EXTENDING MUSIC BASED ON EMOTION AND TONALITY VIA GENERATIVE ADVERSARIAL NETWORK

Romit Mohanty(190720 romit@iitk.ac.in), Devesh Kumar(190277 rdevesh@iitk.ac.in),
Priyal Pathak(190648 priyal@iitk.ac.in), Sahil (190737 malsahil@iitk.ac.in)
Pranjal Pratap Dubey(190626 dpratap@iitk.ac.in)

I. Baseline Paper general details

Title: B-W Tseng, Y-L Shen and T-S Chi, *Extending Music Based On Emotion And Tonality Via Generative Adversarial Network*
Conf. name: ICASSP
Year: 2021

II. Problem solved

A generative model for symbolic music generation from not only temporal and spectral characteristics, but also from high level attributes like emotion and tonality of previous music pieces.

III. Data Set Details

The data set contains a collection of 11380 pop music tabs with C tonality in the form of MIDI files of different lengths. Each tab contains two channels, one for melody and the other one for the chord. The bit-rate of the MIDI audio files is around 230-250 bits/sec. The sampling rate can be found by dividing the size of each sample by the bit rate.

IV. Programming Framework:

They have not mentioned about any specific programming language, however Python with tensorflow/pytorch frameworks can be used to build the desired networks.

V. Similar SOTA Papers:

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, *"Generative adversarial networks,"* arXiv preprint arXiv:1406.2661, 2014.

Note:- In GANs above, the input of G is a vector of random noise, whereas the output of G is the generated signal. In baseline paper settings, the input of G was the 8 latent variables of the previous 8 bars.

E Moliner and V Valimaki, *BEHM-GAN: Bandwidth Extension of Historical Music using Generative Adversarial Networks.* IEEE 2022

Note:- SEANet is a GAN-based model that was used for speech enhancement and bandwidth extension. They opted to train SEANet using the same training objective as above one's. This gave them better performance than the "feature" loss the authors originally applied.

VI. Proposed Method

Model I: The baseline model consists of generator for melody and chord using a standard Least-Squares GAN.

Model II: Baseline model combined with two classifiers for musical emotion and tonality, respectively. The model also learns from the cross-entropy of the above two classifiers.
Only limited amount of data with emotion labels are available, so directly training a model to generate emotional music with high performance is not feasible.