UMSI

Audio Source Separation: Music Feature Extraction

UNIVERSITY OF MICHIGAN

2020 Expo

Students: Hsin-Yuan Wu Advisor: Prof. Paramveer Dhillon

Introduction

The process of audio source separation is to isolate or extract one or more signals from a mixture of audio sources.

- Separate lead vocals from a music recording, like karaoke
- help with the event/speech detection and improve the audio identification

Objective

- Using the extracted features and Deep learning model to predict and generate the vocal signals, i.e. separate vocals from music

Design of the methodology

- Dataset

MUSDB18 dataset contains 150 full songs, 100 in training set and 50 in test set. We separated the sequence with 6 seconds duration and sampling rate is 64 frames per song.

- Baseline Model, open-unmix[1]

a 3-layer bidirectional deep LSTM trains and predicts the magnitude spectrogram from a mixture of magnitude spectrograms by applying a mask on the input, and separates the signals in the post-processing step via a multichannel wiener filter. [1]

- Proposed Model, adding MelSpectrogram:

Based on open-unmix model, we used mixture spectrogram as well as Mel-scale spectrogram, running 2 separate 3 layers BLSTM and then add the feature together to predict vocal spectrograms.

- Mel Frequency Cepstral Coefficients:

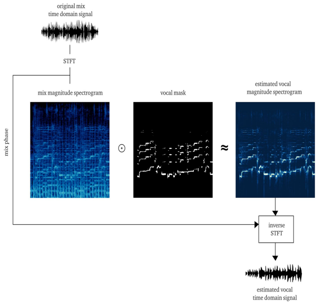
The feature extraction is to convert signals to the mel scale, which frames the audio into short frames and calculates the power spectrum on a non-linear mel scale.

- Hyper parameter tuning:

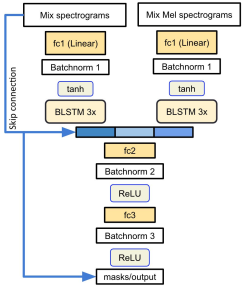
With our proposed model, we use training set with lower sampling rate to tune hyperparameters, like learning rate and decay rate.

- Training and validation:

With limited time and resources, we ran 20 epochs and have MSE loss around 2, similar to the baseline, though not a big improvement.



Isolation process [2]

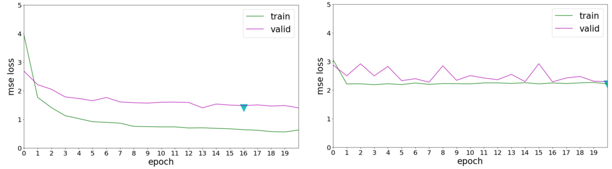


Model

Analysis and Evaluation

- Baseline and Proposed Model: MSE loss

Train_loss: 0.62 vs 2.28; Valid_loss: 1.40 vs 2.34



- Performance result (dB)

median	SDR	SIR	ISR	SAR
Baseline	1.776	0.791	2.399	4.409
Proposed	**1.998**	0.321	**2.447**	**10.384**

Result and Conclusion

- Similar results compared to the Baseline
- Although adding the Mel scale spectrogram seems no big enhancement with this limited epochs, yet we might need to do more epochs for more validation.
- With the property of MFCC to lessen noise, we might investigate more on application.

References

1. https://github.com/sigsep/open-unmix-pytorch
2. https://www.elasticfeed.com/a851a2e8c45813e338ccf90d8fb3178e/