



Adaptive Across-Subcenter Representation Learning for Imbalanced Anomalous Sound Detection

Dong Wang¹, Jiqing Han¹, Guibin Zheng¹, Tieran Zheng¹, Yongjun He¹

¹Faculty of Computing, Harbin Institute of Technology, China

dongwang@stu.hit.edu.cn, jqhan@hit.edu.cn, zhengguibin@hit.edu.cn,
zhengtieran@hit.edu.cn, heyongjun@hit.edu.cn

Abstract

Anomalous Sound Detection requires constructing a distribution using only normal sounds. However, collecting sufficient normal samples across diverse conditions is challenging, leading to sample imbalance within subclasses. Existing subcenter angular margin loss methods use multiple subcenters to capture intra-class diversity but still suffer from under-representation or overfitting. To address this issue, we propose Adaptive Across-Subcenter Representation Learning (AASRL). Unlike existing methods that use either a single or all subcenters, AASRL adaptively selects subcenters based on the representation quality of samples and optimizes their representation across the most relevant subcenters. This ensures efficient representation of each sample and prevents the majority subclass from dominating the representation space. Experiments on the DCASE2023 Challenge Task2 dataset and a constructed imbalanced dataset demonstrate the effectiveness of AASRL.

Index Terms: anomalous sound detection, data imbalance, angular margin loss, subcenter

1. Introduction

The key challenge in Anomalous Sound Detection (ASD) is constructing a compact and discriminative distribution using only normal sounds, which allows for the accurate detection of anomalous sounds that deviate from this distribution [1, 2, 3]. However, in real-world application, normal sounds in each condition are difficult to capture in a balanced way, thus ASD faces the problem of data imbalance [4, 5]. Specifically, ensuring that normal samples from rare conditions are equally compact in the modeled distribution is crucial to avoid false detections and maintain detection performance.

Enhancing the compactness of the distribution is a commonly used strategy to address the imbalance problem among subclasses. One basic method is Outlier Exposure (OE) [6], which compresses the representation space of normal sounds by introducing non-targeted external data as proxy outliers for constructing a classification task [7, 8, 9]. The effectiveness of OE-based method can be greatly improved when combined with Subcenter Angular Margin Loss (SCAML) [10, 11], which creates multiple class centers (or prototypes [12]) for normal sounds to capture intra-class diversity. Based on the way that subcenters are used in SCAML for sample representation, it can be divided into two categories: optimal subclass representation [11] and joint subclass representation [10]. Both approaches have their limitations, the former lacks representational power because it uses only a single subclass center [13], while the latter tends to overfit majority samples [14, 15], although it can capture complex patterns by using all subclass centers [10, 16].

To address the above problems, we propose an Adaptive

Across-Subcenter Representation Learning (AASRL) method. Unlike previous strategies that use a single or all subcenters for representation, AASRL adaptively expands the sample representation subspace based on an angle-based representation quality measurement, such that samples with larger angles are assigned more subcenters, and their representations are optimized across the specific most relevant subcenters. This approach enables AASRL to perform local adaptive optimization of complex patterns in normal sounds while maintaining the compactness of the global distribution. Experimental results on the DCASE2023 Challenge Task2 dataset and a constructed imbalance dataset demonstrate the effectiveness of the proposed method, which improves the AUC and pAUC on both datasets by 1.05% & 0.62%, and 3.07% & 1.61%, respectively, compared with a strong baseline.

