CP-JKU SUBMISSION TO DCASE'21: IMPROVING OUT-OF-DISTRIBUTION DETECTORS FOR MACHINE CONDITION MONITORING WITH PROXY OUTLIERS & DOMAIN ADAPTATION VIA SEMANTIC ALIGNMENT

Technical Report

Paul Primus[1]*, Martin Zwifl, and Gerhard Widmer*[1,2]

[1]Institute of Computational Perception (CP-JKU)
[2]LIT Artificial Intelligence Lab
Johannes Kepler University, Austria

ABSTRACT

This technical report contains a detailed summary of our submissions to the *Unsupervised Anomalous Sound Detection under Domain Shifted Conditions* Task for Machine Condition Monitoring of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events 2021 (DCASE). Our goal was to learn out-of-distribution (OOD) detectors without access to OOD data, i.e., we trained only on recordings of undamaged machines. To this end, we employed a range of popular unsupervised anomaly detection methods based on auxiliary classification, density estimation, and reconstruction error. OOD detectors were trained for each of the seven machine categories included in the development dataset. We then showed that the OOD detectors' performance was enhanced by utilizing metadata labels and other machines' regular sounds as proxy outliers. To further improve detection performance under domain-shifted conditions, we fine-tuned the auxiliary classifiers to semantically align the hidden representations of source and target domain, using the limited target domain data. In addition to this technical description, we release our complete source code to make our submission fully reproducible [1].

Index Terms— Machine Condition Monitoring, Out-of-Distribution Detection, Domain Adaptation, DCASE2021

1. INTRODUCTION

Out-Of-Distribution (OOD) detection aims to detect data points that deviate from a pre-defined set of typical examples. The general framework of OOD detection has many applications, such as monitoring the health status of a patient or detecting fraudulent financial transaction patterns. In this work, we deal with Unsupervised Machine Condition Monitoring (MCM) for the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events 2021 (DCASE) [1]. The main objective of this task was to learn a model from typical machine sounds capable of detecting machine breakdowns.

To develop such a model, the task organizers provided a development dataset containing typical recordings of seven different machine types - fan, pump, slider, valve, toy car, and toy train [2, 3]. The dataset contains six different machine instances per machine type, and each recording in this dataset is labeled the machine's identity (labels 1-6; similar to a serial number). Additionally to the development set, a test set with normal and anomalous sounds for all machine identities and all machine types was provided for model selection and submission ranking. The anomaly labels of test recordings were only given for machine instances 1-3 of each machine type. The test recordings of the remaining three machine instances per machine type (4-6) will be used to rank submissions, and thus no OOD labels were provided. Participants were not permitted to use the test data for training.

Another major challenge of this year's task is the presence of a domain shift in the test data, meaning that a part of the test data was not recorded under the same conditions as the training data. These condition shifts include differences in operation speed, machine load, environmental noise, etc. The authors provided three examples of undamaged machines per domain shift in the development set for adjusting the OOD detectors. There are six possible domain shifts per machine type.

This technical report has been organized in the following way: First, we will describe the methods used for anomaly detection and domain adaptation. Then we will give a detailed account of the conducted experiments. Finally, we will briefly state the results and explain the systems used to obtain the submitted predictions.

2. PROPOSED SYSTEMS

We will use $x \in \mathcal{X}$ to denote inputs to the model and $y \in \{1, \dots C\}$ the denote corresponding machine identity labels ($C = 6$). We use superscripts to indicate whether samples are drawn from the in-distribution or the proxy outlier distribution, i.e., $\mathcal{D}^{in}$ and $\mathcal{D}^{PO}$, respectively. Furthermore, we will use subscripts to indicate whether samples are from the source or target domain, i.e., $\mathcal{D}_s$ and $\mathcal{D}_t$, respectively.

