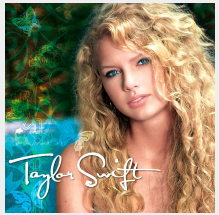
Genre Classification Based on Album Art

Cassidy Newberry & Mateen Saifyan

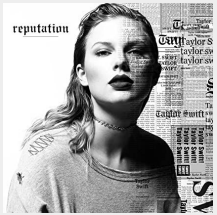
Our Goal

- Music is a major part of our daily lives and culture
  - Many nuances in different music
- Album art sets the tone for songs and merchandise
- Genre classification tends to be done with audio / album name
- Many large artists such as Lady Gaga, Taylor Swift have progression in album art corresponding to a shift in music genre
- **Can we predict the genre of an album based on the album art?**



Country

Synth-Pop



Pop

Pop

Indie/Alt

Data Sourcing

- RGB Images from MuMu: MultiModal Music Dataset (2017)
  - Music gathered from Amazon Reviews Dataset and Million Songs Dataset
    - Album Images from Amazon Reviews
    - Genres Classification from Million Songs
  - Image size 300x300

- Two techniques:
  - From multi-genre songs include all that include top 16 genres
    - Possible Issue: Album art may be included in Pop training AND Country
  - Single genre label only

- 16 Unique Genres Reduced to Top 12 Genres
  - At least 32 images per Genre
  - Drop New Age, Reggae, Classical, Gospel

- 2,267 Total Images

What We Tried

- **Computer Vision**
  - **Baseline Convolutional Neural Network**:
    - 5 convolutional layers
    - Trained for 5 epochs:
      - Training Accuracy: 24%
      - Validation Accuracy: 21%
  - **Pre-trained VGG-16 Model**:
    - Used pre-trained weights
    - Partially frozen
    - Trained for 2 epochs:
      - Training Accuracy: 51%
      - Validation Accuracy: 24%

What Went Wrong / Challenges We Faced

- Downloading many images from the web onto our local computer was really difficult. We had to write a python script to do the web scraping. This was a long, tedious process that took us awhile to figure out.
- We had a very imbalanced dataset. We had too many observations with the Dance & Electronic label, for this to perform better we would need to balance our dataset.
- We did not have the computing power to effectively train the VGG.
    - It started to perform better after two epochs, it likely would have performed much better if we were able to let it train for significantly longer.

What Went Well

- The model did pretty well at predicting the genre for albums that belonged to the most popular genre (Dance & Electronic)
- The VGG performed well on the training data and was an improvement from our baseline convolutional neural network
    - This could be worked on more and likely perform better if we had more computing power

Questions?