Computer Vision Capstone — Isha Mody

Abstract

For my capstone, I was able to learn and explore a variety of topics within computer vision through extra credit assignment and an elaborate final project. Being interested in how social media employs Computer Vision algorithms, my teammates and I did a capstone project called "TikTok Video Edit Generator," that automates the entire pipeline of creating a ready-to-post TikTok music and dance video. I have included our poster for the project below, which includes snapshots of the resulting video and classification accuracy. The outline of the project includes providing a song and a video, either pre-recorded or through the webcam, for the model to classify the genre of the song from 10 pre-selected categories. It then extracts the appropriate features of the song using this classification result and a spectogram. With these features, the model then creates edits to the video in theme with the genre and in time with musical changes of the song. This project streamlines the editing process and outputs a professional-looking TikTok music video. Through my capstone project, I was able to build a CV model that is applicable in modern day social media.

Figure 1. Rock Video Edit

Figure 2. Country Video Edit

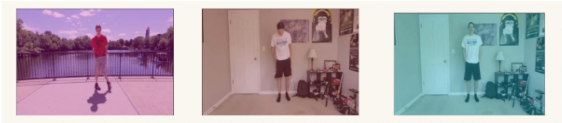


Figure 3. Disco Video Edit

Figure 1: Snapshots of 3/10 Genre Video Edits

TikTok Video Edit Generator

Ethan Asis | Zach Sickles | Sameer Sinha | Isha Mody

BROWN

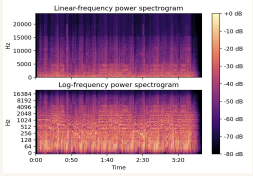
CSCI 1430

Motivation

Massive interest in short video social media

A desire to help new musicians and small creators

The Problem

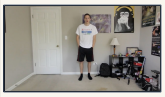


Genre Classification
Utilizing Visual Song Representation

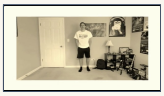
Video Editing
Utilize features of music audio and visualizations to sync genre-based video edits to music

Goals

1. Classify the genre of music based on its spectrogram
2. Using other spectrogram features and the genre, generate video edits that go with the song's genre and features to generate social media (Tiktok) content



Classical
In this case, sepia + video size changes according to music dynamics

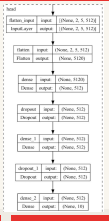


Core Insight

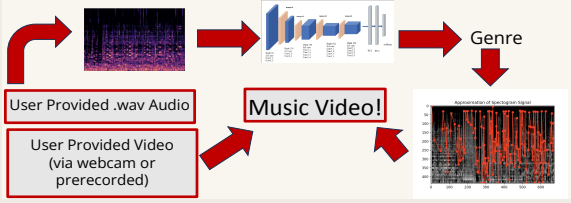
For Classification, rather than training a model from scratch, we fine tuned a VGG with a classification head

Training on an image of a spectrogram generated from a .wav file of music led to 88.89% genre classification accuracy on 2-second clips, and much higher when taking a "majority" vote of such clips to determine Genre

Based on the genre, we checked for appropriate features, which change based on genre and made genre-specific edits to a video.



How it Works



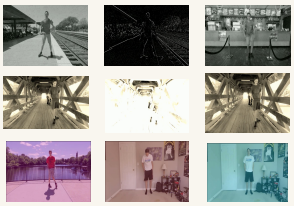
Output Results

Examples (3/10 Genres):

Rock Edits ("In Bloom" by Nirvana): Grayscale Filter + "vintage" looking noise + Sobel Filter on loud percussive frequencies

Country Edits ("Shake It For Me" by Luke Bryan): Sepia + occasional noise bursts on heavy downbeats with the rhythm

Disco Edits ("Dancing Queen" by ABBA): Various color filters align with the beat in classic disco fashion

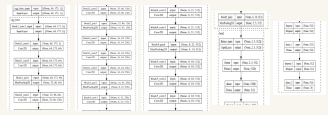


Other results

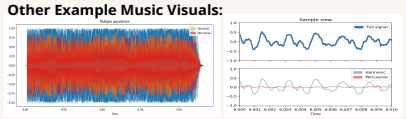
User Interface:

Our finished product allows users to record or upload a video and set it to a song, where appropriate edits will automatically be made based on the detected genre.

Classifier Model:



Other Example Music Visuals:



References

[1] Nirmal; Shahjee Mohan, "Music Genre Classification using Spectrograms" *IEEE*, 2020.
[2] Paul Haeberli, "Matrix Operations for Image Processing" *Grafica Obscura*, Nov. 1993.
[3] Doshi, Sanket. "Music Feature Extraction in Python." *Towards Data Science*, Dec. 2018

Acknowledgements

We would like to thank the creators of the VGG model which we fine tuned, ProjectOneLife for his dancing video, and all artists and producers whose music we used (Nirvana, Bob Marley, Krept and Konan, Luke Bryan, ABBA, Slipknot, and plenty of others).

Figure 2: Poster of Capstone Project