Classifying classical music by era

Sara Soueidan

June 9th, 2020

Problem Statement

Classifying ‘classical’ music by era of composition using audio files.

Classical music eras used in this project:

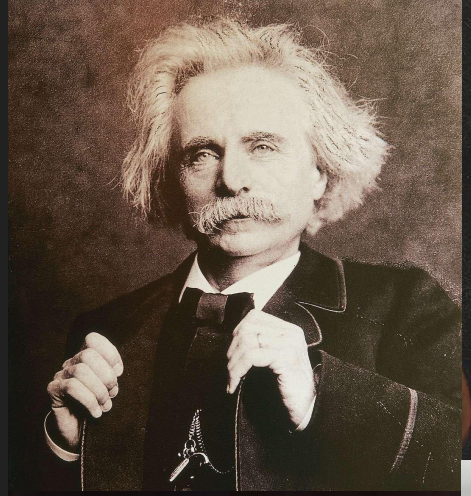
1. Baroque (1600-1750)
2. Classical (1720-1770)
3. Romantic (1750-1820)
4. Modern (1890-1950)

Outline

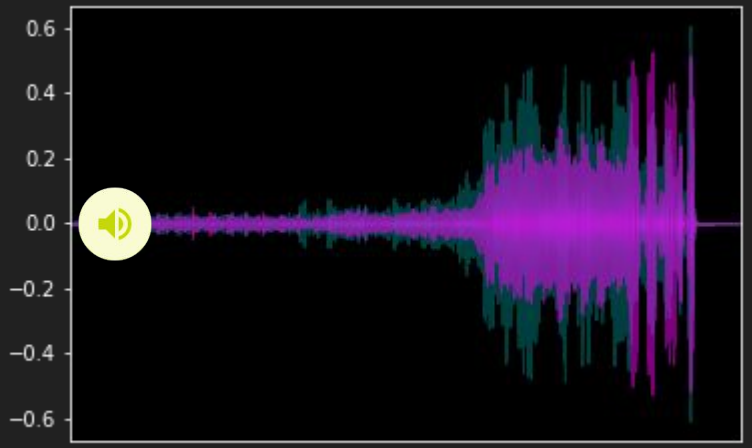
1. Problem statement
2. Raw data (.mp3)
3. Preprocess data (.ipynb)
   - ↳ Amplitudes (.csv)
4. Process data (.ipynb)
   - ↳ Melspectrogram images (.png)
   - ↳ Chromagram images (.png)
5. Model (.ipynb)
   - ↳ Convolutional neural networks (2D)
6. Future directions

Raw Data

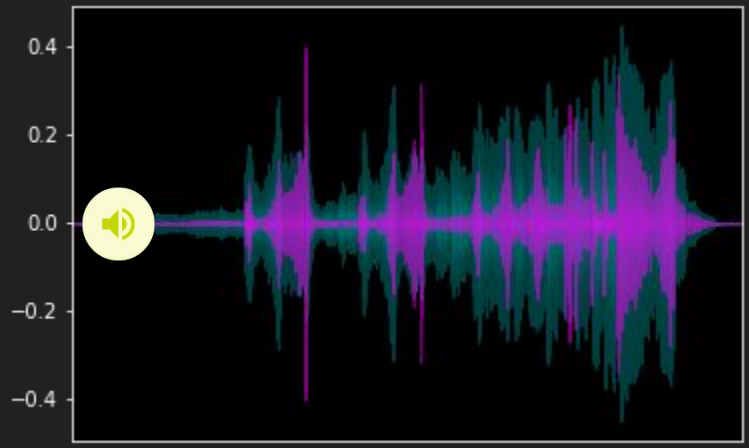
	Size	Songs	Format	Loss
Musopen	8.5 GB	145	.m4a	no
Spotify	63 MB	180	.mp3	yes



In the Hall of the Mountain King
Edvard Grieg



Also sprach Zarathustra
Richard Strauss



Preprocess

1. Load
2. Label
3. Chop


song
sample
sample
sample
sample
sample

csv

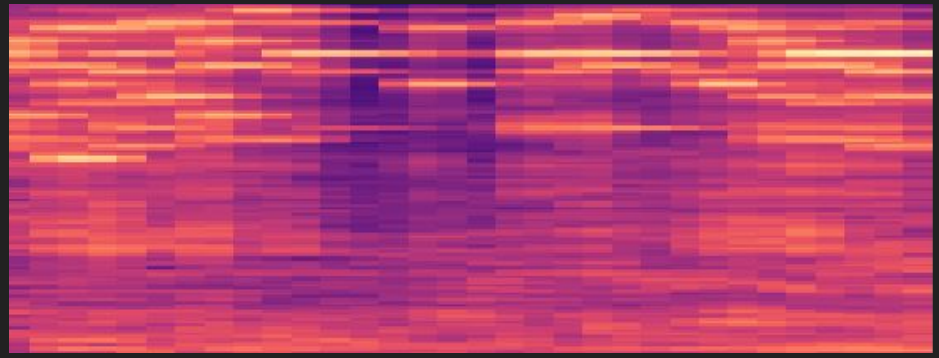
sample	song
sample	song
sample	song
sample	song
sample	song
sample	song

