Music Genres Classification

John Uchal, Haotian Zhou

Methods

Our classification model is based on a deep neural network, which consists of 2 parts: a convolutional neural network and a feed forward network. In order to improve the efficiency and reduce the computation cost, we took advantage of the famous VGG model, and replaced the fully connected layers with our own layers.

The input data is raw audio data, thus some pre-processing and feature extraction process are needed. We concatenate all audio files to 20-second long, from which we compute Mel-spectrogram, MFCC and chromagram. The feature is presented as 2D images, so we feed those results into the VGG-based neural network, then the network will give the prediction.

The baseline model is a simple SVM classifier implemented with MATLAB using MFCC as the feature.

Problem

Genre classification plays an important role in how people consume music. In the ever growing world-music library, it is becoming increasingly difficult to find new artists/groups that a certain person would enjoy. A classifer's ability to analyze a user's audio history and make accurate recommendations for new music can greatly improve a user's music consuming experience. Current applications are not accurate enough to be useful to users.

Results

Baseline Confusion Matrix: Magnatune Dataset - 13 dimension MFCC

	Alt Rock	Ambient	Classical	Electro Rock	Electronica	Hard Rock	Jazz	New Age	World
Alt Rock	23.70%	12.26%	5.26%	16.76%	9.95%	13.94%	6.79%	6.38%	4.96%
Ambient	6.45%	36.75%	11.04%	7.01%	7.33%	3.35%	5.41%	13.83%	8.83%
Classical	5.48%	11.60%	41.50%	3.02%	1.14%	1.57%	8.66%	14.99%	12.04%
Electro Rock	14.90%	5.80%	3.19%	28.81%	17.28%	18.90%	1.70%	3.38%	6.04%
Electronica	13.59%	10.10%	3.06%	19.45%	27.52%	12.93%	3.02%	5.51%	4.82%
Hard Rock	6.72%	1.33%	0.92%	12.66%	3.94%	70.30%	2.16%	0.89%	1.08%
Jazz	14.18%	10.95%	10.49%	6.93%	13.16%	5.23%	15.10%	15.02%	8.95%
New Age	7.76%	17.25%	22.23%	5.80%	7.97%	1.55%	14.55%	17.73%	5.16%
World	6.26%	10.01%	18.68%	11.46%	6.63%	8.96%	6.41%	5.39%	26.20%

Deep Neural Network Confusion Matrix: Magnatune Dataset - MFCC

	Alt Rock	Ambient	Classical	Electro Rock	Electronica	Hard Rock	Jazz	New Age	World
Alt Rock	86%	0%	0%	2%	0%	6%	0%	0%	6%
Ambient	6%	72%	0%	0%	0%	0%	0%	10%	12%
Classical	0%	0%	90%	0%	0%	0%	0%	2%	8%
Electro Rock	12%	2%	0%	66%	0%	10%	2%	4%	4%
Electronica	16%	0%	0%	16%	60%	0%	8%	0%	0%
Hard Rock	2%	0%	0%	0%	0%	98%	0%	0%	0%
Jazz	10%	0%	0%	0%	0%	0%	78%	0%	12%
New Age	4%	12%	0%	0%	0%	0%	2%	64%	18%
World	4%	0%	0%	0%	0%	2%	2%	2%	90%

Deep Neural Network Confusion Matrix: Magnatune Dataset - Mel-Spectrogram

	Alt Rock	Ambient	Classical	Electro Rock	Electronica	Hard Rock	Jazz	New Age	World
Alt Rock	100%	0%	0%	0%	0%	0%	0%	0%	0%
Ambient	0%	96%	4%	0%	0%	0%	0%	0%	0%
Classical	0%	0%	100%	0%	0%	0%	0%	0%	0%
Electro Rock	2%	0%	6%	80%	8%	0%	0%	0%	4%
Electronica	0%	0%	4%	0%	92%	0%	2%	0%	2%
Hard Rock	6%	2%	2%	0%	0%	90%	0%	0%	0%
Jazz	2%	2%	12%	0%	2%	0%	74%	0%	8%
New Age	6%	8%	2%	0%	0%	0%	0%	80%	4%
World	2%	0%	0%	0%	2%	0%	0%	0%	96%

