Handling Domain Shifts for Anomalous Sound Detection: A Review

Kevin Wilkinghoff[1], Takuya Fujimura[2], Keisuke Imoto[3], Jonathan Le Roux[4]

[1] *Aalborg University and Pioneer Centre for AI, Aalborg, Denmark, email: kevin.wilkinghoff@ieee.org*

[2] *Nagoya University, Nagoya, Aichi, Japan email: fujimura.takuya@g.sp.m.is.nagoya-u.ac.jp*

[3] *Doshisha University, Kyotanabe, Kyoto, Japan, email: keisuke.imoto@ieee.org*

[4] *Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, email: leroux@merl.com*

Abstract

When detecting anomalous sounds in possibly complex environments, one of the main difficulties is that trained models need to be sensitive to subtle differences of monitored target signals. At the same time, for many practical applications these models should be insensitive to changes of the acoustic domain. Examples of these domain shifts are changing the microphone or the location of acoustic sensors, which both may have a much stronger impact on the acoustic signal than subtle anomalies. Moreover, users want to train a model only with a relatively large collection of source domain data. Furthermore, such a trained model should be able to generalize well to any unseen domain by only providing very few samples of the target domain to define how acoustic signals in this domain sound like. 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.

Introduction

Anomalous sound detection (ASD) has many applications. Examples are acoustic monitoring of machines [22, 8, 9, 32], health [27], roads [12], smart home environments [45] or public places [19]. The goal of all these applications is to distinguish between normal and anomalous audio recordings. Usually, ASD systems are trained with normal data exclusively as anomalous data is often difficult and costly to obtain.

One of the major difficulties that ASD systems need to overcome are the so-called *domain shifts*. These changes of the recorded audio signals are caused by changes of the acoustic environment, sensors, or properties of monitored sound sources themselves. Inherently, domain shifts have a strong impact on the audio signals and therefore also affect the outcomes of ASD systems if no precautions are taken. However, for many applications, this effect is not desirable and should be suppressed. Ideally, trained ASD systems are completely insensitive to domain shifts. At the same time, ASD systems should be very sensitive to modifications of the monitored target signals that indicate the occurrence of application-dependent anomalies.

This manuscript reviews the latest work on handling domain shifts for ASD in the context of the annual DCASE challenge [26]. To this end, the two main topics *domain adaptation* and *domain generalization* will be discussed by defining these terms, listing publicly available datasets and recently published works related to these topics.

Domain Shifts

In general, samples that serve as input to an ASD system consist of audio signals or features derived from them. It is possible that the acoustic environment changes causing the underlying audio signals to be altered without changing the intrinsic properties that define them to be normal or anomalous. Examples are changing the microphones or their locations, including or removing other sound sources, or modifying certain properties of the monitored sound sources themselves, e.g., changing the settings of monitored machines. Such changes in the acoustic environment occurring between a *source domain* and a *target domain* constitute a so-called domain shift. In addition to the acoustic differences of both domains, the domains also differ in the number of available training samples. For the source domain, there are sufficiently many training samples available and for the target domain, only a few training samples are available because collecting sufficiently many samples after a domain shift occurred is impractical.

The most severe consequence of domain shifts related to the strong differences in signal space is that the anomaly scores are usually also distributed very differently in both domains. This is called *domain mismatch*. Since the embedding models were trained without data belonging to the target domain, the two anomaly score distributions belonging to normal and anomalous samples are not well-separated in the target domain, which inherently leads to a worse ASD performance as in the source domain. Furthermore, the optimal decision thresholds differ in both domains, which degrades the performance even more if a single decision threshold is used.

Domain Adaptation

One way to handle domain shifts is to adapt an existing ASD system that was trained on a source domain to a particular 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, has to be transferred somehow to obtain a well-performing system. Note that once a system is adapted to a target domain, it does not need to still perform well in the source domain for which it was initially trained.

Datasets

A dataset focusing on handling domain shifts for ASD through domain adaptation is the DCASE2021 ASD dataset [22]. This dataset contains 10 s recordings of five different machine types from MIMII DUE [35] and two additional machine types from ToyAdmos2 [17] 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. Furthermore, both the development and evaluation set consist of a training split and a test split. The training splits contain only normal data, of which 1000 samples belong to the source domain and only 3 samples belong to the target domain, resulting in a very imbalanced dataset in view of domain. For each training file, additional attribute information about machine settings or the acoustic environment are provided that can be used to train the 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.

