0.7479	0.7594	0.5539
simple Autoencoder	AUC (target)	0.5123	0.5418	0.6537	0.5243	0.3580	0.6026	0.5621	0.5264	0.5657
mode (FS-AE) [12]	pAUC (src & tgt)	0.5344	0.5420	0.6658	0.4973	0.5136	0.5152	0.5474	0.5355	0.5192
	TOTAL score	**0.5599**	**0.5880**							
Selective	AUC (source)	0.6650	0.7138	0.9401	0.5031	0.6820	0.8510	0.7195	0.7637	0.5371
Mahalanobis	AUC (target)	0.5557	0.5999	0.7965	0.5075	0.4055	0.7724	0.6280	0.5595	0.5296
AE mode [12]	pAUC (src & tgt)	0.5623	0.5763	0.7738	0.5067	0.5133	0.6071	0.5763	0.5408	0.5164
	TOTAL score	**0.5903**	**0.6300**							
DCASE2022T2	AUC (source)	0.6478	0.6762	0.7749	0.5973	0.6433	0.7006	0.7281	0.7548	0.5346
baseline AE [16]	AUC (target)	0.4451	0.4771	0.4761	0.3915	0.3200	0.5762	0.5321	0.4931	0.5505
	pAUC (src & tgt)	0.5263	0.5314	0.5939	0.4969	0.5046	0.5204	0.5519	0.5377	0.5146
	TOTAL score	**0.5272**	**0.5616**							
DCASE2022T2	AUC (source)	0.5758	0.6535	0.6457	0.5340	0.6914	0.5996	0.6333	0.7692	0.7014
baseline	AUC (target)	0.4542	0.5248	0.4621	0.5916	0.3842	0.4433	0.5239	0.5052	0.7629
MobileNetV2 [17]	pAUC (src & tgt)	0.5345	0.5413	0.5419	0.5046	0.5354	0.4820	0.5151	0.5498	0.6601
	TOTAL score	**0.5163**	**0.5732**							

^{*1}*hmean* denotes harmonic mean, and *amean* denotes arithmetic mean.

6. REFERENCES

[1] 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)*, November 2020, pp. 81–85. [Online]. Available: http://dcase.community/documents/workshop2020/proceedings/DCASE2020Workshop_Koizumi_3.pdf

[2] 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. [Online]. Available: https://ieeexplore.ieee.org/document/8937164

[3] 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. [Online]. Available: http://dcase.community/documents/workshop2019/proceedings/DCASE2019Workshop_Purohit_21.pdf

[4] 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 *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, Barcelona, Spain, November 2021, pp. 186–190.

[5] 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 *Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, Nancy, France, November 2022.

[6] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, Barcelona, Spain, November 2021, pp. 1–5.

[7] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, "MIMII DUE: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 21–25, 2021.

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

[9] "DCASE 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," 2023, https://dcase.community/challenge2023/task-first-shot-unsupervised-anomalous-sound-detection-for-machine-condition-monitoring.

[10] 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 arXiv e-prints: 2305.07828*, 2023.

[11] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "First-shot anomaly detection for machine condition monitoring: A domain generalization baseline," *In arXiv e-prints: 2303.00455*, 2023.

[12] N. Harada, Y. Musashijima, and D. Niizumi, "dcase2023_task2_baseline_ae," 2023, https://github.com/nttcslab/dcase2023_task2_baseline_ae.

[13] Y. Kawaguchi, "dcase2020_task2_baseline," 2020, https://github.com/y-kawagu/dcase2020_task2_baseline.

[14] ——, "dcase2022_task2_baseline_ae," 2021, https://github.com/y-kawagu/dcase2021_task2_baseline_ae.

[15] ——, "dcase2022_task2_baseline_mobile_net_v2," 2021, https://github.com/y-kawagu/dcase2021_task2_baseline_mobile_net_v2.

[16] K. Dohi, "dcase2022_task2_baseline_ae," 2022, https://github.com/Kota-Dohi/dcase2022_task2_baseline_ae.

[17] ——, "dcase2022_task2_baseline_mobile_net_v2," 2022, https://github.com/Kota-Dohi/dcase2022_task2_baseline_mobile_net_v2.

[18] "DCASE 2023 challenge task 2 additional training dataset," 2023, https://zenodo.org/record/7830345.

[19] "DCASE 2023 challenge task 2 evaluation dataset," 2023, https://zenodo.org/record/7860847.

[20] "Creative commons attribution 4.0 international public license," https://creativecommons.org/licenses/by/4.0/legalcode.

[21] J. Jie, "Anomalous sound detection based on self-supervised learning," DCASE2023 Challenge, Tech. Rep., June 2023.

[22] "DCASE 2020 challenge task 2: Unsupervised detection of anomalous sounds for machine condition monitoring," 2020, https://dcase.community/challenge2020/task-unsupervised-detection-of-anomalous-sounds.

[23] "DCASE 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions," 2021, https://dcase.community/challenge2021/task-unsupervised-detection-of-anomalous-sounds.

[24] "DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," 2022, https://dcase.community/challenge2022/task-unsupervised-anomalous-sound-detection-for-machine-condition-monitoring.