Anomaly Detection In Raw Audio Using Deep Autoregressive Networks

Ellen Rushe and Brian Mac Namee

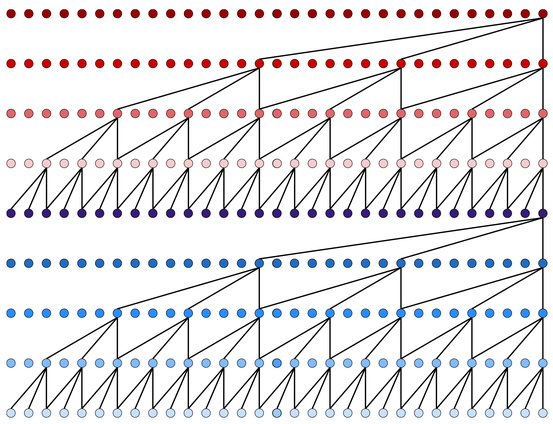
Insight Centre for Data Analytics, University College Dublin

Introduction

For time series modelling autoregressive deep learning architectures such as WaveNet have proven to be powerful generative models, specifically in the field of speech synthesis. In this work, we propose to extend the use of this type of architecture to the area of anomaly detection in raw audio. We compare the performance of this approach to a baseline of a more conventional autoencoder model in experiments using multiple datasets and show superior performance in almost all cases.

WaveNet for Anomaly Detection

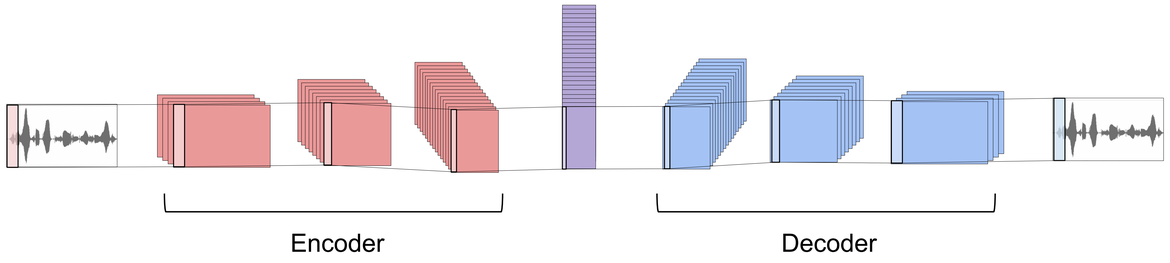
Neural autoregressive models obtain a probability distribution over all outputs by conditioning each on all preceding outputs, assuming a sequential structure in the data [1].



Following the product rule, the output distribution can be factorized into the product of conditionals such that

$$p(x) = \prod_{t=1}^{T} p(x_t | x_1, \ldots, x_{t-1}) \qquad (1)$$

where x_t denotes an output at time t [1]



Baseline: Convolutional Autoencoder

The CAE is composed of 20 layers (10 layer encoder, 10 layer decoder), matching the number of layers in the WaveNet model, with a kernel size of 3 throughout. To recognise anomalies, signal sequences are presented to the trained network and the distance between the predictions generated by the network and the subsequent actual signals is calculated. A small distance is indicative of a normal signal while a large distance is indicative of an anomaly.

Experimental Setup

We use the Task 2 dataset from the 2017 DCASE Challenge [2]. There are three different classes of rare events: ***babycry***, ***glassbreak*** and ***gunshot*** These rare events were artificially mixed with background audio from 15 different environmental settings. We split the dataset into 15 subsets based on setting:

1. beach
2. bus
3. cafe/restaurant
4. car
5. city center
6. forest path
7. grocery store
8. home
9. library
10. metro station
11. office
12. park
13. residential area
14. train
15. tram

Results

We find that WaveNet consistently outperforms the baseline CAE in almost all datasets, with a tie in the *home* and *office* scenarios.

Table: AUC scores for both models on each dataset.

Scene	CAE	WaveNet
beach	0.69	**0.72**
bus	0.79	**0.83**
cafe/restaurant	0.69	**0.76**
car	0.79	**0.82**
city center	0.75	**0.82**
forest path	0.65	**0.72**
grocery store	0.71	**0.77**
home	**0.69**	**0.69**
library	0.59	**0.67**
metro station	0.74	**0.79**
office	**0.78**	**0.78**
park	0.70	**0.80**
residential area	0.73	**0.78**
train	0.82	**0.84**
tram	0.80	**0.87**

It is also noteworthy that performance of both models noticeably varies across the different acoustic scenarios, indicating that the ability of the models to detect anomalies can significantly be affected by different acoustic environments.

Conclusions

In this work we have adapted the WaveNet architecture to the domain of acoustic anomaly detection. We find that we obtain significant performance gains over standard convolutional autoencoders across multiple datasets.

References

[1] A. van den Oord *et al.*, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[2] A. Mesaros *et al.*, "Dcase 2017 challenge setup: Tasks, datasets and baseline system," in *DCASE 2017 - Workshop on Detection and Classification of Acoustic Scenes and Events*, 2017.

Contact

ellen.rushe@insight-centre.org

https://www.github.com/EllenRushe/AudioAnomalyDetectionWaveNet

Acknowledgements

This work has been supported by a research grant by Science Foundation Ireland under grant number SFI/15/CDA/3520.

Insight

UCD DUBLIN

Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn
Science Foundation Ireland
For what's next