MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Handling Domain Shifts for Anomalous Sound Detection: A Review of DCASE-Related Work

Wilkinghoff, Kevin; Fujimura, Takuya; Imoto, Keisuke; Le Roux, Jonathan; Tan, Zheng-Hua; Toda, Tomoki

TR2025-157 November 01, 2025

Abstract

When detecting anomalous sounds in complex environments, one of the main difficulties is that trained models must be sensitive to subtle differences in monitored target signals, while many practical applications also require them to be insensitive to changes in acoustic domains. Examples of such domain shifts include changing the type of microphone or the location of acoustic sensors, which can have a much stronger impact on the acoustic signal than subtle anomalies themselves. Moreover, users typically aim to train a model only on source domain data, which they may have a relatively large collection of, and they hope that such a trained model will be able to generalize well to an unseen target domain by providing only a minimal number of samples to characterize the acoustic signals in that domain. In this work, we review and discuss recent publications focusing on this domain generalization problem for anomalous sound detection in the context of the DCASE challenges on acoustic machine condition monitoring.

Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE) 2025

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Handling Domain Shifts for Anomalous Sound Detection: A Review of DCASE-Related Work

Kevin Wilkinghoff[1,2], *Takuya Fujimura*[3], *Keisuke Imoto*[4], *Jonathan Le Roux*[5], *Zheng-Hua Tan*[1,2], *Tomoki Toda*[3]

[1] Aalborg University, Denmark, [2] Pioneer Centre for AI, Denmark, [3] Nagoya University, Japan
[4] Kyoto University, Japan, [5] Mitsubishi Electric Research Laboratories, MA, USA

***Abstract*—When detecting anomalous sounds in complex environments, one of the main difficulties is that trained models must be sensitive to subtle differences in monitored target signals, while many practical applications also require them to be insensitive to changes in acoustic domains. Examples of such domain shifts include changing the type of microphone or the location of acoustic sensors, which can have a much stronger impact on the acoustic signal than subtle anomalies themselves. Moreover, users typically aim to train a model only on source domain data, which they may have a relatively large collection of, and they hope that such a trained model will be able to generalize well to an unseen target domain by providing only a minimal number of samples to characterize the acoustic signals in that domain. In this work, we review and discuss recent publications focusing on this domain generalization problem for anomalous sound detection in the context of the DCASE challenges on acoustic machine condition monitoring.**

***Index Terms*—anomalous sound detection, domain shift, domain adaptation, domain generalization, machine condition monitoring**

1. INTRODUCTION

Anomalous sound detection (ASD) has a wide range of applications, including acoustic monitoring of machines [1]–[5], health [6], [7], roads [8], smart home environments [9], or public places [10]. The goal in all these applications is to distinguish between normal and anomalous audio recordings. Usually, ASD systems are trained exclusively with normal data, as anomalous data is often difficult and costly to obtain.

One of the major difficulties that ASD systems need to overcome is how to robustly handle so-called *domain shifts*, changes in the recorded audio signal that are caused by changes in acoustic environments, sensors, or properties of monitored sound sources themselves. Inherently, domain shifts have a strong impact on the audio signal and thus also affect the outcome of ASD systems if no precautions are taken. However, in most applications, this effect is not desirable and should be suppressed. Ideally, trained ASD systems should be completely insensitive to domain shifts, while being very sensitive to modifications of the monitored target signals that indicate the occurrence of application-dependent anomalies. As these differences may be very subtle compared to changes caused by domain shifts, especially in noisy environments with many sound sources, developing ASD systems that perform well in domain-shifted conditions is very challenging.

This paper, which extends a non-peer reviewed manuscript [11], reviews the latest works on how to handle domain shifts for ASD in the context of the annual DCASE challenge [12]. We first describe the challenges posed by domain shifts and the resulting *domain mismatch*, illustrated in Fig. 1. We then describe the two main types of approaches to handling them, *domain adaptation (DA)* and *domain generalization (DG)*, defining the terms, listing publicly available datasets, and reviewing recently published works related to these topics. Last but not least, the limitations of different approaches are discussed and future directions of research are identified.

