MUSIC GENERATION USING WAVENET ARCHITECTURE

R Subba Rao[1], P. Sharada[2]
[1](Department of CSE,Gandhi Engineering College, Bhubaneswar)
[2](Department of CSE,Gandhi Engineering College, Bhubaneswar)

ABSTRACT

Music has been one of the most important things in the human life and has been there to make lives easier as soon as humanity came into existence. People started choosing this as a profession which is why they are known and valued. Now, a good taste of music or voice or having a deep knowledge of musical instruments is considered a skill. However, not everyone can have it all, what if someone has a good sense of sound mixing but not the voice needed to enter the world or the art of playing the instrument. This paper is about making music using wavenet design that allows one to express and manage different notes and tones, that is, to combine as many buttons and notes as they want and to create something that allows you to have full art. Wavenet Architecture is one of the most popular methods of deep learning. As an attempt to look into the future, this research paper discusses automatic music generation as a separate genre of music. As more and more people spend their time at home, creating, listening to, and using music in various projects becomes a major part of many lives. The initial success in music production, production and editing by Artificial Intelligence software is amazing and will accelerate this trend further. Their latest discovery comes from producing their music by putting the status quo in a Music Instrument Digital Interface (MIDI) file, which people can think of is like a music sheet that gives tough instructions to artists.

Key words: Deep Learning, Artificial Intelligence, wavenet, music generation

1. INTRODUCTION

This paper focuses on automated music production using a deep learning-based format called wavenet architecture. Advanced Learning is a subset of machine learning that deals with algorithms that promote brain structure and function called artificial neural networks. One doesn’t have to be naturally well equipped with musical instruments or music to be a musician. Anyone with little experience in deep learning and a passion for music can download quality

music using wavenet. WaveNet is an advanced production model for raw audio waveforms. With the creation of a wavenet, convolutional neural network(CNN) takes the green signal as input and integrates one sample at a time. The job here is to capture existing music data and train the model using this existing data. The model should learn the patterns in the provided set of music files. As soon as it finds out, it will start generating new music. The expectation from the model here is to produce quality music.

CNN is a Deep Learning algorithm that can capture imagery, provide value (readable and discriminatory metrics) in various aspects, objects in the image and is able to distinguish one from the other. The initial processing required for CNN is very small compared to other phase algorithms. This means that the network is learning to optimize filters or characters through automatic reading, and in traditional algorithms these filters are hand-made. This independence of previous knowledge and human intervention in the removal of the feature is of great benefit.

WaveNet is a deep neural network for producing raw audio formats. The model is completely feasible and has the power to stand on its own, with the distribution of speculation of each sound sample placed on all previous ones; however it shows that it can be well trained in data with tens of thousands of samples per second of audio. When used text-to-speech, expressing artistic performance, human audiences rate it as a natural sound that is more important than the excellent parametric and concatenative systems of English and Mandarin. One WaveNet can handle the features of many different speakers with equal reliability, and can switch between them by adjusting the identity of the speaker. When they are trained in modelling music, it is found that it produces musical pieces of novels that are often realistic. It also shows that it can be used as a discriminatory model, restoring promising results of phonetic recognition.

2. LITERATURE SURVEY

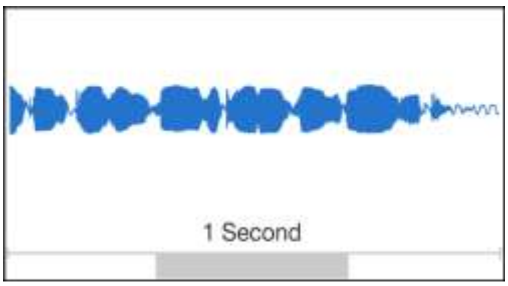
As this research paper emphasizes towards a peek into the future, the topic of discussion over here is automatic music generation. It is like a gateway for those who are naturally not so well equipped with music, although have a very good sense and control over different types of sounds. The focus over here is more towards building a successful basic model, which when implemented upon with different APIs, would give an organized and successful platform.

There have been many previously published papers that helps to find the direction of the project. A paper written by K. Chen, W. Zhang, S. Dubnov, G. Xia and W. Li,[1] and other authors introduces WaveNet as the ultimate training platform for modeling green waveforms, which can be used in text-to-speech in speech the focus is on non-parameter, model-based generation.