Process

1. Load
2. Image Generator
3. Save

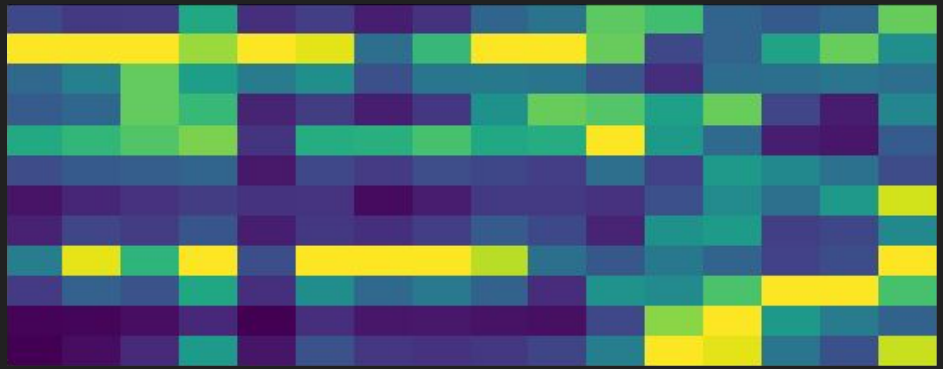
Mel-spectrogram

1. Plot Data - **Waveplot**
   - ↳ X = Amplitude
   - ↳ Y = Time
2. Fourier Transform x 1
   - ↳ X = Absolute amplitude
   - ↳ Y = Frequency band
3. Fourier Transform x All
   - ↳ X = Frequency
   - ↳ Y = Time
   - ↳ Heat = Amplitude
4. Scale - **Mel-Spectrogram**
   - ↳ X = Frequency (non-linear mel-scale)
   - ↳ Y = Time
   - ↳ Heat = Decibels



Chromagram

1. Plot Data - **Waveplot**
    - ↳ X = Amplitude
    - ↳ Y = Time
2. Transform - **Constant-Q Spectrogram**
    - ↳ X = Frequency bin / MIDI pitch
    - ↳ Y = Time
3. Transform - **Chromagram**
    - ↳ X = Pitch (ABCDEFG)
    - ↳ Y = Time
    - ↳ Heat = Energy in signal



Model

1. Load Images
2. Convert to Array
3. Train Neural Net

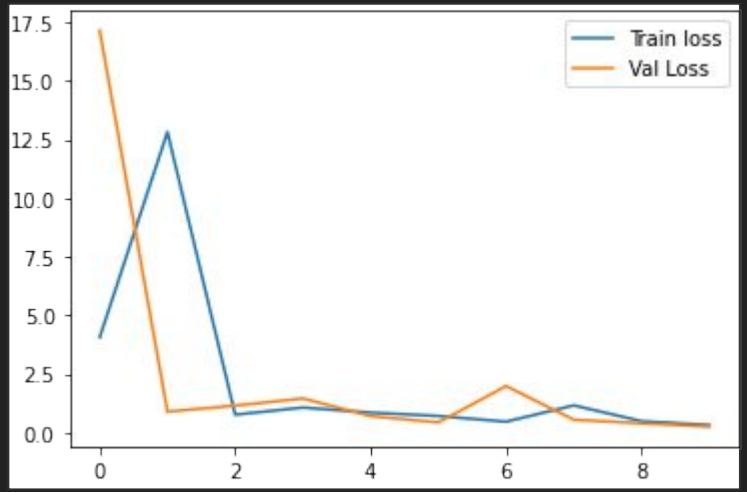
Convolutional 2D Neural Network

1. Sequential
2. Multi-class classifier
3. Layers
    - ↳ Convolutional 2D
        - ↳ Max Pooling
    - ↳ Convolutional 2D
        - ↳ Max Pooling
        - ↳ Flatten
    - ↳ Dense
    - ↳ Dense

Musopen Data

- 145 songs
- Unbalanced classes
- Mel-spectrogram images
- 5 10s samples per song

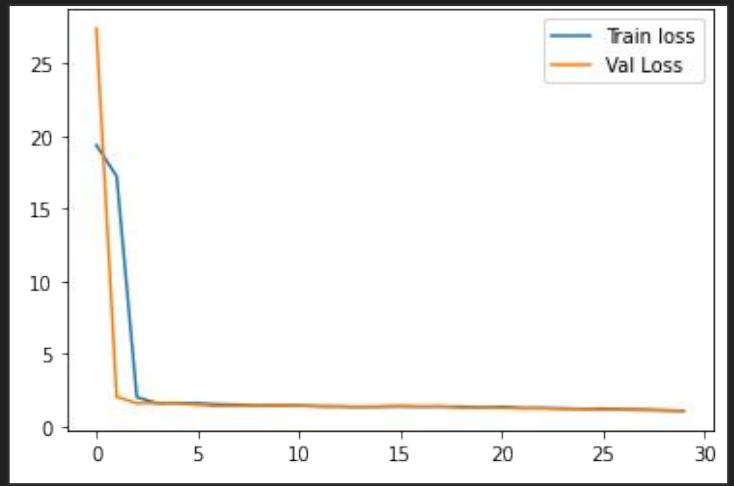
	Train	Validation
Loss	0.3119	0.2445
Accuracy	0.8679	0.9531



Spotify Data

- 200 songs
- Balanced classes
- Mel-spectrogram images
- 5 10s samples per song

	Train	Validation
Loss	1.0673	1.0518
Accuracy	0.5603	0.4844



Future Directions

1. More data (10,000 songs)
2. Beat extraction
3. Instrument extraction
4. Composer classification
5. Interactive application

Thank you!

Questions?