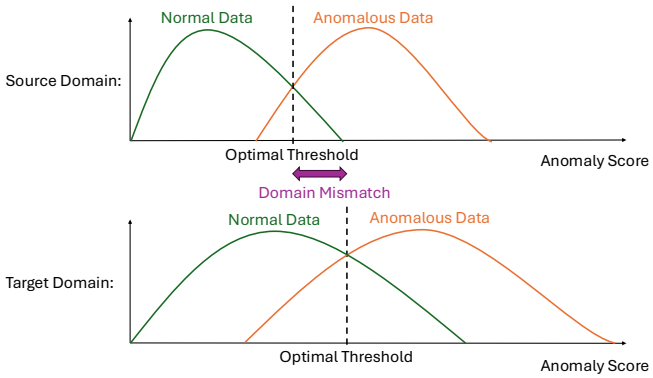

Fig. 1: Illustration of the domain mismatch between the anomaly scores of a source domain and a target domain. Normal and anomalous samples are usually less well separated in the target domain than in the source domain, which decreases domain-independent performance over the performance obtained for the source domain alone. Furthermore, the optimal decision thresholds for separating the scores belonging to the normal and anomalous data of different data domains differ substantially, which significantly decreases performance when using only a single threshold for both domains.

2. DOMAIN SHIFTS

The goal of ASD is to determine, based on a sound recording of a phenomenon, whether that phenomenon is normal or anomalous. In the context of the DCASE Challenges, the objective is to differentiate between normal and anomalous sounds produced by a known type of machine. The normal/anomalous characteristic of a phenomenon is an intrinsic property of that phenomenon: as such, the determination by the ASD system should ideally be independent of the conditions in which the sound was recorded, as long as the change in conditions does not alter the observability and nature of the phenomenon's normal/anomalous characteristic. Such changes include the use of microphones of different types or in different locations, the presence or absence of other sound sources, or modifications in certain properties of the monitored sound sources themselves, e.g., the use of different settings of the monitored machines.

ASD systems are however often placed in a conundrum: their only view of the phenomenon's normal/anomalous characteristic is via a training set of normal data recorded under a certain set of conditions, referred to as a *source domain*, and they are expected to make a determination on the normal/anomalous characteristic on a new sample recorded under a potentially different set of conditions, referred to as a *target domain*, for which they only have a few training samples. Changes in the recording conditions occurring between the source domain and the target domain constitute a so-called *domain shift*. ASD systems thus need to implicitly disentangle the effects on the sound signal of the normal/anomalous characteristic and the recording conditions, in order to be robust to changes in these conditions.

An additional difficulty lies in the large imbalance often observed in

Table 1: DCASE ASD datasets in domain-shifted conditions.

		# machine types			# sections (per machine type)				# recordings (per section)				
									source domain		target domain		
Name	Purpose	total	dev. set	eval. set	total	dev. set	eval. set	split	normal	anomalous	normal	anomalous	supplemental
DCASE2021 [1]	DA	7	7	7	6	3	3	train	1000	0	3	0	0
								test	100	100	100	100	0
DCASE2022 [2]	DG	7	7	7	6	3	3	train	990	0	10	0	0
								test	50	50	50	50	0
DCASE2023 [3]	DG	14	7	7	1	1	1	train	990	0	10	0	0
								test	50	50	50	50	0
DCASE2024 [4]	DG	16	7	9	1	1	1	train	990	0	10	0	0
								test	50	50	50	50	0
DCASE2025 [5]	DG	14	7	7	1	1	1	train	990	0	10	0	100
								test	50	50	50	50	0

practice between the number of available training samples in the source and target domains. Many training samples are typically available for the source domain, while only a few training samples may be available for the target domain. Recording a large amount of new training samples after a domain shift occurred may be impractical, as it requires much effort, and thus is very costly; domain shifts may also frequently or even continuously occur; and operators may not even be aware of such domain shifts.