2.1. Auxiliary Classification Task

Prior work [4] found that classifiers predicts in-distribution examples more confidently than OOD examples and, hence, use the negative maximum softmax probabilities of an auxiliary classifier as anomaly scores.

$$\mathcal{A}_{\text{nmp}}(x, y) \coloneqq -\max_y f_y(x) \tag{1}$$

[1]https://github.com/OptimusPrimus/dcase2021_task2

Here f is an auxiliary classifier and f_y is the predicted class probability for class y. Potential prediction targets for such auxiliary classifiers are metadata labels; other approaches use self-supervised prediction targets such as the degree of augmentation performed on the input [5].

For our submission, we trained separate auxiliary classifiers per machine type to predict a machine's identity from Mel-spectrograms. We trained them to minimize the cross entropy loss $\mathcal{L}_{\text{clf}}$:

$$\mathcal{L}_{\text{clf}} = \mathbb{E}_{(x,y)\sim\mathcal{D}^{in}}[\ -\log f_y(x)\ ]$$

Since the actual machine identity labels are given in both the development and evaluation set, we used the negative probability of the true class instead of Equation 1 as anomaly score:

$$\mathcal{A}_{\text{ntp}}(x,y) \coloneqq -f_y(x)$$

A major advantage of using auxiliary classifiers is that it allowed us to build upon existing systems for audio classification, such as [6].

We also augmented the machine recordings by convexly combining random pairs of raw audio recordings and their one-hot encoded labels. This procedure is similar to Mix-Up [7] and was proposed as a method to synthesize new machine identities for MCM by [8]. As this method did not enhance the detection performance in general, we only applied it to the OOD detector for valves.

2.2. Density Estimation

Density estimators construct a model of the true underlying probability density $p(x)$ from in-distribution data and thus offer a natural way to assess the typicality of examples via log probability estimates. A straightforward way to employ density models for OOD detection is to utilize the negative log probabilities as anomaly scores. In our work, we employed Masked Autoencoder for Distribution Estimation (MADE) [9] and Masked Autoregressive Flows (MAF) [10] because they allow for rapid evaluation of log probabilities. We again trained one density model for each of the seven machine types. Motivated by [11], we additionally made use of the available labels by conditioning the density models on the machine identity, because we hoped that additional information might lead to more precise probability estimates. We trained our models to maximize the log likelihood on the normal data:

$$\mathcal{L}_{\text{density}} = -\mathbb{E}_{(x,y)\sim\mathcal{D}^{in}}[\ \log p(x \mid y)\ ]$$

Furthermore, we used the negative log probabilities as anomaly scores:

$$\mathcal{A}_{\text{denstiy}}(x,y) \coloneqq -\log p(x \mid y)$$

2.3. Reconstruction Error

Reconstruction-error-based OOD methods are trained to reconstruct the in-distribution data from a compressed representation. These methods assume that the models will not generalize to novel patterns and thus fail to reconstruct OOD examples. Our reconstruction-error-based system builds upon the DCASE2021 autoencoder baseline method and enhances it by conditioning the autoencoders on the machine identity. We trained our autoencoders g to minimize the reconstruction error:

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{(x,y)\sim\mathcal{D}^{in}}[\ \|x - g(x,y)\|^2\ ]$$

Anomaly score were computed based on the reconstruction error:

$$\mathcal{A}_{\text{rec}}(x,y) \coloneqq \|x - g(x,y)\|^2$$

2.4. Proxy Outlier Loss

Recent work [12] proposed leveraging auxiliary databases not related to the anomaly detection task at hand to improve the performance of various existing anomaly detection systems, a method that was coined *outlier exposure*. Analogously, we found that training binary classifiers to distinguish between *proxy outliers* and normal machine sounds considerably outperforms density estimation and reconstruction error baselines [13]. This work will show that similar strategies can be adopted to improve the previously introduced OOD methods for MCM with proxy outliers. As all of our models were trained for a specific machine type, the normal sounds of all other machines can be used as proxy outliers.

Similar to [12], we applied distinct strategies to integrate proxy outliers in the training procedure of the previously proposed OOD detectors. For the auxiliary classification-based methods, we encouraged the classifier f to provide uncertain predictions for proxy outlier examples by enforcing a close-to-uniform class probability distribution via the cross-entropy loss H:

$$\mathcal{L}_{\text{PO,clf}} = \mathbb{E}_{(x)\sim\mathcal{D}^{PO}}[\ H(\mathcal{U}, f(x))\ ]$$

where $\mathcal{U}$ is the uniform distribution.

