Deep Convolutional Networks for Music Genre Classification

Kai Middlebrook, Shyam Sudhakaran, Kunal Sonar, David Guy Brizan

UNIVERSITY OF SAN FRANCISCO
Machine Learning | MAGICS LAB

Spotify

Summary

- We explore **end-to-end Convolutional Neural Network (CNN)** and **Long Short-Term Memory (LSTM) Hybrid architectures** for **raw audio genre classification tasks**.
- We **adopt deep architectures** from state-of-the-art image classification and speech recognition networks (residual layers, mixup, dropout, cascading convolution filters).
- We utilize the Spotify API to **introduce an un-curated dataset** that is more balanced than popular genre classification datasets.

Data

- We use our un-curated Spotify dataset to train our models and test them on two popular genre classification datasets (Tagtraum, GTZAN).

Figure 1

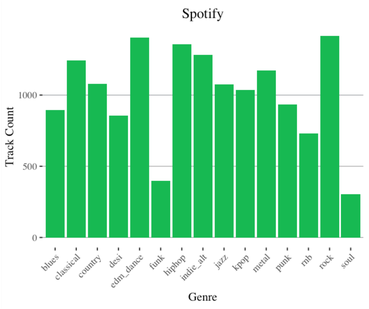
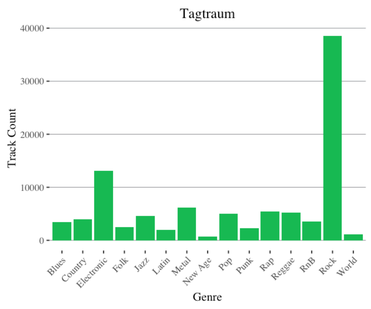


Figure 2



- **Spotify Dataset** (Figure 1)
  - 15,177 songs
  - 15 genres represented
  - 30 seconds of audio for each song
- **Tagtraum Dataset** (Figure 2)
  - 97,516 songs
  - 15 genres represented
  - 30 seconds of audio for each song

Figure 3

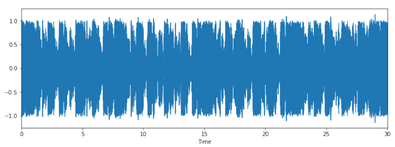
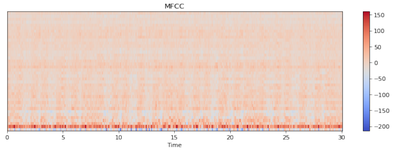
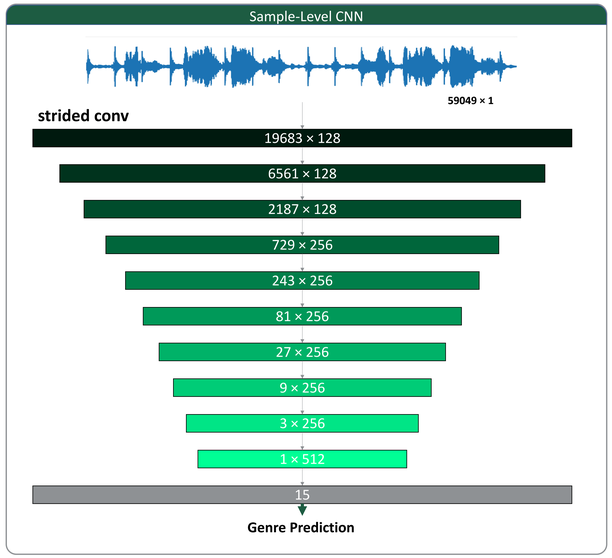