A consequence of domain shifts related to differences in signal space is that anomaly scores are usually also distributed very differently across both domains. This results in a so-called *domain mismatch*, illustrated in Fig. 1. Since the embedding models were trained with no or few samples belonging to the target domain, the anomaly score distributions for the normal and anomalous samples are not well-separated in the target domain, which inherently leads to a worse ASD performance than in the source domain. Furthermore, the optimal decision thresholds differ substantially, further degrading the performance if the same decision threshold is used for both domains.

3. DOMAIN ADAPTATION

One way to handle domain shifts is to adapt an existing ASD system that was trained with sufficient data from a source domain to a specific target domain. Here, the main challenge is that the training set for the target domain consists of only very few samples and thus knowledge from the source domain, which may differ substantially from the target domain, needs to be transferred somehow to obtain a well-performing system. Note that once a system is adapted to a target domain, it may no longer perform well in the source domain it was trained on.

3.1. Datasets

A dataset focusing on handling domain shifts for ASD through DA is the DCASE2021 ASD dataset [1]. This dataset contains 10 s recordings of five different machine types from MIMII DUE [13] and two additional machine types from ToyADMOS2 [14] that are combined with background noise from real factories. The dataset is divided into a development set and an evaluation set, which both contain three sections for each machine type. These sections are specific partitions of the dataset for calculating the performance and may contain recordings from multiple machines of the same type. Both the development and evaluation sets consist of a training split and a test split. The training splits contain only normal data, of which 1,000 samples belong to the source domain and only 3 samples belong to the target domain, resulting in a very imbalanced dataset in terms of domain. For each training file, additional attribute information about machine settings or the acoustic environment is provided, and it can be utilized for training ASD systems. The test splits contain 200 samples for each domain, of which one half are normal and the other half are anomalous. For each test file, it is known whether these files belong to the source or the target domain, but it is unknown whether they are normal or anomalous. Summarized statistics of this dataset can be found in the top part of Table 1.

3.2. Approaches

Approaches for DA mostly focus on first training an ASD system with data from the source domain and then adapting this trained system to a specific target domain. One possibility to do this is to fine-tune an entire model that was trained in the source domain with data belonging to the target domain [15], [16]. This changes the problem of balancing the very differently sized training datasets of both domains to a problem of preventing the model from overfitting to the few target domain samples available for training. In [17], only the parameters of batch normalization layers [36] are fine-tuned to the target domain to minimize the computational costs needed for the adaptation while also reducing overfitting effects. A similar idea based on domain-specific normalization layers is realized in [18] by using AutoDIAL [37]. The authors of [19] propose to use gradient-based meta learning [38] and a prototypical loss [39] in order to more effectively adapt to target domains with only a very small amount of training samples. Another DA approach is to simply train a joint embedding model for both domains but estimate the distributions of each domain individually [16], [20]. Table 2 briefly summarizes all mentioned approaches.

4. DOMAIN GENERALIZATION

Adapting models for each new domain with the possible need of re-training models, fine-tuning hyperparameters, or even replacing system components is very impractical, as it is computationally costly and may require expert knowledge. It is thus much more desirable to strive for domain generalization (DG) [40], [41], with the goal of obtaining a model that performs well in the source domain and generalizes well to unseen target domains by only providing a few samples to define how normal sounds belonging to these domains should sound like. Since DG is literally a generalization of DA to arbitrary domains rather than a specific target domain, it is more difficult to obtain such a model. To highlight the practical differences between DA and DG, the reader is reminded that for DA, models can be fully adapted to a specific target domain, even if performance in the source domain is weak after adaptation. In contrast, DG requires strong performance in the source and target domain, using even the same decision threshold for the anomaly scores in both domains.

4.1. Datasets

Currently, there are four DCASE ASD datasets focusing on DG: the DCASE2022 [2], DCASE2023 [3], DCASE2024 [4], and DCASE2025 [5] datasets, which are all based on MIMII DG [42] as well as,

Table 2: ASD approaches for handling domain shifts.

