Guitar Tone Converter

By Luke Abram

Objective:

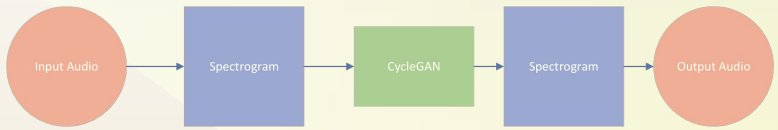
Given a recording of an acoustic guitar playing one note, the model should be able to create output audio which sounds closer to the target instrument than the source

Introduction

- There is currently no good way to convert the audio of one instrument to make it sound like another
- This approach utilizes a short time Fourier transform to convert audio into spectrograms, or 2D frequency vs time arrays.
- Once the audio is converted into a spectrogram, it is fed into a CycleGAN neural network.
- CycleGANs perform image to image translation with unpaired datasets.
- The CycleGAN in this implementation, translates spectrograms of different instruments.

Methods

- The conversion of audio to an image (spectrogram) and back to audio is optimized until there is no noticeable difference between the source and output audio
- Spectrogram creation is done with the librosa STFT package.
- The CycleGAN will be trained and tested on datasets of an acoustic and electric guitar



Before testing the CycleGAN on timbre conversion, the model was tested on chord modification with sine waves as a proof of concept. Figure 1.1 shows spectrograms for the input interval and output chord. The model was able to consistently make accurate transformations on test data.

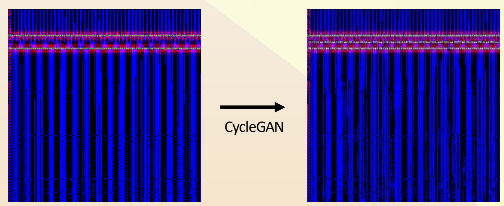


Figure 1.1

The second set of spectrograms in Fig. 2.1 is of an acoustic and electric guitar. The image was originally 1024 pixels long, however due to the limitations of the CycleGAN, it was cropped down to 256x256 pixels. The cropped data was extremely miniscule in comparison to the bulk of the data captured in the lowest 256 frequency bins. For this reason, sufficient translations can be made with ¼ of the data, however, when dealing with sources such as piano, there are much more overtones and harmonics in the upper frequencies.

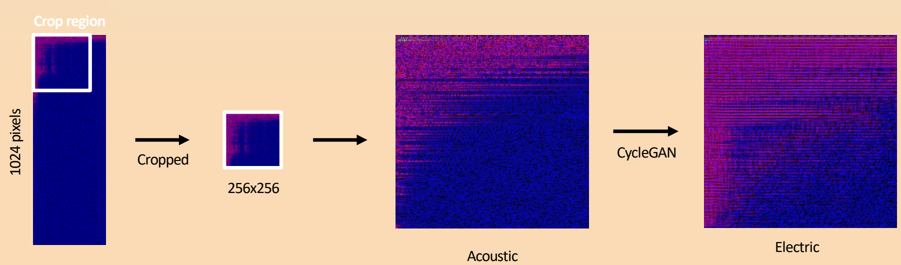


Figure 2.1

Results

- The most promising iteration of the converter performs the bare minimum conversion of an acoustic guitar note to an electric guitar tone.
- The model's timbre conversion problems can be attributed to a lack of consistent training data.

More Methods

- The short time Fourier transform creates a complex number array.
- One approach to this problem was to take the absolute value of the array, converting each complex number to a single positive number.
- Two adjacent pixels were created for each cell from the spectrogram
- When testing the conversion of audio into a spectrogram and back to audio, this new method yielded near-perfect conversion with no audible difference in input and output audio.

References

librosa.stft—Librosa 0.10.0.dev0 documentation. (n.d.). Retrieved September 20, 2022, from https://librosa.org/doc/main/generated/librosa.stft.html

STFT_Notes_ADSP.pdf. (n.d.). Retrieved September 20, 2022, from https://course.ece.cmu.edu/~ece491/lectures/L25/STFT_Notes_ADSP.pdf

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2020). *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks* (arXiv:1703.10593). arXiv. http://arxiv.org/abs/1703.10593