2. Background

2.1. Imbalanced anomalous sound detection

Existing methods for imbalanced anomalous sound detection can be divided into three categories based on the strategies employed: (i) *Data augmentation* methods mitigate the imbalance problem by increasing the number of minority samples. For example, synthetic minority over-sampling technique (SMOTE) [17] creates new minority samples by oversampling. Intra-class mixup [18] enhances model attention to minority samples by interpolating between minority and majority samples. However, these methods heavily depend on the quality of the minority sample. (ii) *Distribution squeezing* promotes the compactness of distribution by squeezing the feature space of normal samples. For example, OE-based [8] or contrastive learning-based [19] methods construct proxy anomalies to compress normal sound distributions. Focal loss (FL) [20], on the other hand, focuses directly on low probability minority samples. However, contrastive learning relies heavily on data augmentation strategies and faces challenges in optimization, while FL is sensitive to hyperparameters and exhibits unstable performance. (iii) *Prototype alignment* method captures the diversity of normal samples by learning prototypes [21], which can be effectively implemented by constructing multiple subclass centers based on angular margin loss (AML) [22, 23]. For instance, SCAdacos [10] introduces multiple subcenters for normal class, enabling samples to be represented by multiple prototypes. This method is widely used for ASD tasks because it is efficient and easy to implement [24, 25]. In practical applications, these methods are often used in combination, which can significantly improve the detection performance [26].

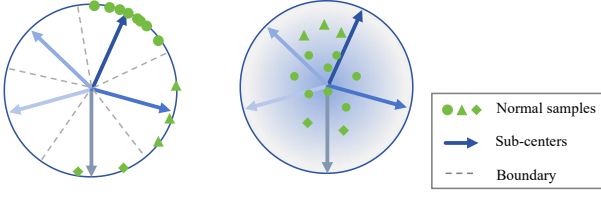


Figure 1: Two existing representation strategies for SCAML. Left: optimal subclass representation. Right: joint subclass representation.

2.2. Subcenter angular margin loss (SCAML)

The AML enhances traditional softmax in several key ways. Firstly, it normalizes both the weight vectors and the feature embeddings ($\|w_j\| = \|h\| = 1$) and setting $b_j = 0$ [22]. Here, $h \in \mathbb{R}^d$ is the sample x embedding, w_j is the j -th column of the weight matrix $W \in \mathbb{R}^{d \times C}$, b_j is the bias term and C is the class number. This transforms the classification problem from Euclidean distance to angular similarity, focusing on the angle θ_j between the weight vector w_j and the feature h . Secondly, AML introduces an angular margin function $f(s, \cos(\theta_i))$ to impose an additional angular penalty on the decision margin between classes [23]. This function enhances the inter-class distinction and improves the compactness of the intra-class distribution by explicitly penalizing the angular differences between classes. The angular margin loss can be formulated in a general form:

$$\mathcal{L}_{AML} = -\log \frac{e^{f(s, \cos(\theta_i))}}{e^{f(s, \cos(\theta_i))} + \sum_{j=1, j \neq i}^C e^{f(s, \cos(\theta_j))}} \quad (1)$$

For AdaCos [27] and SCAdaCos [10], the $f(s, \cos(\theta_i))$ defined as:

$$f(s, \cos(\theta_j))_{AdaCos} = s(t) \cos(\theta_j) \quad (2)$$

where s is the scaling factor, t is the number of iterations, AdaCos adjusts the angular margin penalty according to the iterative process.

Since the diversity of target class samples, only one class center is not enough to represent efficiently, therefore, SCAML is introduced, e.g., SCARCFace [11], SCAdaCos [10]. SCAML creates multiple subclass centers for each class, the class center matrix expands from $d \times C$, to $d \times C \times S$, where S is the number of subcenters in each class.

There are two existing representation strategies for SCAML, as shown in Fig. 1. The optimal subclass representation strategy represents the current sample by measuring the distance of the sample from each subcenter and using a nearest subcenter to represent the current sample [11]. This makes the sample embedding optimally aligned to its closest prototype, and is equivalent to creating distributions for each subclass. The joint subclass representation uses all subcenters to represent the sample, which builds a distribution that captures the complex patterns in the sample [10]. However, due to the imbalance of normal samples, majority samples easily overfit on this distribution [15]. Therefore, both strategies have their limitations in addressing the imbalance problem in ASD tasks, which motivates us to explore a more robust multi-subcenter representation approach in the next section.