Datasets & Features

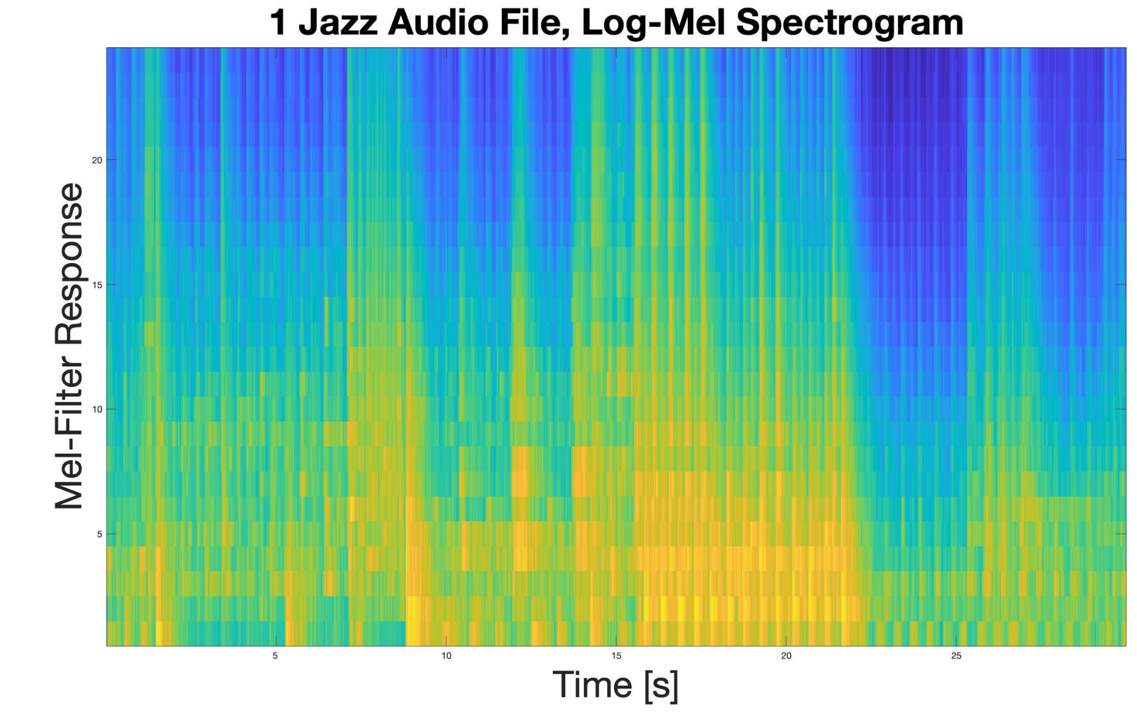
GTZAN dataset (fs = 22,050 Hz): 10 genres each with 100 thirty-second long audio files

Magnatune dataset (fs = 16,000 Hz): 9 genres each with 549 thirty-second long audio files

