Deep note

Karim Kohel, Mohamed Osama

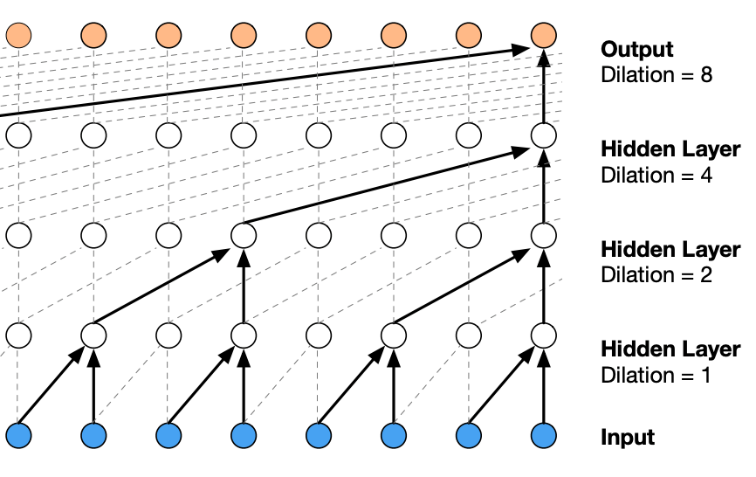
# Abstract

Allowing machines to be indulged in creative fields shows the pinnacle of human-computer interaction. For machines to enter the field and be recognized as a valid competition is another thing. Comes WaveNet, “a deep generative model of raw audio waveforms” (van den Oord and Dieleman, 2016). Letting in the possibility for us to train machines to produce music just as musicians practice their music, but as much as real musicians, the process required time and effort to produce a tolerable output. The model is more than capable of producing real world usable musical scores if trained properly and made easier to try, and this is exactly what we did while testing the model on different music genres.

# WaveNet

###### arXiv:1609.03499v2 [cs.SD] 19 Sep 2016

WaveNet is the leading model in audio generation when it comes to raw audio samples. Modeling raw audio has been avoided by researchers as it can scale exponentially, and what better way to tackle that problem than to make an exponentially growing receptive field for the convolution layer using various dilation factors as The model is fully probabilistic and autoregressive (van den Oord and Dieleman, 2016).

Using real world samples for training, the model generates and validates the generated output as it goes. For sample generation, each sample is is drawn from the probability distribution, then the generated sample is then fed back as input and a new prediction for the next step is made based on the input, thus making output audio sequential in nature as each sample is loosely based on the last one(van den Oord and Dieleman, 2016).

# Implementation

A year later (2017) after WaveNet research paper was published Igor Babuschkin started the open source implementation of the network in tensorflow 1 as tensorflow-wavenet on github under the MIT License.

With the help of 31 other open source contributors, the WaveNet tensorflow implementation was packaged in a neat bundle of tools to allow for easy training and audio generation, or so it seemed as it hasn’t been maintained for the last 5 years.

# How we used the implementation

After recognizing how outdated the Igor implementation was, we were faced with 2 options, either we use the most famous implementation (which was Igor’s) and maintain it ourselves by updating code, tools and packages needed to run the model, Or to go fork hunting for the best up-to-date fork of Igor’s implementation. We did both, after searching for the best fork to use; we recognized that most people working in the field of audio generation are laser-focused on text-to-speech applications for the model, which made our job much harder and much more interesting. After failing in finding the best implementation suited for music generation we had to use the original one and modify it as much as we need by removing old deprecated tensorflow 1 library calls within the code, matching the drivers and their compatibility with each other and hunting down the old tools used in the implementation for music handling in python to just let it work as intended.

# Bibliography

1. van den Oord, A. and Dieleman, S., 2016. WaveNet: A Generative Model for Raw Audio. [online] Deepmind. Available at: <https://deepmind.com/blog/article/wavenet-generative-model-raw-audio> [Accessed 4 January 2022].