GROUP MASKED AUTOENCODER BASED DENSITY ESTIMATOR FOR AUDIO ANOMALY DETECTION

Ritwik Giri, Fangzhou Cheng, Karim Helwani, Srikanth V. Tenneti,
Umut Isik, Arvindh Krishnaswamy

Amazon Web Services, Palo Alto, CA, USA

ABSTRACT

In this paper, we address the problem of detecting previously unseen anomalous audio events, when the training dataset itself does not contain any examples of anomalies. While the traditional density estimation techniques, such as Gaussian Mixture Model (GMM) showed promise in past for the problem at hand, recent advances in neural density estimation techniques, have made them suitable for anomaly detection task. In this work, we develop a novel neural density estimation technique based on the Group-Masked Autoencoder, that estimates the density of an audio time series by taking into account the intra-frame statistics of the signal. Our proposed approach has been validated using the DCASE 2020 challenge dataset (Task 2 - *Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring*). We demonstrate the effectiveness of our approach by comparing against the baseline autoencoder model, and also against recently proposed Interpolating Deep Neural Network (IDNN) model.

Index Terms— Masked Autoencoder, Density Estimation, Anomaly Detection

1. INTRODUCTION

1.1. Background

Anomaly Detection is a problem of detecting anomalous data points that are significantly different from the normal operation data. In audio modality, anomaly detection has a number of important applications. For example, detecting unusual events using audio can nicely complement video based approaches. This is especially true in cases where there is not sufficient illumination, or in the presence of visual occlusions where the performance of video surveillance is impaired.

Most of the current literature we have seen on audio scene monitoring systems [1, 2], propose fully supervised learning methods: which require labeled examples of anomalous sounds.

In real-world scenarios, actual anomalous sounds rarely occur and have very diverse characteristics. Therefore, it is impossible to collect an exhaustive patterns of anomalous sounds to train a supervised anomaly detection system. This magnifies the need of an unsupervised audio anomaly detection system which can be used to detect unknown anomalous sounds, which have not been observed in the given training data.

1.2. Related Work

Unsupervised anomaly detection problem is typically solved by first learning a model or a probability distribution over normal data [3]. Then during testing, this pre-trained model is exploited to determine if a test sample belongs to the learned distribution/model or not. In the literature, several different model choices have been used, and majority of these approaches rely on a Deep Auto Encoder (AE) architecture. The main working principle of these approaches lies in training an AE using normal/expected data, and during testing checking if the network is struggling to decode the encoded test data accurately. I.e., if the system produces a high reconstruction error compared to some threshold, it is then considered anomalous input. In [4], the authors propose a denoising AE structure using both feedforward units and LSTM units for acoustic anomaly detection task. In [5], authors propose a convolutional AE on Mel-spectrograms to detect anomalies in the context of industrial plants and processes. Recently, in [6] authors propose a variant of AE architecture: Interpolating Deep Neural Network (IDNN), where the proposed model utilizes multiple frames of a spectrogram whose center frame is removed as an input, and it predicts an interpolation of the removed frame as an output. Anomalies can be detected based on an interpolation error, that is the difference between the predicted frame and the true frame. Authors show that, IDNN performs significantly better than a baseline AE for machine condition monitoring task, specially for non-stationary sounds.

Another line of work focuses on density estimation methods for anomaly detection. Unsupervised methods model the distribution of all the normal samples during training, and during inference, regard samples in the pdf regions with low probabilities as anomalies. In earlier works, Unimodal Gaussian [7], and Gaussian Mixture models [8] were used to model the normal data distribution. More recently, neural density estimation techniques, such as Normalizing Flow [9] have been used to solve the problem in hand. Neural density estimators can readily provide exact density evaluations unlike generative modeling approaches—such as Variational Auto-Encoder (VAE) [10, 11] and Generative Adversarial Networks (GANs) [12]. This makes them a popular choice for anomaly detection application, where we are more interested in evaluating exact densities during inference to detect anomalous points, rather than generating synthetic data.