Topic	Strategy	Reference	Main idea	DCASE datasets used
DA	-	Chen et al [15]	fine-tune model	2021
DA	-	Kuroyanagi et al [16]	fine-tune models and domain-specific backends	2021
DA	-	Yamaguchi et al [17]	fine-tune batch normalization layers	other
DA	-	Lopez et al [18]	domain-specific normalization with AutoDIAL	2021
DA	-	Chen et al [19]	gradient-based meta learning	2021
DA	-	Wilkinghoff [20]	domain-specific backends	2021
DG	domain specialization	Kuroyanagi et al [21]	domain-specific models and domain classifier	2022
DG	domain specialization	Kuroyanagi et al [21]	balance mini-batches with mixup	2022
DG	domain specialization	Guan et al [22]	balance domains with SMOTE	2022
DG	domain specialization	Junjie et al [23]	balance domains with SMOTE	2023
DG	domain specialization	Chen et al [24]	balance domains with SMOTE and a mixup variant	2022,2023
DG	domain-invariant representations	Deng et al [25]	normalize samples independently for each batch	2022
DG	domain-invariant representations	Nejjar et al [26]	VICreg extension with mixup	2022
DG	domain-invariant representations	Yan et al [27]	minimize covariance differences between domains	2020, 2022
DG	feature disentanglement	Venkatesh et al [28]	discriminative tasks for sections and attributes	2022
DG	feature disentanglement	Dohi et al [29]	disentangle attributes with normalizing flow	other
DG	feature disentanglement	Lan et al [30]	Mahalanobis distances for hierarchical metadata	2022
DG	feature disentanglement	Guan et al [31]	gradient reversal with hierarchical metadata	2022
DG	anomaly score calculation	Harada et al [32]	domain-specific Mahalanobis distance	2022
DG	anomaly score calculation	Wilkinghoff [33]	nearest neighbor based anomaly scores	2022
DG	anomaly score calculation	Wilkinghoff et al [34]	local density based score normalization	2020, 2023, 2024
DG	anomaly score calculation	Saengthong et al [35]	domain-wise standardization of scores	2020, 2023

respectively, ToyADMOS2 [14], ToyADMOS2+ [43], and ToyADMOS2# [44]. A part of the DCASE2024 dataset was recorded with the same setup as IMAD-DS [45]. IMAD-DS is another ASD dataset containing noisy multi-sensor signals of two machines in domain-shifted conditions recorded with a microphone, an accelerometer, and a gyroscope. The fundamental difference with the DCASE2021 dataset, which focuses on DA, is that the domain of individual test samples is unknown during inference. Moreover, the performance for each section is computed jointly for the source and target domains, i.e., a single decision threshold needs to be used for both domains. Other less severe differences include the size of all test sets, which is only half the size of the DCASE2021 test sets, and the availability of 10 normal training samples belonging to the target domain instead of 3. The differences between the DCASE2021 and DCASE2022 datasets end there, making them relatively similar. In contrast, the DCASE2023, DCASE2024, and DCASE2025 datasets only consist of a single section for each machine type, and the development and evaluation sets contain recordings of completely different machine types. Furthermore, the DCASE2024 and DCASE2025 datasets have noise conditions that are exclusively used for specific machine types, and for some machine types no additional attribute information is provided. The major difference between the DCASE2024 and DCASE2025 datasets is that supplemental recordings containing noise-free recordings of normal machine sounds or only background noise are provided for the train splits of the DCASE2025 dataset. Each year, modifications to the datasets result in a more realistic yet more challenging ASD task. For more details about these datasets, the reader is referred to Table 1 and the corresponding references.

4.2. Approaches

DG approaches for ASD in the context of DCASE can be grouped into four different strategies, which will now be discussed. Table 2 contains an overview of these approaches.