Approaches

Approaches for domain adaptation 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 of doing this is to fine-tune an entire model that was trained in the source domain with data belonging to the target domain [6]. 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 [41], only the parameters of the batch normalization layers [20] are fine-tuned to the target domain to minimize the computational costs needed for the adaptation while also reducing overfitting effects. The authors of [5] propose to use gradient-based meta learning [30] and a prototypical loss [34] to be able to more effectively adapt to target domains with only a few training samples. Another domain adaptation approach is to simply train a joint embedding model for both domains but estimate the distributions of each domain individually [38].

Domain Generalization

Adapting models for each unknown domain with the possible need of re-training models, fine-tuning hyperparameters or even replacing system components is very impracticable since it is computationally costly and may require expert knowledge. A model that performs well on the source domain and also generalizes well to unseen target domains is much more favorable. This is called *domain generalization* [37]. However, since domain generalization is literally a generalization of domain adaptation to arbitrary instead of a specific target domain obtaining such a well-generalizing model is much more difficult to achieve.

Datasets

Currently, there are three ASD datasets focusing on domain generalization. These are the DCASE2022 [8], DCASE2023 [9] and DCASE2024 [32] datasets, which are all based on MIMII DG [10] as well as, respectively, ToyADMOS2 [17], ToyADMOS2+ [18], and ToyADMOS2# [31]. A part of the DCASE2024 dataset was recorded with the same setup as IMAD-DS [1]. Compared to the DCASE2021 dataset related to domain adaptation, the fundamental difference is that during inference the domain of individual test samples is unknown. Moreover, the performance for each section is computed jointly for the source and target domain, i.e. a single decision threshold needs to be used for both domains. Other less severe differences are that the size of all test sets is only half the size of the DCASE2021 test sets and that there are 10 normal training samples belonging to the target domain available instead of 3. Otherwise, the DCASE2021 and DCASE2022 datasets are very similar. In contrast, the DCASE2023 and DCASE2024 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 dataset has noise conditions that are exclusively used for specific machine types and for some machine types no additional attribute information are provided. For more details about the structures of these datasets, the reader is referred to the corresponding references.

Approaches

Domain Specialization

A simple approach to reduce 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 [23] or using more sophisticated approaches [15, 21] such as SMOTE [4]. 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 domain generalization* or *domain specialization*. Note that in contrast to domain adaptation, ASD systems need to work well for all domains without having knowledge about the domain a given sample belongs to. Another way to handle this is to train domain-specific models and a domain classifier [23]. A closely related variant is to minimize the difference between data of different domains. The authors of [29] propose DG-mix, an extension of variance-invariance-covariance regularization (VICreg) [2] for self-supervised pre-training, using a loss term that minimizes the difference between domains and virtual domains created by mixup [43] before fine-tuning the model.

Domain-Invariant Representation Learning

Domain-invariant representation learning [3] or a domain-mixing-based approach [8] reduces the variance between multiple source domains based on the assumption that this also reduces the variance to an arbitrary target domain. In [7], individual samples of a batch are normalized independently of each other to avoid overemphasizing the source domain due to the highly imbalanced number of samples. The authors of [29] use a loss term that minimizes the difference between domains and virtual domains created by mixup [43] before fine-tuning the model. Similarly, [42] aims to reduce 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 [28] to target domain samples to create more diverse domain shifts for training.

Feature Disentanglement

The main idea of *feature disentanglement* [44] is to decompose the data into domain-related and domain-unrelated features, the latter are independent of the domain and thus also generalize well to unseen domains. In most works on ASD, this is achieved by focusing on the attribute information provided. The authors of [36] use two discriminative tasks for the sections and attributes and in [11] attribute information are disentangled in a normalizing flow-based ASD model. In [24], 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 [14] with gradient reversal-based [13] feature disentanglement of attribute information and using a focal loss [25].

Anomaly Score Calculation

Last but not least, domain generalization capabilities can be improved by modifying the anomaly score computation. This approach has the advantage that no expensive re-training of neural networks that serve as the basis of state-of-the-art ASD systems is needed and thus can be labeled as *strong domain generalization.* To this end, [16] 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 [39]. 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 [40] or by a domain-wise standardization of the anomaly score distributions [33].