We used a margin ranking loss for the density estimation and reconstruction-error-based models to ensure that the proxy outliers receive a lower log probability and a higher reconstruction error, respectively. We compute the margin loss between a normal sample (x, y) and a proxy outlier (x', y') via:

$$\text{margin}(x,y,x',y') = \max\{0, m + \mathcal{A}(x',y') - \mathcal{A}(x,y)\}$$

The margin ranking loss is then computed as follows:

$$\mathcal{L}_{\text{PO,rank}} = \mathbb{E}_{(x,y)\sim D^{\mathbf{in}}}[\ \mathbb{E}_{(x',y')\sim\mathcal{D}^{PO}}[\ \text{margin}(x,y,x',y')\ ]\ ]$$

As these methods require a conditioning label y' for the proxy outliers, we randomly choose one from $\{1,\ldots,6\}$. Additionally, we expect the conditional models to give a lower log probability or higher reconstruction error when conditioned on the wrong machine identity. However, this is not the case in general, as we will show in the experiment section, and we, therefore, enforced this property by adding in-distribution examples with randomly altered machine identity labels to the proxy outlier set.

When jointly training the OOD detectors with their primary loss and outlier exposure, we use a weight λ to control the influence of the proxy outliers.

2.5. Contrastive Semantic Alignment

A major challenge of this year's MCM task was the domain shift between training and test data and the limited target domain examples available for training. The method we employ is based on the contrastive semantic alignment loss proposed by Motiian et al. [14].

In particular, we fine-tuned our auxiliary classifiers to semantically align the hidden representations ϕ of source and target domain by minimizing the pairwise distance between examples of the same class but from different domains:

$$\mathcal{L}_{sem} = \sum_{y=1}^{C} \mathbb{E}_{(x,x')\sim\mathcal{D}_{\mathrm{s}|y},\mathcal{D}_{\mathrm{t}|y}} \left[\; \left\| \phi(x) - \phi(x') \right\|^2 \; \right]$$

$\mathcal{D}_{\mathrm{s}|y}$ and $\mathcal{D}_{\mathrm{t}|y}$ are source and target distribution for a specific machine identity y. Simultaneously, we enforced a margin between the hidden representation of examples from different classes and different domains to preserve class separability.

$$\mathcal{L}_{cont} = \sum_{y=1}^{C} \mathbb{E}_{(x,x')\sim\mathcal{D}_{\mathrm{s}|y},\mathcal{D}_{\mathrm{t}|\neg y}} \left[\; \mathtt{max}\{0, n + \left\| \phi(x') - \phi(x) \right\|\}^2 \; \right]$$

$\mathcal{D}_{\mathrm{t}|\neg y}$ are samples from the target domain which are not form class y. For fine-tuning, we created a convex combination of the domain adaptation loss and the previously introduces losses:

$$\mathcal{L}_{\mathrm{ft}} = \alpha \cdot \left(\mathcal{L}_{\mathrm{clf}} + \lambda \cdot \mathcal{L}_{\mathrm{PO,clf}} \right) + (1 - \alpha) \cdot \left(\mathcal{L}_{sem} + \mathcal{L}_{cont} \right)$$

3. EXPERIMENTS

We conducted a series of experiments to determine the best configuration for our submission. First, we trained all previously mentioned OOD detection methods (Sections 2.1, 2.2 and 2.3) for each machine type and without outlier exposure to establish a baseline. Then, we included proxy outliers in the training procedure as described in section 2.4. For the auxiliary-classification-based OOD method, we used the sounds of other machines as proxy outliers. For the conditional models (density estimators and autoencoders), we conducted experiments with two proxy outlier sets: The first one contained only sounds of the same machine with false machine identity labels. The second proxy outlier set additionally contained the sounds of all other machines. In our final experiment, we fine-tuned the best auxiliary classifier for each machine type to account for the domain shift in the test data. The following sections give a detailed account of the network architectures, audio prepossessing, and training procedure.