3. Proposed method

3.1. Principles of multi-center representation

The optimal subclass representation strategy is a direct extension of the original AML, which has been discussed in [23]. In this paper, we focus on joint subclass representation strategy. Assume that for each class i , there exist S_i subclass centers, denoted as $w_i^1, w_i^2, \dots, w_i^{S_i}$. The goal of the multi-center representation is to align samples with multiple subclasses within the normal class. Thus, given a sample x and its class index i , the loss function \mathcal{L}_{SCAML} is defined as follows:

$$\mathcal{L}_{SCAML} = -\log \frac{\sum_{m=1}^{S_i} e^{sh^\top w_i^m}}{\sum_{j=1}^C \sum_{m=1}^{S_j} e^{sh^\top w_j^m}} \quad (3)$$

where the numerator is the sum of the exponents of the sample x belonging to all subcenters under its class i , m is the index of the subcenter, and the denominator is the sum of the exponents of all subcenters under all classes, j is the class index.

P_i is defined as the probability that sample x belongs to class i , calculated as the sum of the probabilities of x belonging to each existing subcenter within class i :

$$P_i = \sum_{m=1}^{S_i} p_i^m = \sum_{m=1}^{S_i} \frac{e^{sh^\top w_i^m}}{\sum_{j=1}^C \sum_{m'=1}^{S_j} e^{sh^\top w_j^{m'}}} \quad (4)$$

Thus, the loss function can be simplified to:

$$\mathcal{L}_{SCAML} = -\log P_i \quad (5)$$

Compute the gradient of the loss function \mathcal{L}_{SCAML} with respect to the feature h :

$$\nabla_h \mathcal{L}_{SCAML} = -\frac{1}{P_i} \nabla_h P_i \quad (6)$$

Further expanding $\nabla_h P_i$:

$$\nabla_h P_i = s \nabla_h \left(\sum_{m=1}^{S_i} p_i^m \right) = s \sum_{m=1}^{S_i} \nabla_h p_i^m \quad (7)$$

According to the gradient formula of probability:

$$\nabla_h p_i^m = s(p_i^m w_i^m - p_i^m \sum_{j=1}^C \sum_{m'=1}^{S_j} p_j^{m'} w_j^{m'}) \quad (8)$$

Therefore:

$$\nabla_h P_i = s \left(\sum_{m=1}^{S_i} p_i^m w_i^m - P_i \sum_{j=1}^C \sum_{m'=1}^{S_j} p_j^{m'} w_j^{m'} \right) \quad (9)$$

Substituting the above expression into the Eq. (6), the gradient $\nabla_h \mathcal{L}_{SCAML}$ is:

$$\begin{aligned} \nabla_h \mathcal{L}_{SCAML} &= -\frac{s}{P_i} \left(\sum_{m=1}^{S_i} p_i^m w_i^m - P_i \sum_{j=1}^C \sum_{m'=1}^{S_j} p_j^{m'} w_j^{m'} \right) \\ &= s \left(\sum_{j=1}^C \sum_{m'=1}^{S_j} p_j^{m'} w_j^{m'} - \frac{\sum_{m=1}^{S_i} p_i^m w_i^m}{P_i} \right) \end{aligned} \quad (10)$$

The first term, $\sum_{j=1}^C \sum_{m'=1}^{S_j} p_j^{m'} w_j^{m'}$, represents the weighted sum of all subclass centers, where $p_j^{m'}$ denotes the

