Using GANs and RNNs To Synthesize Music In The Time Domain

Michael Wisnewski
University of Iowa
Iowa City, IA 52242
mwisnewski@uiowa.edu

March 13, 2019

I am interested in experimenting with novel ways of generating noise in the time domain. I want to see if it is possible to achieve this with GANs and RNNs. In the past, many have sufficiently generated MIDI files using GANs and a few have tackled time domain music synthesis with RNNs. I seek to combine the different approaches and use GANs to train an RNN that can generate music in the time domain. This is worthy of research because nobody has created a network that can generate music comparable to that commercially sold. The best results sound like a poorly recorded live performance.

To achieve this, I plan to train the RNN initially on sub-segments of songs from the same album. This will hopefully keep the style and sonic qualities as similar as possible. The dataset is largely irrelevant, as it will only affect the style of the output and hopefully not the quality of the output. My goal is to generate results more convincing than that of Dadabots in their 2017 SampleRNN model. As stated earlier, they were comparable to poor live performances. This was due to a large amount of static noise present in all samples. I aim to create a model that can learn to "filter" out this noise and produce cleaner results. I will also implement the technique introduce in the 2018 IntroVAE paper of seeding the generative model with the output of an autoencoder. This drastically reduced blurriness of output images in their tests. I believe that this may be critical in managing ambient/static noise in output samples. Due to the largely subjective measure of the quality of output samples, I will mostly measure success by presence of noise and how "musical" the output sounds.