3.1. Network Architecture

As auxiliary classifier, we chose the model architecture introduced by Koutini et al. [6], a receptive-field-regularized, fully convolutional, residual network (ResNet) [15], which has been successfully adopted for various audio-related classification tasks [16, 17]. We tuned the receptive field such that the initial anomaly detection performance across all machine types without outlier exposure was maximized. The exact architecture of the ResNet can be found in our GitHub repository.

Our implementation of MAF and MADE is based on a public GitHub repository[2]. Based on the results reported in [11], we used MAFs with four autoregressive blocks, one hidden layer per block and 2048 units per hidden layer. Similarly, we use MADEs with four hidden layers and 2048 units per layer.

The conditional autoencoder matches the architecture of the DCASE autoencoder baseline [1]. We condition each layer of the autoencoder on the machine identity label by adding the output of a learnable linear projection of the one-hot encoded label to the output after applying the non-linearity.

3.2. Audio Preprocessings

Following the DCASE 2021 Challenge task 2 baseline system [1], we re-sampled the audio signals to 16000Hz and computed a mono-channel Short Time Fourier Transform using 1024-sample windows and a hop-size of 512 samples. We weighted the resulting power spectrogram with a Mel-scaled filterbank of 128 filters and applied the logarithm to dampen large outliers.

3.3. Training

We trained the auxiliary classification on entire 10-second audio clips; the density estimators and the autoencoder were trained on 5-frame snippets. All models were trained only on the source domain data with the Adam update rule [18] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We used batch updates with 14 and 512 examples for the auxiliary classifiers and other methods, respectively. We doubled the batch size for training with proxy outliers and stratified the batches to contain equal numbers of in-distribution and OOD samples. During one epoch, we iterate over all normal samples; proxy outliers were drawn randomly. We set the initial learning rate to 10^{-4}, kept it constant for 30 epoch, linearly decayed it to 10^{-5} over a period of 60 epochs, and continued training for ten more epochs. To weight the PO loss, we use $\lambda = 0.5$ for the auxiliary-classifier based methods and $\lambda = 1.0$ for the margin ranking loss based methods. The minimal margin is set to $m = 0.5$.

For fine-tuning the auxiliary classifier, we randomly sampled batches of 14 target domain examples and resumed training for three epochs with the learning kept steady at 10^{-5}. We apply the congestive semantic loss on the hidden representations after the last stage. $\alpha = 0.9$ and $n = 2.0$ were chosen based on grid search.

3.4. Prediction

To obtain the anomaly score from the auxiliary classifiers, we inserted the whole 10-second audio segments and used the negative probability of the true class as an anomaly score. For the other OOD detectors, we first obtained the anomaly score for each 5-frame Mel-spectrogram segment with a hop size of 1 and then averaged the scores over the whole 10-second recording.

3.5. Evaluation

We evaluate our models with the Ω-score used for ranking submissions. The Ω-score is defined as the harmonic mean (hm) of the AUC and pAUC scores over all the machine types $\mathcal{M}$, machine identities C, and domains $\mathcal{D}$:

$$\Omega \coloneqq \mathrm{hm}(\{\mathrm{AUC}_{j,i,d}, \mathrm{pAUC}_{j,i,d} \mid j \in \mathcal{M}, i \in \mathcal{C}, d \in \mathcal{D}\}) \quad (2)$$