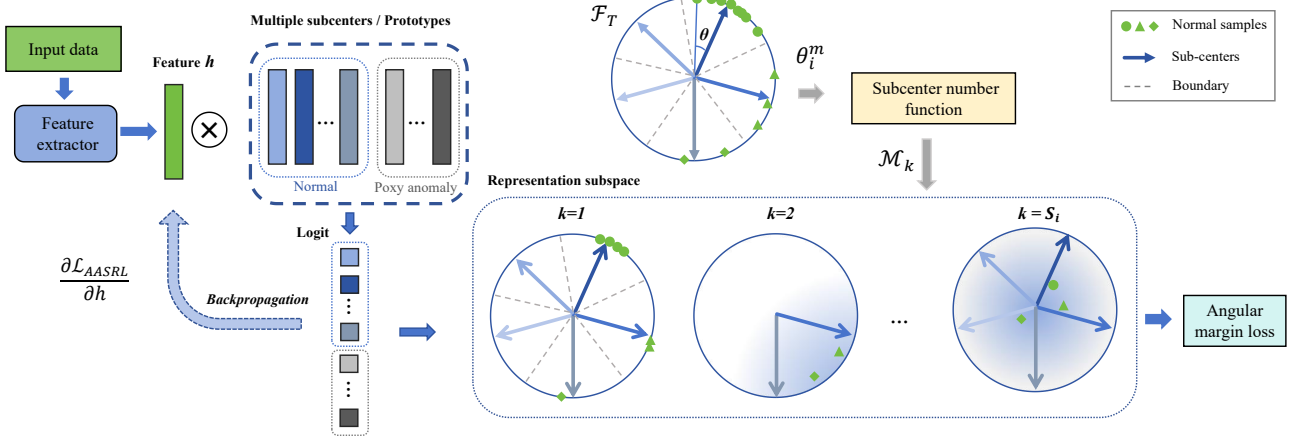


Figure 2: The schematic diagram of AASRL. \mathcal{F}_T is the model in representation quality measurement phase. \mathcal{M}_k is the subcenter set for sample representation.

probability of the sample x belonging to $w_j^{m'}$, the m' -th subclass of class j . The second term, $\frac{\sum_{m=1}^{S_i} p_i^m w_i^m}{P_i}$, represents the weighted average of the subclass centers under the class i to which the sample x belongs, normalized by the total probability P_i of sample x belonging to class i . Therefore, the gradient $\nabla_h \mathcal{L}_{SCAML}$ drives the sample x to approach the weighted average of the subclass centers of its corresponding class i , while also considering the overall structure of all subclass centers. This ensures that the sample x is aligned with the subclass centers of its class without deviating from the global weighted structure.

3.2. Adaptive across-subcenter representation learning for imbalanced ASD

From Eq. (10), the contribution of each subclass center w_i^m to the gradient is weighted by its corresponding probability p_i^m ($p_i^m \geq 0$). Consider two subclass centers w_i^1, w_i^2 with completely opposite orientations. Without probability weighting, their gradient contributions would cancel each other out. However, since the probability values p_i^1 and p_i^2 are mutually exclusive (i.e., they sum to 1), the probability weighting mechanism effectively prevents gradient conflicts. Consequently, increasing the number of subclass centers does not lead to gradient conflicts, as the influence of each subcenter is proportionally adjusted by its probability. This provides a theoretical basis for steadily improving the representation capability by increasing the number of subclass centers.

According to [16], samples that are less well-represented should utilize more subclass centers to capture their complex patterns. When using the optimal subclass representation strategy, the angle between a sample and its nearest subcenter can reflect the representation quality of a sample in the overall distribution [28, 29]. Therefore, AASRL initially employs the optimal subclass representation strategy to train the model with T iterations. This captures intra-class diversity and provides a basis for measuring the representation quality for each sample. The set of angles θ_i^m between the sample x of class i and each sub-center w_i^m can be obtained as follows:

$$\theta_i^m = \arccos \left(\frac{\mathcal{F}_T(x) \cdot w_i^m}{\|\mathcal{F}_T(x)\| \cdot \|w_i^m\|} \right) \quad (11)$$

where \mathcal{F}_T is the model after T iterations.

A subcenter number function is designed according to the sample representation quality:

$$\alpha = \begin{cases} 1 - \cos \theta & \theta \leq 90^\circ \\ 1 & \theta > 90^\circ \end{cases} \quad (12)$$

$$k = 1 + \lfloor (S_i - 1) \cdot \alpha^\tau \rfloor \quad (13)$$

where $\theta = \min(\theta_i^m)$, τ is a temperature parameter, $\tau > 0$, which controls the speed of subcenter expansion, with larger values expanding more slowly.

Lastly, the k most similar subclass centers $\{w_i^m \mid m \in \mathcal{M}_k\}$ are selected as the final representation subspace for the current sample:

$$\mathcal{M}_k = \underset{m \in \{1, 2, \dots, S_i\}}{\operatorname{arg\,min}} \theta_i^m \quad (14)$$

The schematic diagram of AASRL is shown in Fig. 2. In AASRL, samples are adaptively represented by k ($k \leq S^i$) subcenters, rather than by all S^i subcenters. This selective approach prevents majority samples from dominating the representation space and ensures that minority samples are also effectively represented.