1.3. Contribution

There are primarily two families of neural density estimators that are both flexible and tractable: autoregressive models [13] and normalizing flows [14]. Motivated from the success of Autoregressive models in modeling audio data [15, 16], in this work we focus on autoregressive model based density estimators. These models usually decompose the joint density as a product of conditionals over individual dimensions, and model each conditional as a parametric density where a neural network outputs the parameters of that density. Our work builds on one such approach named: Masked Autoencoder for Density Estimation (MADE) [18]. Like all autoregressive models, MADE is also very sensitive to the order of the variables. In this work we address this issue for audio anomaly detection task, by developing a novel Group Masked Autoencoder (Group MADE) architecture where the joint distribution can be decomposed as conditionals over groups/frames instead of individual dimensional conditional. We model each conditional by either a Gaussian distribution (Group MADE-G) or a mixture of Gaussians (Group MADE-MOG). We also show that, with suitable choice of group ordering, our approach can be interpreted as the probabilistic version of recently proposed state-of-the-art approach for audio anomaly detection: IDNN [6]. Finally, we demonstrate the effectiveness of our proposed approach by presenting extensive experimental results using the publicly available DCASE2020 Challenge Task 2 dataset [17].

The rest of the article is organized as follows: In Section 2, the proposed method is presented in detail. In Section 3, a brief description of challenge dataset, that have been used in this article is given. In Section 4, we present evaluation results of our proposed method and other competing methods over challenge dataset, and finally Section 5 concludes the paper and talks about some future research directions.

2. PROPOSED APPROACH

Our method builds on previous work on Masked Autoencoder for Distribution Estimation (MADE). We provide a brief description of MADE in the following subsection. More details about MADE can be found in original publication [18].

2.1. Masked Autoencoder for Distribution Estimation

In [18], authors propose a simple way of adapting an autoencoder architecture to develop a competitive and tractable neural density estimator. The key idea lies in masking the weighted connections between layers of a standard autoencoder to convert it into a tractable density estimator. Authors show that by designing appropriate masks, the output of the autoencoder satisfies the autoregressive property for a given ordering of inputs, i.e., each input dimension is reconstructed solely from the dimensions preceding it in the ordering. Multiple layers with non linearity can be added in this structure, which will result in a highly capable neural density estimator.

By using MADE, density of an input vector $\mathbf{x}$ is calculated by means of the decomposition according to the probability chain rule. In an autoregressive setting this will be,

$$p(\mathbf{x}) = \prod_{d=1}^{D} p(x_d|\mathbf{x}_{<d}) \tag{1}$$

Hence, in the autoencoder output, each dimension can be interpreted as one of the D conditional probability distributions as shown above, and each output unit $\hat{x}_d$ only depends on the previous input units, $\mathbf{x}_{<d}$, and not the other units, $\mathbf{x}_{\geq d} = [x_d, ..., x_D]^T$. This model is trained by minimizing the negative log likelihood for all training data points,

$$Cost = -\log p(\mathbf{x}) = \sum_{d=1}^{D} -\log p(x_d|\mathbf{x}_{<d}). \tag{2}$$

2.2. Group Masked Autoencoder for Distribution Estimation

For the audio anomaly detection problem, we operate in log mel-spectrogram feature space. Instead of using each frame as an input to the network, we concatenate T frames to provide more temporal context to the model. Let's assume that, there are M number of mel bands, hence, the input space is $T \times M$ dimensional. Since for this task, we are interested in the autoregressive ordering across frames (not across each dimension of the input), we design a Group MADE architecture, where the joint distribution can be decomposed as conditionals over groups/frames, instead of individual dimensional conditional. Also, note that in our architecture, the mel bins in one frame are conditionally independent when conditioned on all previous frames.