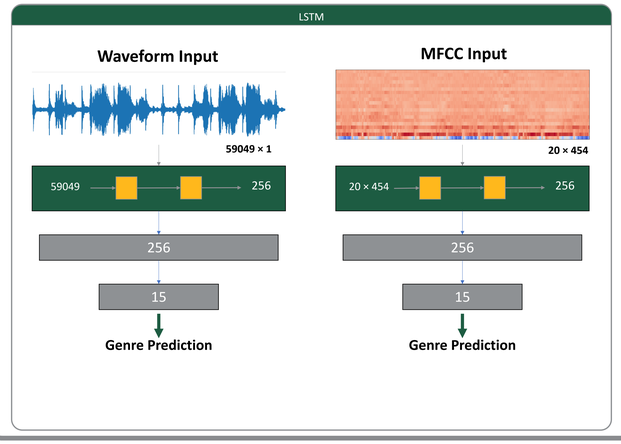
Figure 4

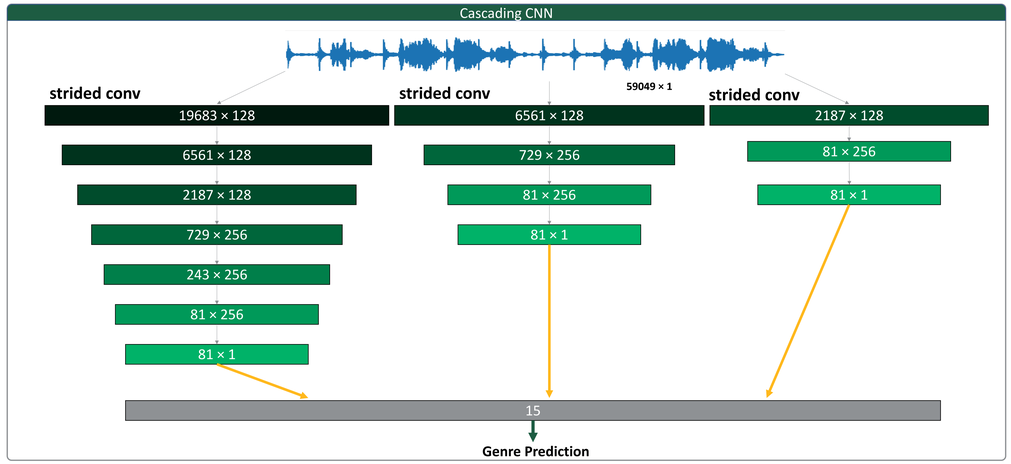


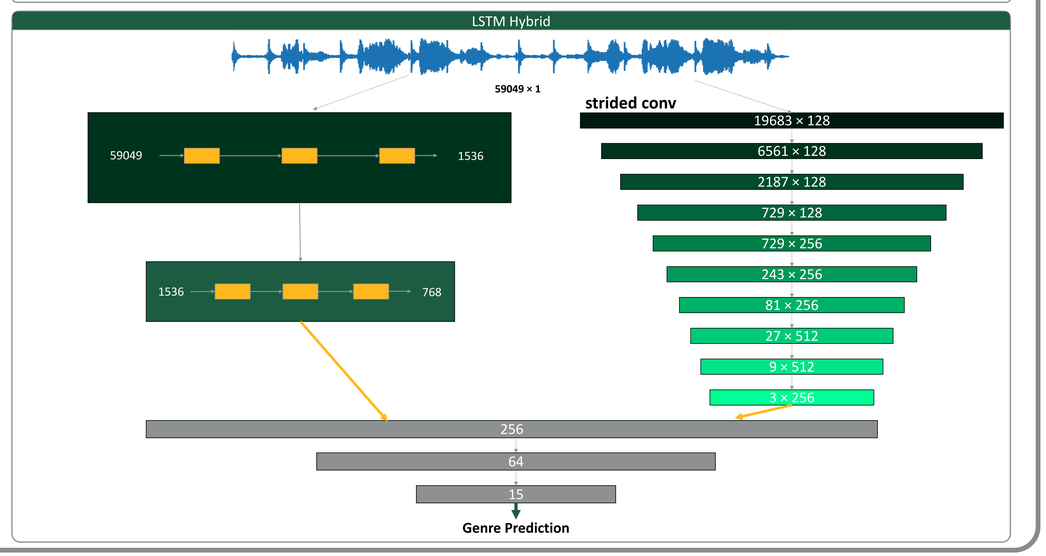
- Raw waveform audio (Figure 3) or preprocessed audio in the form of spectrogram (Figure 4) were used as the input for our models

Models









Results

Model	Train Accuracy	Test Accuracy
Sample-Level CNN	99.6%	51.54%
LSTM Hybrid	42.3%	41.80%
Cascading CNN with mixup	48.50%	38.60%
LSTM (MFCC input)	29.2%	26.2%
LSTM (waveform input)	16.3%	16.3%

Conclusion

Our models are promising — we achieve close to state-of-the-art test accuracy on our Cascading CNN and LSTMHybrid architectures. However, our models suffer from some **mild under- and overfitting**.

Our **future work** will focus on the following:

- Experiment with **downsampling**-- testing variable sample rate/duration waveforms
- Test **various model hyperparameters**--imitating more time series models

We believe that **with further experimentation**, these architectures **can achieve state-of-the-art results** for music genre classification.