4. Experimental results

4.1. Experimental setup

We conducted experiments on two datasets: (i) The *DCASE2023 Challenge Task 2 dataset* consists of real-world samples collected from 14 device types. The samples in this dataset naturally exhibit imbalance. Details can be found in [30]. (ii) The test set of the first dataset is balanced in the source and target domains, but the imbalance within subclasses makes the test conditions insufficient to fully meet our research objectives. To better evaluate model performance under imbalanced conditions, we constructed an *Imbalanced dataset*. This dataset was created by replacing the Toy car and Toy train data in the first dataset with samples from ToyADMOS2 [31]. Using the tools¹ provided by ToyADMOS2, a total of 1000 training and 200 test samples were obtained for each type, with each type containing 10 configurations (using mic1 only). The device speed label was set to be unobtainable during training to

¹<https://github.com/nttclab/ToyADMOS2-dataset>

simulate intra-class imbalance. For training, the ratio of “speed level 1:speed level 2” was set to 3:2 for Toy Train and 4:1 for Toy Car, while the ratio was 1:1 for testing. The data quality was set at 0 dB.

Evaluation metrics are the area under the receiver operating characteristic (ROC) curve (AUC) and partial AUC (pAUC) with $p = 0.1$. pAUC is computed under the low false-positive rate range $[0, p]$ to prioritize high true positive rates and avoid frequent false alarms.

The baseline system adopted SSLASD [26], an OE-based system that integrates several self-supervised learning methods and achieves outstanding performance. The network architecture and training settings for all experiments were aligned with the baseline to ensure manageable comparison results. For SCAML, each class has 16 subclass centers. To ensure a fair comparison, all methods were trained for 10 epochs.

4.2. Performance comparison with existing methods

We first compared different types of imbalanced ASD methods on the DCASE2023 Challenge Task2 dataset, based on the model of [32]. Then, three different subclass representation strategies, including optimal (“**”)/joint (*baseline*) subclass representation, proposed AASRL ($\tau = 0.5$, $T = 1200$) are performed on both datasets. In addition, to avoid interference from the performance of other devices, only the results of Toy train and Toy car in the imbalance dataset are considered.

Table 1: Harmonic mean of AUC (%) and pAUC (%) of different methods on two Datasets. Results are the mean and standard deviation from five independent experiments.

Method	AUC(%)	pAUC(%)
DCASE2023 Task2 Dataset		
OE [6]	64.61 ± 0.73	54.19 ± 0.58
+FL [20]	64.81 ± 0.41	55.01 ± 0.93
+SMOTE [17]	65.27 ± 1.06	55.39 ± 0.75
+mixup(intra-calss) [18]	66.21 ± 0.51	56.34 ± 0.32
+SCAdaCos [10]	69.57 ± 0.82	57.45 ± 0.84
+SCAdaCos* [11]	68.94 ± 0.52	56.95 ± 0.54
+AASRL	70.31 ± 0.21	57.77 ± 0.71
SSLASD (<i>baseline</i>) [26]	72.86 ± 0.82	59.39 ± 0.74
SSLASD (SCAdaCos*) [11]	72.16 ± 0.91	58.59 ± 0.87
SSLASD (AASRL)	73.91 ± 0.59	60.01 ± 0.64
Imbalanced Dataset		
SSLASD (<i>baseline</i>) [26]	67.82 ± 0.76	55.65 ± 0.62
SSLASD (SCAdaCos*) [11]	68.91 ± 0.85	55.88 ± 0.51
SSLASD (AASRL)	70.89 ± 0.47	57.26 ± 0.53

The experimental results are shown in Table 1. The proposed AASRL achieves the highest performance. On the DCASE2023 Task2 Dataset, AASRL shows improvements of 1.05% in AUC and 0.62% in pAUC compared to the baseline. The enhancements are more pronounced on the Imbalanced dataset, with AUC and pAUC improvements reaching 3.07% and 1.61%, respectively.

