Music Genre Classification Using Deep Learning

{ jobn and mattmeza }@stanford.edu

Objective

Classify music into eight genres using deep learning.

- Use an approach that does not require lots of prior knowledge to hand-tailor the features
- Simplify task into an image classification problem to take advantage of CNNs

Dataset

- Free Music Archive (FMA) `http://freemusicarchive.org/`
- Used the *fma_small* data set described in table below and in [1]

Label	Genre	Number of Tracks
0	Electronic	1000
1	Experimental	1000
2	Folk	1000
3	Hip-Hop	1000
4	Instrumental	1000
5	International	1000
6	Pop	1000
7	Rock	1000

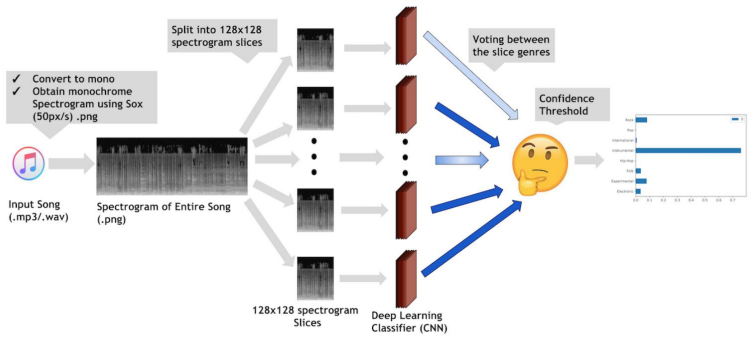
- Use SoX to get raw spectrograms from audio at 50px/s.
- Split each spectrogram into 128x128 pixels slices (2.56s of audio) resulting in about 80k images
- Split dataset into Train/dev/test in the ratio 0.90/0.05/0.05
- Use 5.12s audio clips as augmented data (Zoom out)

References

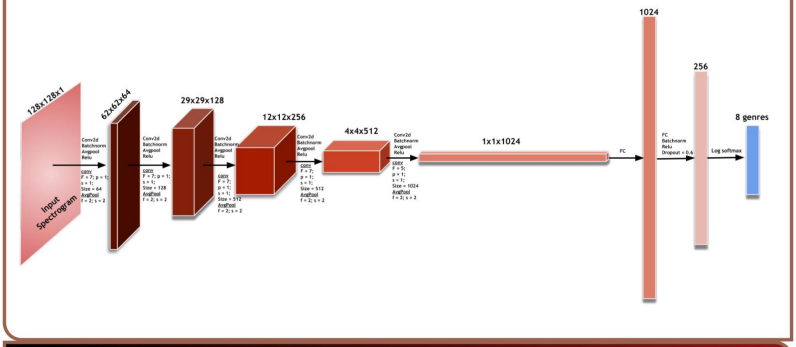
[1] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson "FMA: A Dataset for Music Analysis"

[2] S. Dielman, B. Schrauwen, "End to End Deep Learning for Music Audio," IEEE International conference on Acoustic, Speech and Signal Processing, 2014

Pipeline



CNN Architecture



Design Choices/Experiments

1. Optimizer: Adam, lr = 0.0001, BGD (Batch Size = 32)
2. Loss: Cross Entropy: $\ell = ylog(\hat{y})$. Realized using Pytorch logsoftmax.
3. Regularization:
   - Dropout rate of 0.6 for the FC layer
   - L2 regularization (W_{decay}) = $5e-05$
   - Data augmentation - 'Zoomed out spectrograms'

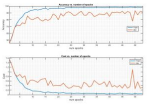
Results

Deep model: No regularization except dropout in FC layer (DR = 0.8)
Fusion: Searched Dropout rates and picked 0.6
WD: Added L2 regularization by searching weight decay values. Used Weight decay value of 1e-5
Best: See design choices above

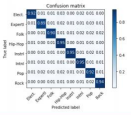
Performance summary

	Deep model	Fusion	WD	Best
Train	98	98.8	93.6	97.7
Eval	87.2	89.5	86.7	93.4
Test	87.3	89.2	86.9	92.7

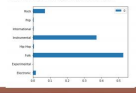
"Best" model



Performance on test set.



Enya - Gladiator



A Future Direction

- Modify the cost function to penalize misclassification (See confusion table)
- Figure out the jumps in the dev learning curve
- Multi-genre classification of Music
- Attempt Music Style Transfer between two genres

Source Code

The source code can be found in our Github repository at `https://github.com/jonalkn/cs230_final_project`