Let's assume, that one input sample can be represented as $\mathbf{t} = [\mathbf{t_{i+1}}, \mathbf{t_{i+2}}, ..., \mathbf{t_{i+T}}]^T \in \mathbb{R}^{(T\times M)\times 1}$, where i^{th} frame is $\mathbf{t_i} \in \mathbb{R}^{M\times 1}$. Hence the joint density will be decomposed as,

$$p(\mathbf{t}) = \prod_{i=1}^{T} p(\mathbf{t_i}|\mathbf{t}_{<i}) = \prod_{i=1}^{T}\prod_{j=1}^{M} p(t_{ij}|\mathbf{t}_{<i}) \tag{3}$$

Hence, all the mel bins in an output frame $\mathbf{t}_i$ depends on all the mel bins from previous frames but not on other units, i.e., not on mel bins of the i th frame, or on the mel bins of the future frames. Because of this group masking nature, we name our approach as Group Masked Autoencoder for Density Estimation (Group MADE). To compare against the baseline model provided by DCASE 2020 challenge task 2 [17], we set $T = 5$ frames, and $M = 128$ mel bands.

So far, we have assumed that the conditionals modeled by Group MADE were consistent with the causal frame ordering, but in this work we use, 3 different orderings of the input dimensions, and use the ensemble of these three models to compute the anomaly score.

- **Ordering 1 (IDNN):** In this case we predict the middle frame conditioned on 4 other frames, i.e.,

$$p(\mathbf{t}) = p(\mathbf{t}_3|\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_4, \mathbf{t}_5)p(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_4, \mathbf{t}_5) \tag{4}$$

  Note that this ordering represents a probabilistic counterpart of recently proposed start-of-the-art IDNN [6] approach, which predicts the middle frame, conditioned on rest of the 4 frames.

- **Ordering 2 (LR):** In this case we use causal forward AR ordering i.e.,

$$p(\mathbf{t}) = \prod_{i=1}^{5} p(\mathbf{t_i}|\mathbf{t}_{<i}) \tag{5}$$

- **Ordering 3 (RL):** In this case we use backward AR ordering i.e.,

$$p(\mathbf{t}) = \prod_{i=1}^{5} p(\mathbf{t_i}|\mathbf{t}_{>i}) \tag{6}$$

In [18], authors have only considered binary observations, and in this work we extend that to real valued observations. In our first approach we parametrize each conditional distribution as a single Gaussian, and we name this approach as: Group MADE-G, where the autoencoder outputs mean, variance for each Gaussian conditional. In our second approach, we model each conditional as a mixture of C Gaussians, i.e., the autoencoder outputs mean, variance and the mixture component probabilities, and we name this approach as: Group MADE-MOG. Group MADE-G can also be

Table 1: DCASE 2020 Task 2 Experimental Results over Dev Data

Algorithm	Toy Car	Toy Conveyor	Fan	Pump	Slider	Valve
Baseline	78.77 (67.58)	72.53 (60.43)	65.83 (52.45)	72.89 (59.99)	84.76 (66.53)	66.28 (50.98)
IDNN [6]	76.95 (**70.01**)	**76.46** (**62.07**)	69.62 (**53.55**)	74.83 (62.70)	90.16 (72.54)	92.09 (75.13)
Group MADE-G (IDNN)	74.77 (67.35)	74.74 (60.34)	68.02 (53.05)	72.09 (61.66)	92.84 (77.97)	94.78 (84.23)
Group MADE-G (LR)	78.33 (65.34)	69.19 (55.01)	67.17 (52.39)	72.42 (62.26)	93.57 (81.42)	70.87 (62.61)
Group MADE-G (RL)	79.51 (68.41)	72.77 (56.66)	67.53 (52.34)	74.12 (**66.23**)	94.41 (83.65)	95.54 (84.65)
Group MADE-G (mean ensemble)	79.50 (68.40)	74.70 (60.30)	68.00 (53.10)	74.10 (66.20)	**94.40** (**83.70**)	**95.60** (**85.50**)
Group MADE-G (max ensemble)	79.51 (68.41)	74.74 (60.34)	68.02 (53.05)	74.12 (66.23)	94.40 (83.65)	95.60 (84.99)
Group MADE-MOG (IDNN)	79.07 (69.05)	75.14 (61.25)	70.26 (53.16)	74.23 (62.02)	92.71 (77.15)	92.82 (76.33)
Group MADE-MOG (LR)	78.19 (65.57)	70.37 (55.74)	68.44 (52.63)	74.51 (64.66)	93.59 (81.65)	82.38 (67.72)
Group MADE-MOG (RL)	79.74 (68.44)	73.99 (57.68)	68.28 (52.62)	74.96 (65.92)	94.00 (83.01)	93.36 (76.15)
Group MADE-MOG (mean ensemble)	**80.20** (69.70)	75.10 (61.30)	**70.30** (53.20)	74.96 (65.92)	94.00 (83.00)	93.36 (76.15)
Group MADE-MOG (max ensemble)	79.74 (68.44)	75.14 (61.25)	70.26 (53.16)	**75.00** (66.10)	94.00 (83.01)	93.40 (76.40)