4.3. Hyperparameter impact on performance

The number of iterations T affects the accuracy of the sample representation quality assessment, and the temperature coefficients τ in the subclass number function determine the speed of subclass expansion. A sensitivity analyses were conducted by alternatively fixing each parameter, the results are shown in Fig. 3. It can be concluded that the first phase of AASRL requires a certain number of iterations to establish the accuracy

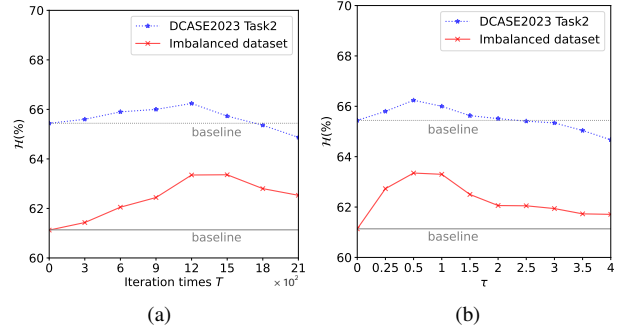


Figure 3: Harmonic mean of combined AUC and pAUC \mathcal{H} (%) on DCASE2023 Task2 and Imbalanced datasets by varying the hyperparameters value. (a) τ fixed at 0.5; (b) T fixed at 1200.

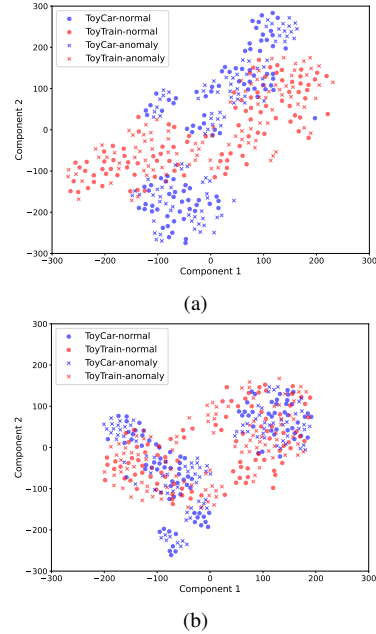


Figure 4: The t-SNE plots of the test set in the Imbalanced dataset: (a) baseline, (b) SSLASD (AASRL).

of the sample representation quality assessment. In addition, a faster subclass expansion speed is beneficial to the performance improvement.

4.4. Visualization

As can be seen from Fig. 4, compared to the baseline method, AASRL can better capture the similarities of normal data, making them more clustered in the feature space.

5. Conclusion

This paper proposed AASRL, a novel method that adaptively increases the number of subcenters based on the representation quality of each sample. AASRL effectively expands the representation space for under-represented samples while mitigating overfitting to majority samples, thereby achieving superior performance compared to existing methods with a fixed number of subcenters. The effectiveness of AASRL is demonstrated through experiments on the DCASE2023 Challenge Task2 dataset and a constructed imbalanced dataset.