Conclusion

In this manuscript, recent work on how to handle domain shifts for ASD tasks was reviewed. To this end, the topics *domain adaptation* and *domain generalization* were discussed by motivating and defining these terms, presenting relevant datasets and collecting different works related to these topics. For future work, it is planned to compare all presented techniques in extensive experimental evaluations.

References

[1] Davide Albertini et al. “IMAD-DS: A Dataset for Industrial Multi-Sensor Anomaly Detection Under Domain Shift Conditions”. In: *Proc. DCASE.* 2024, pp. 1–5.

[2] Adrien Bardes, Jean Ponce, and Yann LeCun. “VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning”. In: *Proc. ICLR.* 2022.

[3] Shai Ben-David et al. “Analysis of Representations for Domain Adaptation”. In: *Proc. NeurIPS.* 2006, pp. 137–144.

[4] Nitesh V. Chawla et al. “SMOTE: Synthetic Minority Over-sampling Technique”. In: *J. Artif. Intell. Res.* 16 (2002), pp. 321–357.

[5] Bingqing Chen, Luca Bondi, and Samarjit Das. “Learning to Adapt to Domain Shifts with Few-shot Samples in Anomalous Sound Detection”. In: *Proc. ICPR.* 2022, pp. 133–139.

[6] Han Chen et al. “Self-Supervised Representation Learning for Unsupervised Anomalous Sound Detection Under Domain Shift”. In: *Proc. ICASSP.* 2022, pp. 471–475.

[7] Yufeng Deng et al. “Ensemble of Multiple Anomalous Sound Detectors”. In: *Proc. DCASE.* 2022, pp. 21–25.

[8] Kota Dohi et al. “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.

[9] Kota Dohi et al. “Description and Discussion on DCASE 2023 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring”. In: *Proc. DCASE.* 2023, pp. 31–35.

[10] Kota Dohi et al. “MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task”. In: *Proc. DCASE.* 2022, pp. 31–35.

[11] Kota Dohi, Takashi Endo, and Yohei Kawaguchi. “Disentangling physical parameters for anomalous sound detection under domain shifts”. In: *Proc. EUSIPCO.* 2022, pp. 279–283.

[12] Pasquale Foggia et al. “Audio Surveillance of Roads: A System for Detecting Anomalous Sounds”. In: *IEEE Trans. Intell. Transp. Syst.* 17.1 (2016), pp. 279–288.

[13] Yaroslav Ganin et al. “Domain-Adversarial Training of Neural Networks”. In: *JMLR* 17 (2016), 59:1–59:35.

[14] Jian Guan et al. “Disentangling Hierarchical Features for Anomalous Sound Detection Under Domain Shift”. In: *Proc. ICASSP.* 2025.

[15] Jian Guan et al. “Time-Weighted Frequency Domain Audio Representation with GMM Estimator for Anomalous Sound Detection”. In: *Proc. ICASSP.* 2023.

[16] Noboru Harada et al. “First-Shot Anomaly Sound Detection for Machine Condition Monitoring: A Domain Generalization Baseline”. In: *Proc. EUSIPCO.* 2023, pp. 191–195.

[17] Noboru Harada et al. “ToyADMOS2: Another Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection under Domain Shift Conditions”. In: *Proc. DCASE*. 2021.

[18] Noboru Harada et al. “ToyADMOS2+: New Toyadmos Data and Benchmark Results of the First-Shot Anomalous Sound Event Detection Baseline”. In: *Proc. DCASE*. 2023, pp. 41–45.

[19] Tomoki Hayashi et al. “Anomalous Sound Event Detection Based on WaveNet”. In: *Proc. EUSIPCO*. 2018, pp. 2494–2498.

[20] Sergey Ioffe and Christian Szegedy. “Batch normalization: accelerating deep network training by reducing internal covariate shift”. In: *Proc. ICML*. Vol. 37. 2015, pp. 448–456.

[21] Wang Junjie et al. *Anomaly Sound Detection System Based on Multi-Dimensional Attention Module*. Tech. rep. DCASE2023 Challenge, 2023.

[22] Yohei Kawaguchi et al. “Description and Discussion on DCASE 2021 Challenge Task 2: Unsupervised Anomalous Detection for Machine Condition Monitoring Under Domain Shifted Conditions”. In: *Proc. DCASE*. 2021, pp. 186–190.