seen as a special case of Group MADE-MOG with C=1. For example in case of Group MADE-MOG, for D dimensional input, number of outputs for the model will be, $D \times C \times 3$. For all our experiments we set $C = 10$ for Group MADE-MOG.

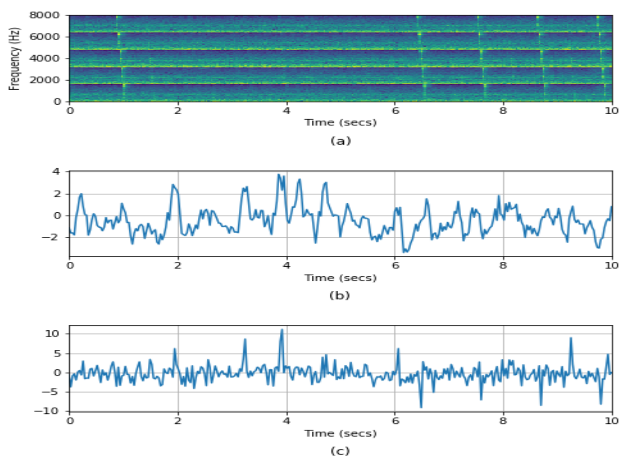

Figure 1: (a) Log mel spectrogram, (b) Anomaly score for baseline model, (c) Anomaly score for Group MADE-MOG (IDNN) for an audio clip from normal valve recording.

2.3. Construction of Masks

Above discussed orderings can be implemented by designing corresponding masks. To explain more, let's consider the causal ordering case, i.e., **Ordering 2**. This can be implemented by making sure that there is no computational path between output frame $\hat{\mathbf{t}}_\mathbf{i}$ and any of the input frames that come after it, $\hat{\mathbf{t}}_{\mathbf{i+1}}, ..., \hat{\mathbf{t}}_{\mathbf{i+5}}$. This will guarantee the conditional represented by the output frame $\hat{\mathbf{t}}_\mathbf{i}$ only depends on the preceding frames $\mathbf{t}_{<i}$. As proposed in [18], convenient way of zeroing connections is to elementwise multiply weight matrices by a binary mask matrix. By setting the corresponding binary mask matrix elements to zero, we can remove our desired connections, and achieve the previously discussed orderings. We

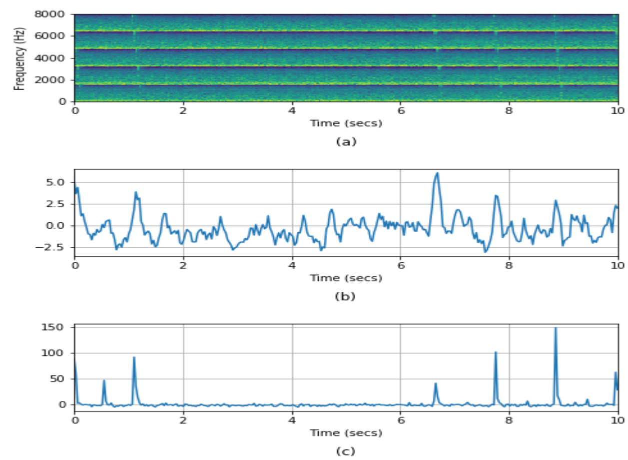

Figure 2: (a) Log mel spectrogram, (b) Anomaly score for baseline model, (c) Anomaly score for Group MADE-MOG (IDNN) for an audio clip from anomalous valve recording.

follow the similar mask construction procedure as described in [18], with the difference being, masks are formed to zero out connections between groups instead of individual units.