6. References

- [1] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda *et al.*, “Description and discussion on dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring,” *arXiv preprint arXiv:2006.05822*, 2020.
- [2] V. Zavrtanik, M. Marolt, M. Kristan, and D. Skočaj, “Anomalous sound detection by feature-level anomaly simulation,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2024, pp. 1466–1470.
- [3] 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 Transactions on Emerging Topics in Computational Intelligence*, 2024.
- [4] Y. Zhou and Y. Long, “Attribute classifier with imbalance compensation for anomalous sound detection,” *Tech. Rep., DCASE2023 Challenge*, 2023.
- [5] P. Primus, V. Haunschmid, P. Praher, and G. Widmer, “Anomalous sound detection as a simple binary classification problem with careful selection of proxy outlier examples,” *Tech. Rep., DCASE2020 Challenge*, 2020.
- [6] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” *Proceedings of the International Conference on Learning Representations*, 2019.
- [7] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Improvement of serial approach to anomalous sound detection by incorporating two binary cross-entropies for outlier exposure,” in *European Signal Processing Conference*. IEEE, 2022, pp. 294–298.
- [8] Y. Tachioka, “Outlier exposure with efficient division of positive and negative examples for anomalous sound detection,” in *European Signal Processing Conference*. IEEE, 2024, pp. 76–80.
- [9] A. Almuđévar, A. Ortega, L. Vicente, A. Miguel, and E. Lleida, “Variational classifier for unsupervised anomalous sound detection under domain generalization,” in *Proceedings of INTERSPEECH*, 2023, pp. 2823–2827.
- [10] K. Wilkinghoff, “Sub-cluster adacos: Learning representations for anomalous sound detection,” in *2021 International Joint Conference on Neural Networks*. IEEE, 2021, pp. 1–8.
- [11] J. Deng, J. Guo, T. Liu, M. Gong, and S. Zafeiriou, “Sub-center arcface: Boosting face recognition by large-scale noisy web faces,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer, 2020, pp. 741–757.
- [12] J. Deng, J. Guo, J. Yang, A. Lattas, and S. Zafeiriou, “Variational prototype learning for deep face recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 906–11 915.
- [13] M. Zhu and A. M. Martinez, “Subclass discriminant analysis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1274–1286, 2006.
- [14] H. Wan, H. Wang, G. Guo, and X. Wei, “Separability-oriented subclass discriminant analysis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 2, pp. 409–422, 2017.
- [15] M. Hayaeian Shirvan, M. H. Moattar, and M. Hosseinzadeh, “Deep generative approaches for oversampling in imbalanced data classification problems: A comprehensive review and comparative analysis,” *Applied Soft Computing*, vol. 170, p. 112677, 2025.
- [16] H. Yan, Y. Qian, F. Peng, J. Luo, F. Li *et al.*, “Neural collapse to multiple centers for imbalanced data,” in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [18] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Two-stage anomalous sound detection systems using domain generalization and specialization techniques,” *Proc. DCASE*, 2022.
- [19] H. Hojjati and N. Armanfard, “Self-supervised acoustic anomaly detection via contrastive learning,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2022, pp. 3253–3257.
- [20] T.-Y. Ross and G. Dollár, “Focal loss for dense object detection,” in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2980–2988.
- [21] X.-M. Zeng, Y. Song, I. McLoughlin, L. Liu, and L.-R. Dai, “Robust prototype learning for anomalous sound detection,” in *Proceedings of the 24th Annual Conference of the International Speech Communication Association*. Dublin, Ireland: ISCA, August 2023, pp. 261–265.
- [22] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.
- [23] K. Wilkinghoff and F. Kurth, “Why do angular margin losses work well for semi-supervised anomalous sound detection?” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.
- [24] J. Wu, F. Yang, and W. Hu, “Unsupervised anomalous sound detection for industrial monitoring based on arcface classifier and gaussian mixture model,” *Applied Acoustics*, vol. 203, p. 109188, 2023.
- [25] S. Choi and J.-W. Choi, “Noisy-arcmix: Additive noisy angular margin loss combined with mixup for anomalous sound detection,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2024, pp. 516–520.
- [26] K. Wilkinghoff, “Self-supervised learning for anomalous sound detection,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2024, pp. 276–280.
- [27] X. Zhang, R. Zhao, Y. Qiao, X. Wang, and H. Li, “Adacos: Adaptively scaling cosine logits for effectively learning deep face representations,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 823–10 832.
- [28] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang, “Curricularface: adaptive curriculum learning loss for deep face recognition,” in *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5901–5910.
- [29] W. Hu, Y. Huang, F. Zhang, and R. Li, “Noise-tolerant paradigm for training face recognition cnns,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 887–11 896.
- [30] 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 *Detection and Classification of Acoustic Scenes and Events*, Tampere, Finland, September 2023, pp. 21–22.
- [31] 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*, Barcelona, Spain, November 2021, pp. 1–5.
- [32] K. Wilkinghoff, “Design choices for learning embeddings from auxiliary tasks for domain generalization in anomalous sound detection,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2023, pp. 1–5.