AI for Media – Assignment 3 – Crysern Smith

LSTM Audio Generation

Goals

Set out the intention for your project. It can be creative aim or more practical. It may be helpful to describe a creative or industry setting in which your generative model might be useful?

To generate the sounds and music created by African tribes.

The development of African countries means the loss of these traditional sounds and their creation, and we want to try to generate the music using existing sounds and recreate it using LSTMs.

Exploration:

The following outputs were generated with a model that took an approximately 1 hour to run.

Sample Input 1: Zulu Tribal Dance in South Africa - https://www.youtube.com/watch?v=HxhhF_nHxIs&t=63s

Drum beat replicated: 1:06 minutes

Sample output: output1.wav

Sample output1 - The first sample output exceeded expectations; the interpretation of drumbeat and speed turned out well in comparison to the dataset it was given.

Sample Input 2: Zulu Song of the warriors- https://www.youtube.com/watch?v=1mTsu8HvC1g

Vocal replicated: 0:13 seconds

Sample output: output2.wav

Sample output2: The output on the second model did not perform as well as the first output; when capturing the voices, the model struggled with replication.

However, just too much white noise is in the output; the general tone and speed of unison singing are captured to a certain extent.

Reflection

What advantages would this approach bring over traditional (non-generative) techniques?

Some advantages and creative use cases for this audio model is in music composition. Using an LSTM audio model to learn the patterns of a particular song or genre of music, a composer could create new pieces of music that are more faithful to the style of the original music.

How well has this approach met your aims (creative or otherwise)?

As stated before, the reasoning for the chosen dataset is for the preservation and recreation of cultural musicianship unique to tribes in Africa.

The aims of recreation were not met, but rather an opportunity had arisen from this experiment that if trained on a more accurate dataset, the possibility for historical preservation is possible.

What improvements in either control of the process and quality of outputs would you make it better for the task you have described?

The improvements needed could be the sound quality and clarity of the dataset. However, this could be a challenge when working with historical data; hence we could try to record the music with appropriate equipment and derive a qualitative dataset.

If the above is not possible, another solution could be to extract a single voice, a beat, then feed into the model one by one for replication with each input at the same rate, speed, and length.

We then should train the model for a lot longer, and with more epochs, we could then layer each output to create a close replication of the original sound.