

Music Genres Classification Derived from Album Cover Using Convolutional Neural Networks

Tyler Gutowski

November 16, 2023

1 Introduction

Music genres often reflect unique artistic, cultural, and historical elements. The visual design of album covers can be an indicator of these genres, providing a rich dataset for machine learning. This project explores the classification of music genres using visual cues from album covers with the application of Convolutional Neural Networks (CNNs). We hypothesize that CNNs can effectively classify images into music genres, offering insights into the relationship between visual art and music.

2 Data Preparation

2.1 Data Collection

Album cover images were initially collected using the Spotify API. The API provided access to a diverse range of albums and artists. A dataset was constructed containing paths to album images and their corresponding genre labels, leveraging the metadata provided by Spotify. Unfortunately, albums don't have specific genres, but artists do. This meant that for each album requested, an artist request must be sent, which doubled the number of Spotify API calls. Due to the large number of images needed for the training of the network, it was decided that APIs were not the best method. Instead, the MusicOSet dataset was used, which provides 26,522 albums and 11,518

artists which are stored in separate .csv files. The .csv files contained links to images, so no API calls were necessary.

2.2 Data Preprocessing

The preprocessing of the data was crucial for the success of the model. Album cover images were resized to a standard dimension of 128 by 128 pixels, with 3 channels to ensure uniformity. A script was employed to automate the downloading and resizing of these images. It is standard practice to separate classification into unique folders, but because some albums consisted of multiple genres, all of the albums were stored in a single folder. Artists such as Joey Bada\$\$ had multiple genres, such as conscious hip hop, gangster rap, hip hop, pop rap, rap, southern hip hop, trap music, and underground hip hop. Due to the sheer amount of artists in the dataset, a total of 1378 genres existed, which was far too many to train in a reasonable amount of time. Additionally, most of these can be categorized into a single genre. Gangster rap and pop rap can both fit under the rap category. To fix this issue, only the genres that occurred in at least 5% of the songs were kept, which left 24 genres.

3 Methodology

3.1 Model Architecture

The architecture of the CNN was critical to the project's success. The model consisted of multiple layers, each designed to capture different aspects of the image data. The Conv2D layers were responsible for extracting features from the images, while the MaxPooling2D layers reduced the dimensionality of these features. The Flatten layer converted the 2D features into a 1D vector, which was then processed through Dense layers for classification. Dropout layers were included to prevent overfitting.

3.2 Training Process

The model's training involved several steps. Firstly, the preprocessed dataset was loaded into the model. The genre labels were one-hot encoded to facilitate classification. The dataset was then split into training and testing sets.

The model was compiled with a categorical cross-entropy loss function and optimized using the Adam optimizer. The training process was monitored for accuracy and loss metrics to ensure effective learning.

4 Results and Discussion

4.1 Model Evaluation

The model’s performance was evaluated using the testing set. The evaluation focused on accuracy and loss metrics, providing a quantitative measure of the model’s effectiveness. A confusion matrix was generated to visualize the model’s performance across different genres.

4.2 Visualization

Visualization tools were used to illustrate the model’s classification accuracy per genre. The confusion matrix provided insights into the genres that were most accurately classified and those that posed challenges.

5 Conclusion

This research demonstrates the potential of using visual cues from album covers for music genre classification. The CNN model showed promising results, indicating that album cover imagery contains significant information relevant to genre classification. This study opens avenues for further research in the field of music classification, particularly in exploring the relationship between audio and visual data in understanding music genres.