4.2.1. Domain Specialization: A simple approach for reducing domain mismatch is to balance the number of training samples belonging to the source and target domains. This can be achieved by balancing the domains in each mini-batch [21] or using more sophisticated approaches [22]–[24] such as synthetic minority oversampling technique (SMOTE) [46]. However, since this approach trains models for specific domain shifts and thus requires to re-train the entire ASD system for each possible domain shift, this can be referred to as *weak DG* or *domain specialization*. Note that in contrast to DA, ASD systems need to work well for all domains without knowledge about the domain a given sample belongs to. Another way to handle this is to train domain-specific models and a domain classifier [21].

4.2.2. Domain-invariant Representations: The idea of domain-invariant representation learning [47] or domain-mixing-based approaches [2] is to reduce the variance between multiple domains based on the assumption that this also reduces the variance to other arbitrary target domains. In [25], individual samples of a batch are normalized independently of each other instead of using batch normalization [36] to avoid learning statistics that over-emphasize the source domain due to the highly imbalanced number of samples. The authors of [26] propose DG-mix, an extension of variance-invariance-covariance regularization (VICreg) [48] for self-supervised pre-training, using a loss term that minimizes the difference between domains and virtual domains created by mixup [49] before fine-tuning the model. Motivated by [50], which reduces the difference between second-order statistics of intermediate representations from domain-specific convolutional neural networks, [27] aims at reducing the difference between second-order statistics of source domain features and target domain features. To this end, the authors applied augmentations such as pitch shifting, time shifting, time stretching, adding white noise, and Filteraugment [51] to target domain samples to create more diverse domain shifts for training.

4.2.3. Feature Disentanglement: The main idea of feature disentanglement [52] is to decompose representations into domain-related features that are invariant for different domains and domain-unrelated features that capture the variability of different domains. The latter are independent of the domain and thus also generalize well to unseen domains. A possible way to achieve this is to use a domain-classification-based approach [2]. In most works on ASD in the DCASE context, this is achieved by focusing on the provided

Table 3: Comparison of different types of DG approaches.

DG approach	Effort	Requires labels	Effectiveness
Domain Specialization	high	yes	high
Domain-invariant Representations	medium	no	medium
Feature Disentanglement	low	yes	medium
Anomaly Score Calculation	very low	sometimes	low

attribute information. The authors of [28] use two discriminative tasks for the sections and attributes. In [29], attribute information is disentangled in a normalizing-flow-based ASD model. In [30], a combination of a hierarchical metadata structure and attribute-specific Mahalanobis distances is used to learn more domain-related features. This approach is extended by [31] with gradient-reversal-based feature disentanglement of attribute information [53] and the use of a focal loss [54].

4.2.4. Anomaly Score Calculation: Last but not least, DG capabilities can also be improved by modifying the anomaly score computation. This approach has the advantage of not requiring expensive re-training of the neural networks that serve as the basis of state-of-the-art ASD systems. To this end, [32] uses an autoencoder and calculates the Mahalanobis distance between input data and reconstruction using domain-specific covariance matrices. For discriminative ASD systems, it was shown that simple nearest-neighbor-based anomaly scores lead to better results than estimating domain-specific distributions [33]. Further improvements can be obtained by normalizing the anomaly scores to reduce the domain mismatch. Examples are normalizing the anomaly scores based on local densities of normal reference samples [34], [55] or by a domain-wise standardization of the anomaly score distributions [35].

5. LIMITATIONS AND FUTURE DIRECTIONS

All presented DG approaches have limitations. A comparison of the approaches can be found in Table 3. Although domain specialization approaches are highly effective, they can only adapt to specific domains and thus the entire ASD system, including the embedding model, needs to be re-trained for each domain shift, which is highly impractical as it is a strenuous process requiring access to domain labels. Obtaining domain-invariant representations does not require domain labels, but requires to simulate a diverse set of realistic domain shifts, which is a difficult task by itself and thus may limit the effectiveness. Feature disentanglement does not require much additional effort, but does require access to labeled data of various domains to be effective, which may not always be available. Focusing only on the calculation of the anomaly scores usually does not require one to modify the training of the embedding models. However, this approach is still sensitive to domain shifts if these also affect the obtained embeddings, thus limiting its effectiveness, even when making use of domain labels. Depending on the application and the available data, different DG approaches can be jointly used to mitigate the limitations of individual approaches. For example, anomaly score calculation approaches for DG can be used regardless of the learned representations, and simulated domain shifts, as used when learning domain-invariant representations, can also be used for feature disentanglement. However, the underlying issues of individual approaches persist and future work on DG is needed to solve these problems and identify potential synergies between specific approaches.

