Generating Song Covers in Different Styles

Ben Heller, Abraham Ryzhik, Zarah Tesfai

Motivation

It is relatively straightforward for humans to distinguish between different artists' singing styles, and many people also enjoy creating covers of songs in their own personal style. This project focuses on musical style transfer and generating song covers in the style of a specific artist, and also classifies songs based on artist.

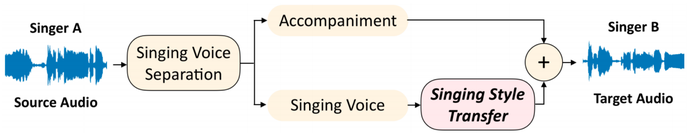
Data

Vocal Stems

- For the neural network, we used vocal stems from *voclr.it* of Beyoncé and The Weeknd
- We used 13 samples from each artist, and these samples were at *34,100 Hz*

Features

To preprocess our data, we constructed spectrograms which are images of time vs frequency graphs. These graphs give us a more useful form of the audio for us to run the various models on.



Models

Cycle-BEGAN

- Wu et. al. [2] uses cycle consistent Boundary Equilibrium Generative Adversarial Networks, with $Cycle\ Gan = L_{GAN} + L_{cyc}$:

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

Support Vector Machines with Principal Component Analysis

- We first use PCA to reduce the dimensions of our data, since it starts with several thousand dimensions. We then used an SVM

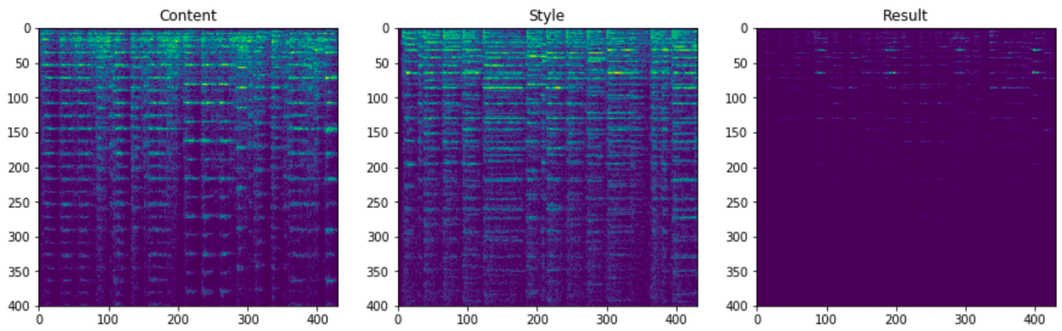
Convolutional Neural Network

- We use a convolutional neural network to classify by genre due to the shift invariance of CNNs.

Results

CNN

- Number of iterations: 300
- Final loss: 568.6033



Discussion

None of the neural networks fully achieved our goal, but we saw progress in our efforts. The first 2D CNN showed some promise when trying to merge instrumental music, but when we tried to apply this method to vocals it was not clear whether or not the network had achieved anything. When we tried the 2D CNN modified to compute Gram matrices over the time axis, the output was still not excellent, but there was a more noticeable effect. The output still sounded much more like the content, but there were several changes we could tell came from the style audio. The CycleBEGAN method worked the best: there was a clear total change to the content audio to match the style of the style audio. However even with this method, the audio quality was very poor in the output.

Future Work

Of course, we want to improve the performance of our singer classifier to more accurately determine the singer of existing songs. However, the main goal in the future is to develop better models to perform singing style transfer, since all existing models are either very specific or largely ineffective. Judging the efficacy of style transfer algorithms can help guide the development of these algorithms in the future.

References

[1] Noam Mor, Lior Wolf, Adam Polyak, Yaniv Taigman (Facebook AI Research). *A Universal Music Translation Network.*
[2] Cheng-Wei Wu, Jen-Yu Liu, Yi-Hsuan Yang, Jyh-Shing R. Jang. *Singing Style Transfer Using Cycle-Consistent Boundary Equilibrium Generative Adversarial Networks.*
[3] Dmitry Ulyanov and Vadim Lebedev. *Audio texture synthesis and style transfer.*
[4] Toya Akira. *Voice style transfer with random CNN.*
[5] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. *A Neural Algorithm of Artistic Style.*
[6] Jeremy F. Alm and James S. Walker. *Time-Frequency Analysis of Musical Instruments.*

Emails: {bheller2, aryzhik, zaraht} @ stanford.edu