[2] https://github.com/kamenbliznashki/normalizing_flows

					fan		gearbox		pump		slider		toy car		toy train		valve	
Method	PO	Ω	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$	$\mathcal{D}_s$	$\mathcal{D}_t$
CLF	-	.648	.715	.593	.705	.645	.656	.626	.743	.676	.819	.580	.627	.596	.784	.482	.711	.589
CLF	all	.667	.750	.600	.839	.644	.676	.644	.760	.639	.811	.540	.646	.696	.780	.493	.778	.597
CLF-DA	all	.692	.747	.644	.831	.698	.699	.654	.763	.682	.800	.611	.648	.727	.724	.527	.801	.657
MAF	-	.581	.615	.550	.606	.575	.578	.629	.679	.569	.707	.542	.627	.529	.595	.513	.542	.514
MAF	same	.616	.663	.576	.695	.673	.639	.644	.699	.640	.739	.540	.663	.556	.697	.499	.545	.528
MAF	all	.615	.656	.578	.775	.710	.552	.569	.726	.623	.750	.574	.666	.580	.666	.496	.542	.537
MADE	-	.584	.612	.558	.611	.583	.587	.637	.669	.552	.696	.559	.629	.537	.594	.535	.530	.516
MADE	same	.618	.662	.580	.699	.667	.635	.644	.682	.611	.743	.546	.688	.575	.683	.518	.544	.531
MADE	all	.614	.652	.580	.771	.704	.549	.574	.714	.597	.752	.586	.663	.585	.652	.511	.542	.538
AE	-	.567	.589	.547	.614	.577	.569	.615	.629	.538	.652	.550	.571	.537	.599	.521	.511	.506
AE	same	.605	.641	.573	.707	.679	.594	.617	.670	.610	.712	.561	.644	.567	.671	.500	.530	.517
AE	all	.604	.636	.575	.774	.700	.531	.558	.706	.611	.730	.583	.621	.554	.641	.525	.534	.529
BL-MN	-	.582	.606	.559	.633	.615	.655	.607	.622	.573	.654	.521	.571	.558	.588	.507	.539	.551
BL-AE	-	.572	.600	.547	.591	.567	.564	.610	.638	.530	.670	.557	.594	.547	.643	.519	.524	.514

Table 1: Experiment results overview. The Ω-score is described in Section 3.5. Other reported scores are computed accordingly over subsets of the test set. Proxy outlier sets 'same' and 'all' are detailed in Section 3.

We report the results for machine instances $i \in \{1, 2, 3\}$ in the results section; scores for the remaining machine instances $i \in \{4, 5, 6\}$ will be used by the task organizers to rank the submissions.

4. RESULTS

A comprehensive overview of the results is given in Table 1. When comparing the conditional (AE) and unconditional autoencoders (Bl-AE), we observe no notable benefit of conditioning on the machine identity labels for anomaly detection. However, additionally using recordings from the same machine type with randomly altered labels as proxy outliers improved the performance in the source domain for every machine type compared across all conditional OOD methods. A similar trend can be observed in the target domain. Adding the remaining machines' normal sound to the proxy outlier set led to a notable increase in detection performance for some machine types (e.g. MAF fan) while worsening the detection performance for others (e.g. MAF gearbox).

We observe a similar trend for the auxiliary-classifier-based methods: While overall, the detection performance improved compared to the baseline when training with outlier exposure (see CLF fan), for some machines, the performance in the source domain worsend slightly (e.g. CLF toy train). Larger deteriorations can be observed in the target domain (e.g.. CLF pump).

After fine-tuning the auxiliary classifiers with the contrastive semantic alignment loss, the detection performance on the target domain test samples increased for all machine types (cf. Table 1, CLF-DA).

4.1. Submissions

Based on the previous results, our predictions for the secret test set were created using the following systems:

- **Submission 1 (MADE)**: We selected the best MADE model for each machine type according to the results in Table 1 and used them to create predictions in both the source and target domain.
- **Submission 2 (MAF)**: We picked the best MAF model for each machine type according to the results in Table 1 and used them to create predictions in both the source and target domain.
- **Submission 3 (ResNet)**: We choose the best auxiliary classifier for each machine type in the source domain according to the results in Table 1 and used them to create predictions in the source domain. We then fine-tuned these classifiers using the contrastive semantic alignment loss and used them to obtain predictions for the target domain.
- **Submission 4 (Ensemble)**: We normalize the outputs of all our systems for each machine instance separately by scaling and shifting the outputs with the mean and standard deviation computed over the anomaly scores of the training data. These normalized scores are then combined using a weighted average.