2.4. Anomaly Scoring and Ensembling

During inference, we use the negative log likelihood as anomaly score for each input instance, i.e., 5 frames. Finally we average the anomaly scores for each file (10 secs segment), to provide a scalar anomaly score for each recording file.

To ensemble across multiple Group MADE orderings, we transform the anomaly scores of each model into a standardized scale, before combining them. The standardization transformation for any given model is applied in a per-machine ID fashion, by computing the mean and variance of its anomaly scores over the training data for that machine ID. The anomaly scores are then transformed to have zero mean and unit variance over the training data of that machine ID. Standardized anomaly scores across different models are then combined using mean or max ensembling.

3. DATASET

We use the publicly available development dataset released as part of DCASE2020 Task 2: *Unsupervised Detection of Anomalous Sounds for Machine Condition Monitoring*. The data used for this task comprises parts of ToyADMOS [19] and the MIMII [20] Dataset consisting of the normal/anomalous operating sounds of six types of toy/real machines. Each recording is a single-channel (approximately) 10-sec length audio that includes both a target machine's operating sound and environmental noise. The following six types of toy/real machines are used in this task:

- Toy-car (ToyADMOS)
- Toy-conveyor (ToyADMOS)
- Valve (MIMII Dataset)
- Pump (MIMII Dataset)
- Fan (MIMII Dataset)
- Slide rail (MIMII Dataset)

The sampling rate of all signals has been converted to 16 kHz. Challenge organizers also add environmental noise for different SNRs to the target machine sound, and only noisy recordings are provided to increase the difficulty of the problem.

4. EXPERIMENTAL RESULTS

4.1. Setup

The proposed Group MADE model is trained using the negative log likelihood as cost function, using all the normal training data across all IDs for a specific machine. During inference, we use the negative log likelihood as anomaly score for each test sample. We use a fully connected network as the architecture where the number of hidden layers and the corresponding hidden units in each layer follow this structure: $[128, 128, 128, 128, 32, 128, 128, 128, 128]$. Finally the output layer has $640 \times 2 = 1280$ units for Group-MADE-G and $640 \times 10 \times 3 = 19200$ units for Group-MADE-MOG. We use Adam optimizer with 0.001 learning rate for training.

We compare our proposed method with the baseline autoencoder model provided by the challenge organizers. We also implement recently proposed IDNN [6], model using the same architecture as Group MADE models, i.e., hidden units in each layer follows this structure: $[128, 128, 128, 128, 32, 128, 128, 128, 128]$.

Each 10s input file from training data is split into frames of length 64ms, with hop length of 32ms between frames. 1024-FFT and 128 mel bins are used to featurize each frame. We use the log mel-band energies as our input feature space and 5 frames are concatenated, resulting in $5 \times 128 = 640$ dimensional input.

4.2. Results

In Table 1, we report the evaluation results of all competing algorithms along with proposed Group MADE models, over the development set of DCASE 2020 challenge Task 2. All the models are evaluated with the area under the receiver operating characteristic (ROC) curve (AUC) and the partial-AUC (pAUC). The pAUC is an AUC calculated from a portion of the ROC curve over the pre-specified range of interest. In our metric, the pAUC is calculated as the AUC over a low false-positive-rate (FPR) range $[0, p]$, and following the challenge we also set $p = 0.1$. AUC and pAUC have been reported for all 6 machines averaged across IDs.

We observe that autoregressive density estimator based approaches show most improvement over the baseline deep autoencoder approach for non-stationary sounds i.e. for valve and slide rail. Similar trend has been noted by the authors in [6] for IDNN. We also note that, Group-MADE-MOG (IDNN) performs better than baseline model for all 6 machines, and it also outperforms IDNN for 4 out of 6 machines in our experiments.

