CS543: Music Genre Recognition using Audio Samples

Aditya Vyas - Vedant Choudhary

Group 35

The Problem

- Lots of previous research in genre recognition but mostly using **lyrics**
- The problem is that using only lyrics does not take into account the musical features of a song - frequency, betas, tone etc...
- Using direct audio samples can help in removing these problems and learn these musical features.

The Data

- GTZAN - One of the most widely used dataset for genre recognition tasks
- However, we did not use this data - has only 1000 audio samples, lots of mislabelling, repetitions and distortions and finally it is very old
- FMA (Free Music Archive) - A huge dataset of audio samples having around 917 GB of data. We used a small subset of 8000 audio samples containing 8 balanced genres.

Count of Tracks per Genre
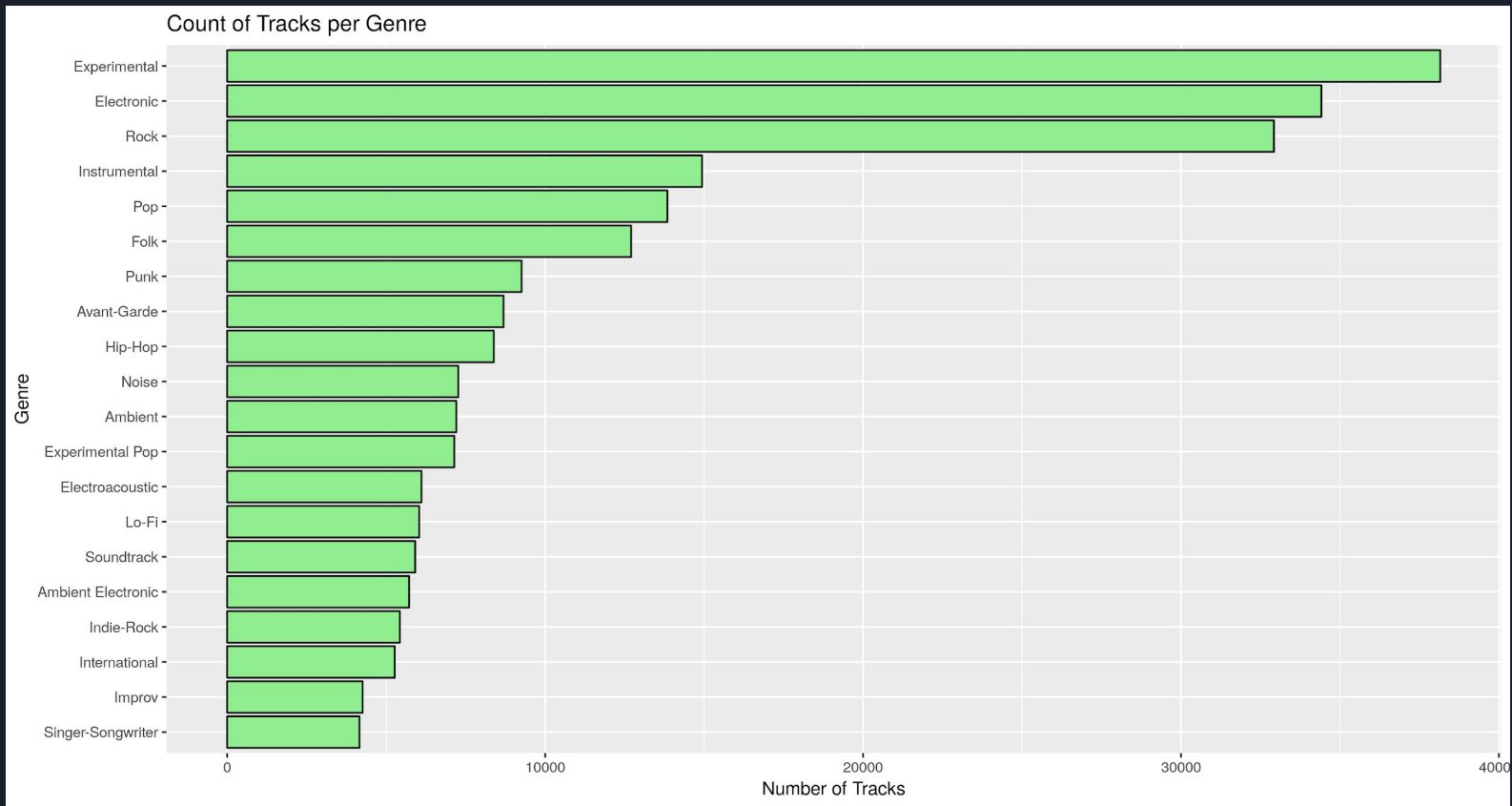
The Architecture