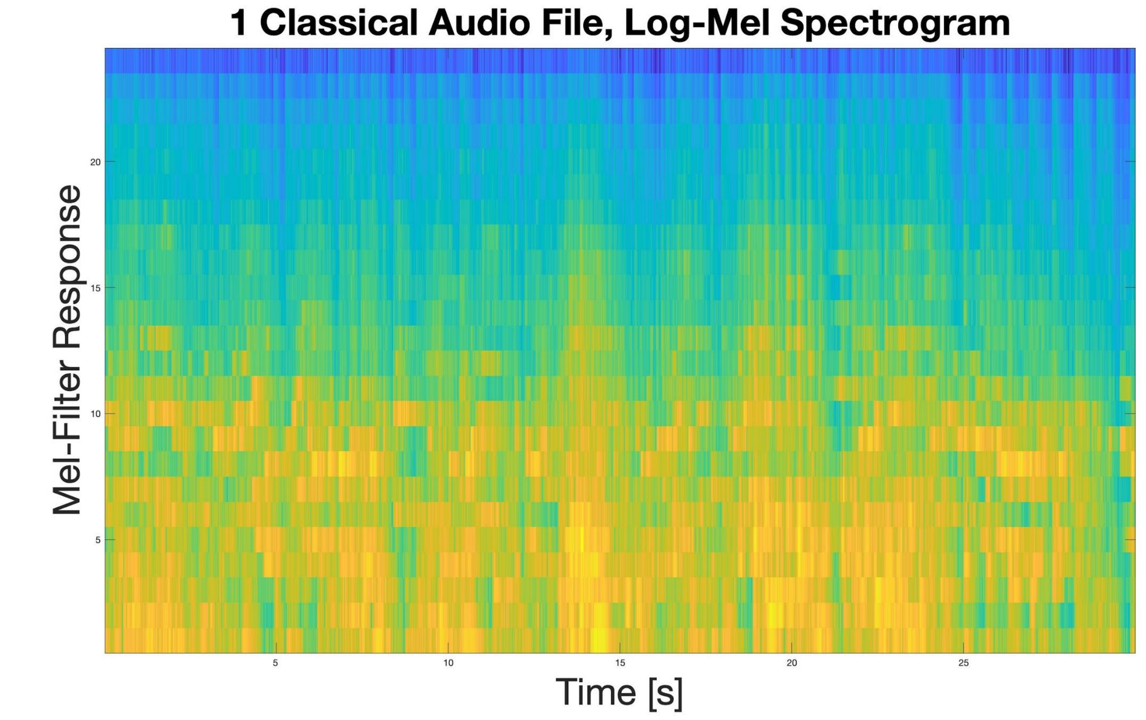
Features

- Mel-Spectrogram
- MFCC
- Chromagram

1 Jazz Audio File, Log-Mel Spectrogram



1 Classical Audio File, Log-Mel Spectrogram



Conclusion

	Precision	Recall	F1 Score	Support
Alt Rock	0.85	1.00	0.92	50
Ambient	0.89	0.96	0.92	50
Classical	0.77	1.00	0.87	50
Electro Rock	1.00	0.80	0.89	50
Electronica	0.88	0.92	0.90	50
Hard Rock	1.00	0.90	0.95	50
Jazz	0.97	0.74	0.84	50
New Age	1.00	0.80	0.89	50
World	0.84	0.96	0.90	50
Micro Avg.	0.90	0.90	0.90	450
Macro Avg.	0.91	0.90	0.90	450
Weighted Avg.	0.91	0.90	0.90	450

From the result above, we conclude that our model performed a state-of-the-art result on our dataset, as the average precision, recall and F-score all reached 90%.

Acknowledgements

Thank you to Professor Duan and all TAs for their assistance in this class and for this project.

Reference:

[1] **Bahuleyan, H.**, 2018. Music Genre Classification using Machine Learning Techniques. https://arxiv.org/pdf/1804.01149.pdf. Accessed: 2019-04-30

[2] **McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., Nieto, O.**, 2015. librosa: Audio and Music Signal Analysis in Python. http://conference.scipy.org/proceedings/scipy2015/pdfs/brian_mcfee.pdf. Accessed: 2019-04-30

[3] **Feng, T.**, 2014. Deep learning for music genre classification. https://courses.engr.illinois.edu/ece544na/fa2014/Tao_Feng.pdf. Accessed: 2019-04-30

[4] **Lachmish, M.**, 2018. MusicGenreClassification. https://github.com/mlachmish/MusicGenreClassification. Accessed: 2019-04-30