Fig. 1 shows the log mel-spectrogram (a), anomaly score produced by baseline autoencoder (b), and anomaly score produced by Group MADE-MOG (IDNN) model for an audio clip recording measured from a normal operating valve. Fig. 2 shows the log mel-spectrogram (a), anomaly score produced by baseline autoencoder (b), and anomaly score produced by Group MADE-MOG (IDNN) model for an audio clip recording measured from a faulty valve. It is evident from these two figures, that proposed model captures the anomalous sound better than baseline by producing a high anomaly score for faulty valve recording.

5. CONCLUSION

We have presented an unsupervised approach for audio anomaly detection, using a novel Group Masked autoencoder based density estimation approach. Previously proposed autoencoder based approaches solve the problem by modeling the normal audio recordings, and they detect anomalous sounds only when the reconstruction/model mismatch error is above a certain threshold. In this work, we modify the autoencoder network to determine the joint distribution where the outputs are the conditional probabilities over T frames. We extend the previously proposed MADE to model dependencies across time/frames by introducing the concept of group masking. We also show, how the recently proposed state-of-the-art IDNN is a special case of Group-MADE and how its performance can be further improved by modeling each conditional as a mixture of Gaussians.

6. REFERENCES

[1] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: a review," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192–204, 2005.

[2] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*. IEEE, 2007, pp. 21–26.

[3] R. Giri, A. Krishnaswamy, and K. Helwani, "Robust non-negative block sparse coding for acoustic novelty detection," in *Workshop on Detection and Classification of Acoustic Scenes and Events*, 2019.

[4] E. Marchi, F. Vesperini, S. Squartini, and B. Schuller, "Deep recurrent neural network-based autoencoders for acoustic novelty detection," *Computational intelligence and neuroscience*, vol. 2017, 2017.

[5] B. Bayram, T. B. Duman, and G. Ince, "Real time detection of acoustic anomalies in industrial processes using sequential autoencoders," *Expert Systems*, p. e12564, 2020.

[6] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Anomalous sound detection based on interpolation deep neural network," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 271–275.

[7] B. Abraham and G. E. Box, "Bayesian analysis of some outlier problems in time series," *Biometrika*, vol. 66, no. 2, pp. 229–236, 1979.

[8] D. Agarwal, "Detecting anomalies in cross-classified streams: a bayesian approach," *Knowl. Inf. Syst.*, vol. 11, no. 1, pp. 29–44, 2007. [Online]. Available: https://doi.org/10.1007/s10115-006-0036-4

[9] M. Yamaguchi, Y. Koizumi, and N. Harada, "Adaflow: Domain-adaptive density estimator with application to anomaly detection and unpaired cross-domain translation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 3647–3651.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2014. [Online]. Available: http://arxiv.org/abs/1312.6114

[11] Y. Kawachi, Y. Koizumi, and N. Harada, "Complementary set variational autoencoder for supervised anomaly detection," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*. IEEE, 2018, pp. 2366–2370. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8462181

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[13] B. Uria, M.-A. Côté, K. Gregor, I. Murray, and H. Larochelle, "Neural autoregressive distribution estimation," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 7184–7220, 2016.

[14] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," *arXiv preprint arXiv:1505.05770*, 2015.

[15] B. Uria, I. Murray, S. Renals, C. Valentini-Botinhao, and J. Bridle, "Modelling acoustic feature dependencies with artificial neural networks: Trajectory-rnade," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 4465–4469.

[16] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[17] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," in *arXiv e-prints: 2006.05822*, June 2020, pp. 1–4. [Online]. Available: https://arxiv.org/abs/2006.05822

[18] 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, F. R. Bach and D. M. Blei, Eds., vol. 37. JMLR.org, 2015, pp. 881–889. [Online]. Available: http://proceedings.mlr.press/v37/germain15.html

[19] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "Toyadmos: A dataset of miniature-machine operating sounds for anomalous sound detection," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA 2019, New Paltz, NY, USA, October 20-23, 2019*. IEEE, 2019, pp. 313–317. [Online]. Available: https://doi.org/10.1109/WASPAA.2019.8937164

[20] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection," *CoRR*, vol. abs/1909.09347, 2019. [Online]. Available: http://arxiv.org/abs/1909.09347