A future direction of research is continuous DA [56]. Here, the discrete indices for the domains are replaced with continuous ones to model the underlying relations between different domains. This improves consistency and may even allow to handle domains for which no training samples are available. Another direction is continual DA [57] that aims at handling domain shifts in non-stationary environments where domains are not static and may change continuously over time. In these environments, only adapting to a fixed domain is not sufficient. Note that continuous domain indices and continuously adapting to non-static domains can also be realized for DG. As a third direction, explaining the decisions of ASD systems is an open problem [58], [59], particularly in domain-shifted conditions. Here, future work may focus on explaining decisions of ASD in unknown target domains [60] or explaining domain shifts themselves [61]. Explainable ASD systems that focus on particular characteristics of the data may even offer better DG capabilities [62]. Last but not least, the semi-supervised ASD setting in the DCASE context can be relaxed to an unsupervised setting where even the training dataset may contain unlabeled anomalies [63].

6. CONCLUSION

In this work, we reviewed recent work on how to handle domain shifts for ASD tasks related to DCASE. We first defined domain shifts in the context of ASD, then discussed the topics of *domain adaptation (DA)* and *domain generalization (DG)* by motivating and defining these terms, presenting relevant datasets, and collecting different works related to these topics. In particular, DG approaches were grouped into *domain specialization*, *domain-invariant representations*, *feature disentanglement*, and *anomaly score calculation*. Furthermore, *continuous* and *continual DA* as well as *explainable DG* were identified as possible future research directions. For future work, we plan to compare all presented techniques with ASDKit [64] through extensive experiments to identify the best approaches and potential synergies.

REFERENCES

[1] 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*, 2021.

[2] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, 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*, 2022.

[3] 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*, 2023.

[4] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, 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*, 2024.

[5] ——, "Description and discussion on DCASE 2025 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," *arXiv preprint arXiv:2506.10097*, 2025.

[6] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, "A robust interpretable deep learning classifier for heart anomaly detection without segmentation," *IEEE J. Biomed. Health Informatics*, vol. 25, no. 6, 2021.

[7] S. N. Murthy and E. Agu, "Deep learning anomaly detection methods to passively detect COVID-19 from audio," in *Proc. ICDH*, 2021.

[8] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, and M. Vento, "Audio surveillance of roads: A system for detecting anomalous sounds," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, 2016.

[9] C. Zieger, A. Brutti, and P. Svaizer, "Acoustic based surveillance system for intrusion detection," in *Proc. AVSS*, 2009.

[10] T. Hayashi, T. Komatsu, R. Kondo, T. Toda, and K. Takeda, "Anomalous sound event detection based on WaveNet," in *Proc. EUSIPCO*, 2018.

[11] K. Wilkinghoff, T. Fujimura, K. Imoto, and J. Le Roux, "Handling domain shifts for anomalous sound detection: A review," in *Proc. DAS/DAGA*, 2025.

[12] A. Mesaros, R. Serizel, T. Heittola, T. Virtanen, and M. D. Plumbley, "A decade of DCASE: achievements, practices, evaluations and future challenges," in *Proc. ICASSP*, 2025.

[13] 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," in *Proc. WASPAA*, 2021.

[14] 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*, 2021.

[15] H. Chen, Y. Song, L. Dai, I. McLoughlin, and L. Liu, "Self-supervised representation learning for unsupervised anomalous sound detection under domain shift," in *Proc. ICASSP*, 2022.

[16] I. Kuroyanagi, T. Hayashi, Y. Adachi, T. Yoshimura, K. Takeda, and T. Toda, "An ensemble approach to anomalous sound detection based on conformer-based autoencoder and binary classifier incorporated with metric learning," in *Proc. DCASE*, 2021.