5. CONCLUSION

In this work, we used a variety of OOD detection methods to establish baselines and showed that outlier exposure improved the detection performance of these methods. Our experiments also showed that conditioning on machine identity labels does not necessarily yield better detection results. To solve this issue, we introduced an additional loss which ensures that in-distribution examples with wrong labels are ranked accordingly. Finally, to account for the domain shift present in training, we fine-tuned our OOD detectors based on auxiliary classifiers with a contrastive semantic domain alignment loss, which led to increased detection performance in the target domain. Finally, we submitted the outputs of four systems to the DCASE challenge, three based on single system predictions and a combination of the methods introduced in this work.

6. ACKNOWLEDGMENT

We thank Khaled Koutini for helpful discussion and for making the implementation of the receptive-field-regularized ResNet available. The LIT AI Lab is financed by the Federal State of Upper Austria.

7. 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 sound detection for machine condition monitoring under domain shifted conditions," *CoRR*, vol. abs/2106.04492, 2021. [Online]. Available: https://arxiv.org/abs/2106.04492

[2] 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," *CoRR*, vol. abs/2106.02369, 2021. [Online]. Available: https://arxiv.org/abs/2106.02369

[3] 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," *CoRR*, vol. abs/2105.02702, 2021. [Online]. Available: https://arxiv.org/abs/2105.02702

[4] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. [Online]. Available: https://openreview.net/forum?id=Hkg4TI9xl

[5] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, 2019, pp. 15 637–15 648. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/a2b15837edac15df90721968986f7f8e-Abstract.html

[6] K. Koutini, H. Eghbal-zadeh, M. Dorfer, and G. Widmer, "The receptive field as a regularizer in deep convolutional neural networks for acoustic scene classification," in *27th European Signal Processing Conference, EUSIPCO 2019, A Coruña, Spain, September 2-6, 2019*, 2019, pp. 1–5. [Online]. Available: https://doi.org/10.23919/EUSIPCO.2019.8902732

[7] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. [Online]. Available: https://openreview.net/forum?id=r1Ddp1-Rb

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

[9] M. Germain, K. Gregor, I. Murray, and H. Larochelle, "MADE: masked autoencoder for distribution estimation," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, ser. JMLR Workshop and Conference Proceedings, vol. 37. JMLR.org, 2015, pp. 881–889. [Online]. Available: http://proceedings.mlr.press/v37/germain15.html

[10] G. Papamakarios, I. Murray, and T. Pavlakou, "Masked autoregressive flow for density estimation," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 2017, pp. 2338–2347. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html

[11] V. Haunschmid and P. Praher, "Anomalous sound detection with masked autoregressive flows and machine type dependent postprocessing," DCASE2020 Challenge, Tech. Rep., July 2020.

[12] D. Hendrycks, M. Mazeika, and T. G. Dietterich, "Deep anomaly detection with outlier exposure," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. [Online]. Available: https://openreview.net/forum?id=HyxCxhRcY7

[13] 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," *CoRR*, vol. abs/2011.02949, 2020. [Online]. Available: https://arxiv.org/abs/2011.02949

[14] S. Motiian, M. Piccirilli, D. A. Adjeroh, and G. Doretto, "Unified deep supervised domain adaptation and generalization," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 2017, pp. 5716–5726. [Online]. Available: https://doi.org/10.1109/ICCV.2017.609

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[16] K. Koutini, H. Eghbal-zadeh, and G. Widmer, "Receptive-field-regularized CNN variants for acoustic scene classification," *CoRR*, vol. abs/1909.02859, 2019. [Online]. Available: http://arxiv.org/abs/1909.02859

[17] K. Koutini, S. Chowdhury, V. Haunschmid, H. Eghbal-zadeh, and G. Widmer, "Emotion and theme recognition in music with frequency-aware rf-regularized CNNs," *CoRR*, vol. abs/1911.05833, 2019. [Online]. Available: http://arxiv.org/abs/1911.05833

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. OpenReview.net, 2015. [Online]. Available: https://openreview.net/forum?id=8gmWwjFyLj