In a paper by Michael Furner[2] mention that, for each size, it is the continuation of comparative analysis of different models and techniques and there are several different types of typing.

3. METHODOLOGY

The Training Phase – INPUT: WN takes a slice of green sound waves as input. The green sound wave means the representation of a wave in the domain of a timeline. In the series-time domain, the sound wave is represented in the form of amplitude values recorded at different time periods:



ISSN: 2278-4632
Vol-11 Issue-02 2021

Figure 1 Input Waves

OUTPUT: Given the amplitude value sequence, WaveNet attempts to predict consecutive amplitude values. Let's understand this with the help of an example. Imagine a 5 second audio wave with a sample rate of 16,000 (i.e. 16,000 samples per second). Now, there are 80,000 samples recorded at various times for 5 seconds. Let's divide the sound into equal distances, say 1024 (which is hyperparameter). The diagram below shows the sequence of installation and removal of the model:

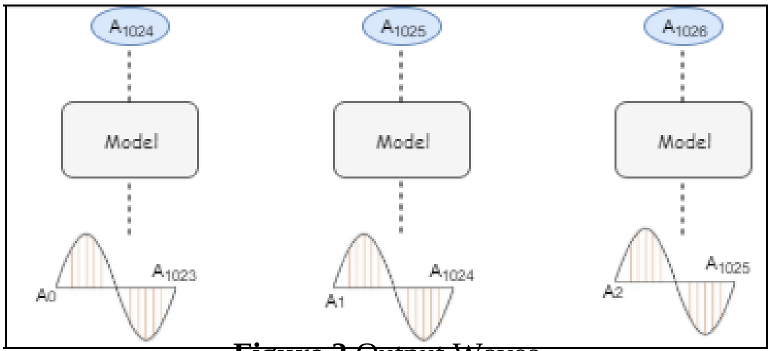


Figure 2 Output Waves

Inference Phase

In the inference phase, new samples are produced in the following procedure:

1. Select a random list of sample values as the start of the model
2. Now, the model releases the distribution of opportunities across all samples
3. Select a value for maximum chances and add it to the list of samples
4. Remove the first item and pass as the next iteration entry
5. Repeat steps 2 and 4 with a certain number of repetitions.

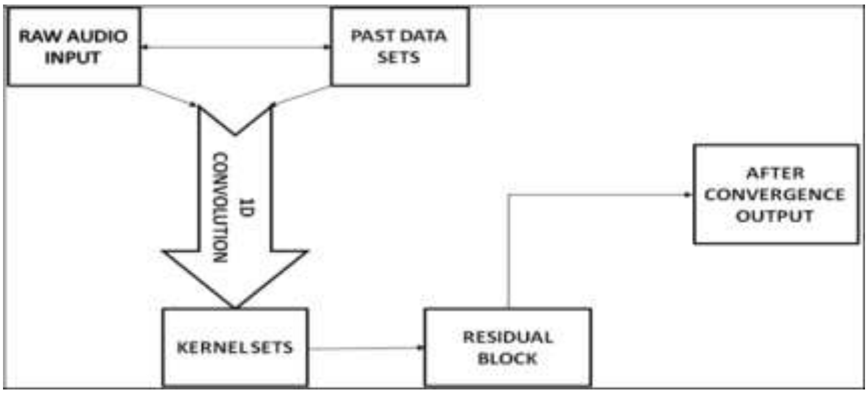


Figure 3 Mathematical Representation

ISSN: 2278-4632
Vol-11 Issue-02 2021

Basis of the wavenet model**,** WaveNet building blocks are the layers of Causal Dilated 1D Convolution. One of the main reasons for using convolution is to remove the input features. Conversion is a mathematical operation that combines two tasks. In the case of image processing, convolution is a straightforward combination of certain parts of an image and a kernel.

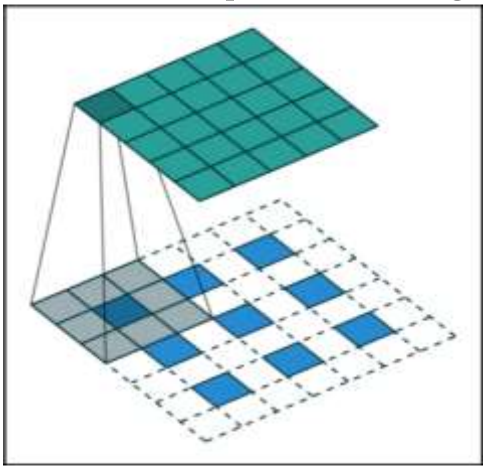


Figure 4 Linear combination of certain parts of an image with the kernel

Convulsive discharge depends on kernel size, installation status, padding type, and stride. Now, after going through various types of attachments, it highlight the importance of using the Dilated Causal 1D Convolution layers.