[17] M. Yamaguchi, Y. Koizumi, and N. Harada, "AdaFlow: Domain-adaptive density estimator with application to anomaly detection and unpaired cross-domain translation," in *Proc. ICASSP*, 2019.

[18] J. A. Lopez, G. Stemmer, P. Lopez-Meyer, P. Singh, J. A. del Hoyo Ontiveros, and H. A. Cordourier, "Ensemble of complementary anomaly detectors under domain shifted conditions," in *Proc. DCASE*, 2021.

[19] B. Chen, L. Bondi, and S. Das, "Learning to adapt to domain shifts with few-shot samples in anomalous sound detection," in *Proc. ICPR*, 2022.

[20] K. Wilkinghoff, "Combining multiple distributions based on sub-cluster AdaCos for anomalous sound detection under domain shifted conditions," in *Proc. DCASE*, 2021.

[21] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Two-stage anomalous sound detection systems using domain generalization and specialization techniques," in *Proc. DCASE*, 2022.

[22] J. Guan, Y. Liu, Q. Zhu, T. Zheng, J. Han, and W. Wang, "Time-weighted frequency domain audio representation with GMM estimator for anomalous sound detection," in *Proc. ICASSP*, 2023.

[23] W. Junjie, W. Jiajun, C. Shengbing, S. Yong, and L. Mengyuan, "Anomaly sound detection system based on multi-dimensional attention module," DCASE2023 Challenge, Tech. Rep., 2023.

[24] S. Chen, Y. Sun, J. Wang, M. Wan, M. Liu, and X. Li, "A multi-scale dual-decoder autoencoder model for domain-shift machine sound anomaly detection," *Digit. Signal Process.*, vol. 156, 2025.

[25] Y. Deng, A. Jiang, Y. Duan, J. Ma, X. Chen, J. Liu, P. Fan, C. Lu, and W. Zhang, "Ensemble of multiple anomalous sound detectors," in *Proc. DCASE*, 2022.

[26] I. Nejjar, J. Meunier-Pion, G. Frusque, and O. Fink, "DG-Mix: Domain generalization for anomalous sound detection based on self-supervised learning," in *Proc. DCASE*, 2022.

[27] J. Yan, Y. Cheng, Q. Wang, L. Liu, W. Zhang, and B. Jin, "Transformer and graph convolution-based unsupervised detection of machine anomalous sound under domain shifts," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 8, no. 4, 2024.

[28] S. Venkatesh, G. Wichern, A. S. Subramanian, and J. Le Roux, "Improved domain generalization via disentangled multi-task learning in unsupervised anomalous sound detection," in *Proc. DCASE*, 2022.

[29] K. Dohi, T. Endo, and Y. Kawaguchi, "Disentangling physical parameters for anomalous sound detection under domain shifts," in *Proc. EUSIPCO*, 2022.

[30] H. Lan, Q. Zhu, J. Guan, Y. Wei, and W. Wang, "Hierarchical metadata information constrained self-supervised learning for anomalous sound detection under domain shift," in *Proc. ICASSP*, 2024.

[31] J. Guan, J. Tian, Q. Zhu, F. Xiao, H. Zhang, and X. Liu, "Disentangling hierarchical features for anomalous sound detection under domain shift," in *Proc. ICASSP*, 2025.

[32] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline," in *Proc. EUSIPCO*, 2023.

[33] K. Wilkinghoff, "Design choices for learning embeddings from auxiliary tasks for domain generalization in anomalous sound detection," in *Proc. ICASSP*, 2023.

[34] K. Wilkinghoff, H. Yang, J. Ebbers, F. G. Germain, G. Wichern, and J. Le Roux, "Local density-based anomaly score normalization for domain generalization," *arXiv preprint arXiv:2509.10951*, 2025.

[35] P. Saengthong and T. Shinozaki, "Deep generic representations for domain-generalized anomalous sound detection," in *Proc. ICASSP*, 2025.