- Input - The input to the network is the audio converted to a spectrogram.
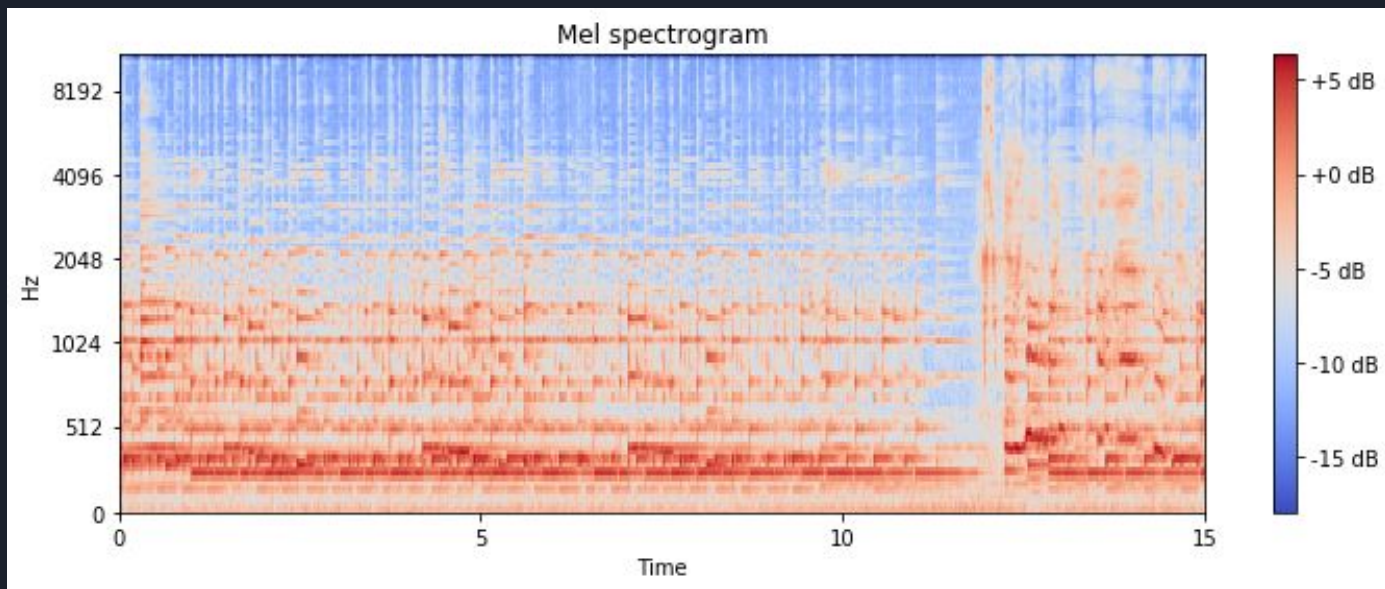
The Architecture

- A spectrogram captures the frequency spectrum of a song with time.
- Points with high peaks are notes played over a short time. Chords are long stretches of horizontal lines as they played over time.
- **3 Convolutional Layers**
- Followed by a **Time-Distributed Layer** - getting probabilities for all 8 genres at each timestep.
- **Softmax Activation**

Conclusion and Future Work

- Our network predicts the genre with **60% accuracy.**
- Improvements can be done to this result by trying better architectures.
- Future work - Once a network learns to recognise the genre of the song, it can be made to produce a sound clip of a particular genre - **music generation.**