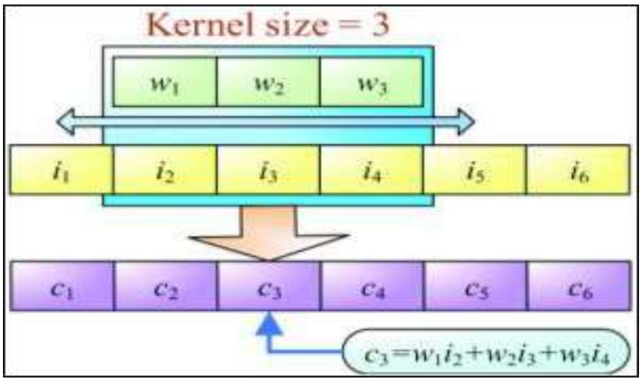


Figure 5 Padding of Kernel

4. DESIGN AND IMPLEMENTATION

THEORETICAL IMPLEMENTATION: When the paste is set to active, the sequence of input and output varies in length. Output length less than input:

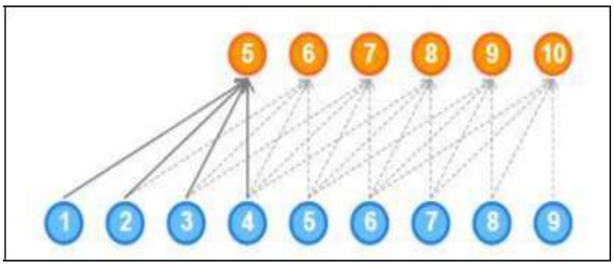


Figure 6

ISSN: 2278-4632
Vol-11 Issue-02 2021

When the padding is set evenly, zeroes are inserted on each side of the insertion sequence to make the insertion and output lengths equal:

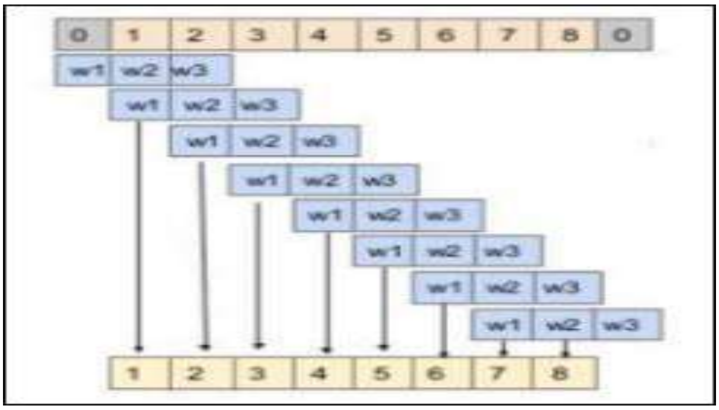


Figure 7

For WaveNet Workflow, Input is placed in causal 1D convolution and the output is then given two different 1-D layers of convolution with sigmoid and *tanh* activations. The clever repetition of the numbers of two different operational values leads to a missed connection and a clever addition of the skipping connection and the release of causal 1-D results to the remaining block.

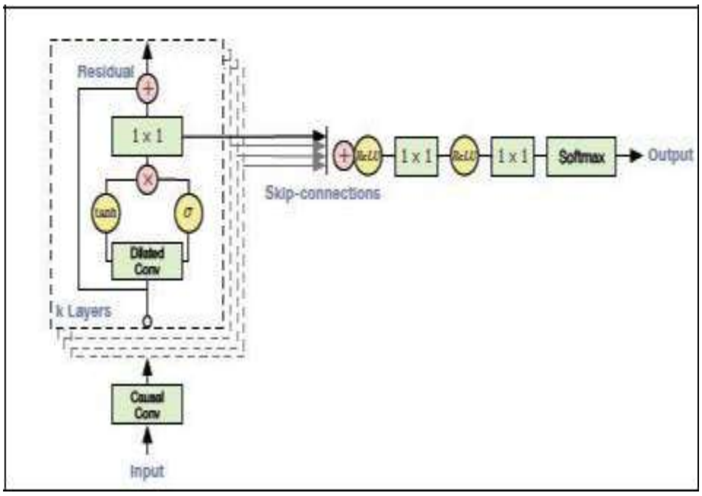


Figure 8 A building block contains Residual and Skip connections which are just added to speed up the convergence of the model