[36] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015.

[37] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulò, "AutoDIAL: Automatic domain alignment layers," in *Proc. ICCV*, 2017.

[38] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," *arXiv preprint arXiv:1803.02999*, 2018.

[39] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. NeurIPS*, 2017.

[40] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, 2023.

[41] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, 2023.

[42] 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*, 2022.

[43] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "ToyADMOS2+: New toyadmos data and benchmark results of the first-shot anomalous sound event detection baseline," in *Proc. DCASE*, 2023.

[44] D. Niizumi, N. Harada, Y. Ohishi, D. Takeuchi, and M. Yasuda, "ToyADMOS2#: Yet another dataset for the DCASE2024 challenge task 2 first-shot anomalous sound detection," in *Proc. DCASE*, 2024.

[45] D. Albertini, F. Augusti, K. Esmer, A. Bernardini, and R. Sannino, "IMAD-DS: A dataset for industrial multi-sensor anomaly detection under domain shift conditions," in *Proc. DCASE*, 2024.

[46] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, 2002.

[47] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Proc. NeurIPS*, 2006.

[48] A. Bardes, J. Ponce, and Y. LeCun, "VICReg: Variance-invariance-covariance regularization for self-supervised learning," in *Proc. ICLR*, 2022.

[49] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *Proc. ICLR*, 2018.

[50] B. Sun and K. Saenko, "Deep CORAL: correlation alignment for deep domain adaptation," in *Proc. ECCV*, 2016.

[51] H. Nam, S. Kim, and Y. Park, "Filteraugment: An acoustic environmental data augmentation method," in *Proc. ICASSP*, 2022.

[52] H. Zhang, Y. Zhang, W. Liu, A. Weller, B. Schölkopf, and E. P. Xing, "Towards principled disentanglement for domain generalization," in *Proc. CVPR*, 2022.

[53] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky, "Domain-adversarial training of neural networks," *JMLR*, vol. 17, 2016.

[54] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. ICCV*, 2017.

[55] K. Wilkinghoff, H. Yang, J. Ebbers, F. G. Germain, G. Wichern, and J. Le Roux, "Keeping the balance: Anomaly score calculation for domain generalization," in *Proc. ICASSP*, 2025.

[56] H. Wang, H. He, and D. Katabi, "Continuously indexed domain adaptation," in *Proc. ICML*, 2020.

[57] Q. Wang, O. Fink, L. V. Gool, and D. Dai, "Continual test-time domain adaptation," in *Proc. CVPR*, 2022.

[58] K. T. Mai, T. Davies, L. D. Griffin, and E. Benetos, "Explaining the decision of anomalous sound detectors," in *Proc. DCASE*, 2022.

[59] S. Tsubaki, Y. Kawaguchi, T. Nishida, K. Imoto, Y. Okamoto, K. Dohi, and T. Endo, "Audio-change captioning to explain machine-sound anomalies," in *Proc. DCASE*, 2023.

[60] J. Gao, X. Ma, and C. Xu, "Learning transferable conceptual prototypes for interpretable unsupervised domain adaptation," *IEEE Trans. Image Process.*, vol. 33, 2024.

[61] S. Bobek, S. Nowaczyk, S. Pashami, Z. Taghiyarrenani, and G. J. Nalepa, "Towards explainable deep domain adaptation," in *Proc. ECAI Workshops*, 2023.

[62] A. Zunino, S. A. Bargal, R. Volpi, M. Sameki, J. Zhang, S. Sclaroff, V. Murino, and K. Saenko, "Explainable deep classification models for domain generalization," in *Proc. CVPR Workshops*, 2021.

[63] C. Aggarwal, *Outlier Analysis*, 2nd ed. Springer, 2017.

[64] T. Fujimura, K. Wilkinghoff, K. Imoto, and T. Toda, "ASDKit: A toolkit for comprehensive evaluation of anomalous sound detection methods," *arXiv preprint arXiv:2507.10264*, 2025.