[23] Ibuki Kuroyanagi et al. “Two-stage anomalous sound detection systems using domain generalization and specialization techniques”. In: *Proc. DCASE*. 2022.

[24] Haiyan Lan et al. “Hierarchical Metadata Information Constrained Self-Supervised Learning for Anomalous Sound Detection under Domain Shift”. In: *Proc. ICASSP*. 2024, pp. 7670–7674.

[25] Tsung-Yi Lin et al. “Focal Loss for Dense Object Detection”. In: *Proc. ICCV*. 2017, pp. 2999–3007.

[26] Annamaria Mesaros et al. “A decade of DCASE: Achievements, practices, evaluations and future challenges”. In: *Proc. ICASSP*. 2025.

[27] Shreesha Narasimha Murthy and Emmanuel Agu. “Deep Learning Anomaly Detection methods to passively detect COVID-19 from Audio”. In: *Proc. ICDH*. 2021, pp. 114–121.

[28] Hyeonuk Nam, Seong-Hu Kim, and Yong-Hwa Park. “Filteraugment: An Acoustic Environmental Data Augmentation Method”. In: *Proc. ICASSP*. 2022, pp. 4308–4312.

[29] Ismail Nejjar et al. “DG-Mix: Domain Generalization for Anomalous Sound Detection Based on Self-Supervised Learning”. In: *Proc. DCASE*. 2022.

[30] Alex Nichol, Joshua Achiam, and John Schulman. “On First-Order Meta-Learning Algorithms”. In: *arXiv preprint arXiv:1803.02999* (2018).

[31] Daisuke Niizumi et al. “ToyADMOS2#: Yet Another Dataset for the DCASE2024 Challenge Task 2 First-Shot Anomalous Sound Detection”. In: *Proc. DCASE*. 2024, pp. 106–110.

[32] Tomoya Nishida et al. “Description and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring”. In: *Proc. DCASE*. 2024, pp. 111–115.

[33] Phurich Saengthong and Takahiro Shinozaki. “Deep Generic Representations for Domain-Generalized Anomalous Sound Detection”. In: *Proc. ICASSP*. 2025.

[34] Jake Snell, Kevin Swersky, and Richard S. Zemel. “Prototypical Networks for Few-shot Learning”. In: *Proc. NeurIPS*. 2017, pp. 4077–4087.

[35] Ryo Tanabe et al. “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, pp. 21–25.

[36] Satvik Venkatesh et al. “Improved Domain Generalization via Disentangled Multi-Task Learning in Unsupervised Anomalous Sound Detection”. In: *Proc. DCASE*. 2022.

[37] Jindong Wang et al. “Generalizing to Unseen Domains: A Survey on Domain Generalization”. In: *IEEE Trans. Knowl. Data Eng.* 35.8 (2023), pp. 8052–8072.

[38] Kevin Wilkinghoff. “Combining Multiple Distributions based on Sub-Cluster AdaCos for Anomalous Sound Detection under Domain Shifted Conditions”. In: *Proc. DCASE*. 2021, pp. 55–59.

[39] Kevin Wilkinghoff. “Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection”. In: *Proc. ICASSP*. 2023.

[40] Kevin Wilkinghoff et al. “Keeping the Balance: Anomaly Score Calculation for Domain Generalization”. In: *Proc. ICASSP*. 2025.

[41] Masataka Yamaguchi, Yuma Koizumi, and Noboru Harada. “AdaFlow: Domain-adaptive Density Estimator with Application to Anomaly Detection and Unpaired Cross-domain Translation”. In: *Proc. ICASSP*. 2019, pp. 3647–3651.

[42] Jingke Yan et al. “Transformer and Graph Convolution-Based Unsupervised Detection of Machine Anomalous Sound Under Domain Shifts”. In: *IEEE Trans. Emerg. Top. Comput. Intell.* 8.4 (2024), pp. 2827–2842.

[43] H. Zhang et al. “Mixup: Beyond empirical risk minimization”. In: *Proc. ICLR*. 2018.

[44] Hanlin Zhang et al. “Towards Principled Disentanglement for Domain Generalization”. In: *Proc. CVPR*. 2022, pp. 8014–8024.

[45] Christian Zieger, Alessio Brutti, and Piergiorgio Svaizer. “Acoustic Based Surveillance System for Intrusion Detection”. In: *Proc. AVSS*. 2009, pp. 314–319.