PRACTICAL IMPLEMENTATION: First different available datasets are taken and then start parsing the MIDI files into a language so that the system can understand it, then these files are plotted using the combined concepts of kernel and 1D convolutions and hence, the outputs are displayed which are chosen on the basis of highest frequencies. The plot shows that that most of the notes have a very low frequency. So, only the top frequent notes are kept and the low-frequency ones are ignored. Here, a threshold value has to be determined. Nevertheless, the parameters are variable.

ISSN: 2278-4632
Vol-11 Issue-02 2021

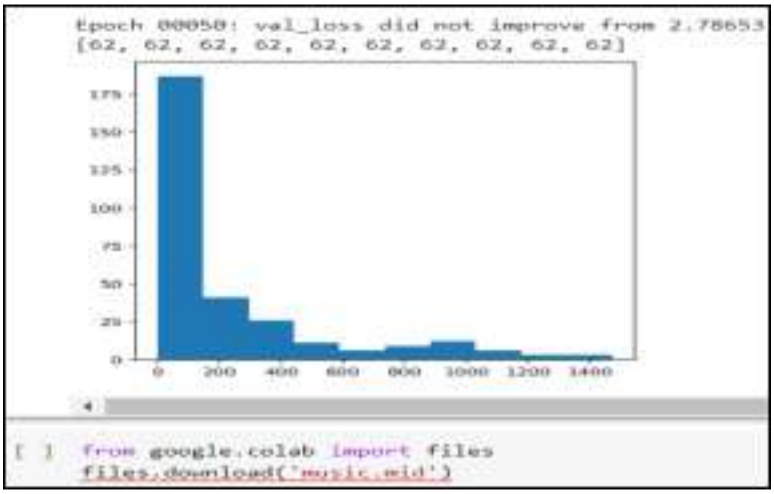


Figure 9 This plot gives top frequent note and removes the low frequency ones

Figure 10 The values after being mapped and at approx 50 epoch generates a combined music which is then downloaded on the system.

4. CONCLUSION

The paper shows the utilization of a simple WaveNet based network to make music production. The results turn out to be very successful showing that the neural network can be used to create music and has the potential to produce complex musical layers. WaveNet has proven to be a good example of catching long-term dependencies. By providing music note material to the network, it has been able to learn the style of music that has been used to produce music. Future work can be done with a variety of WaveNet and integration models that will require more powerful GPUs and the APIs can be implemented along with it to build a platform /application for emerging artists to work on and respect their music preferences.

REFERENCES

[1] K. Chen, W. Zhang, S. Dubnov, G. Xia and W. Li, "The Effect of Explicit Structure Encoding of Deep Neural Networks for Symbolic Music Generation," 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), 2019, pp. 77-84, doi: 10.1109/MMRP.2019.00022.

ISSN: 2278-4632
Vol-11 Issue-02 2021

[2] Michael Furner, Md Zahidul Islam, Chang-Tsun Li. (2021) Knowledge Discovery and Visualisation Framework using Machine Learning for Music Information Retrieval from Broadcast Radio Data. Expert Systems with Applications 89, pages 115236

[3] Daniel Rivero, Iván Ramírez-Morales, Enrique Fernandez-Blanco, Norberto Ezquerra, Alejandro Pazos. (2020) Classical Music Prediction and Composition by Means of Variational Autoencoders. Applied Sciences 10:9, pages 3053.

[4] Wang Shuo, Mu Ming. (2021) Exploring online intelligent teaching method with machine learning and SVM algorithm. Neural Computing and Applications 20.

[5] Joo-Wha Hong, Qiyao Peng, Dmitri Williams. (2020) Are you ready for artificial Mozart and Skrillex? An experiment testing expectancy violation theory and AI music. New Media & Society 19, pages 146144482092579.