MS3 Audio Separation Using Deep Neural Networks

ECE685 Introduction to Deep Learning
Fall 2020

- Students: Jiajun Song, Heather Qiu
- TA: Mohammadreza Soltani

Background & Problem

- **Audio Separation:**

  Techniques can be extended to other disciplines with similar goals (e.g. image processing, digital communications).

- **Blind Source Separation**:

  Separating different source signals from a set of mixed signals.

Cocktail Party Problem

"I hear he won the Pulitzer for mathematics. Or was it the Emmy?"

(figure from MIP-Frontiers post)

Methods - ICA

Independent Component Analysis

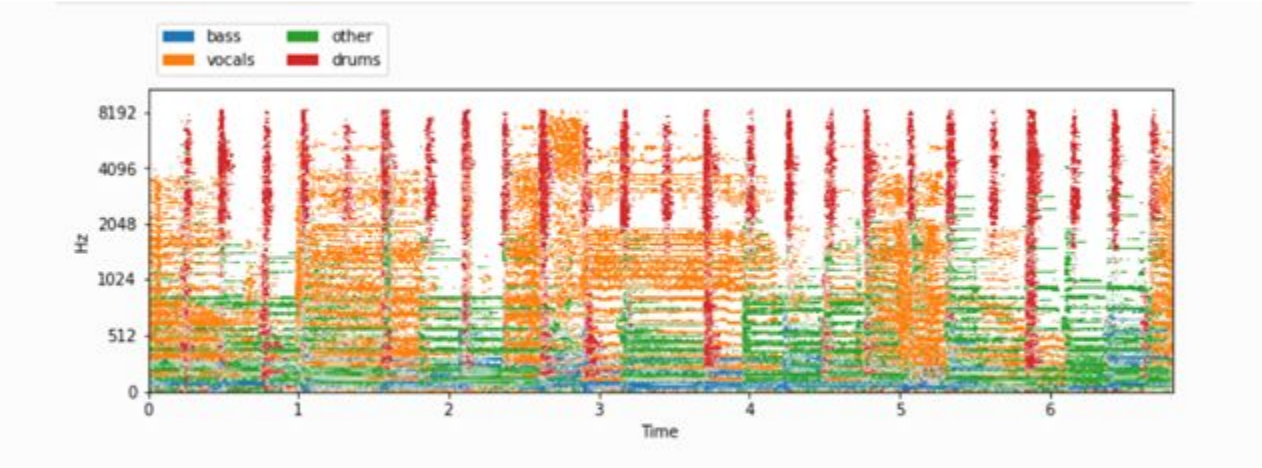
- Based on a **linear** combination of **statistically independent** sources
- Tries to maximize the **independence of output signals**

Drawbacks

- Restrictive **assumption** that rarely holds in reality
- Too **simple** to fully represent the true data
- Bad performance in generalization and sensitive to noise

Dataset: MUSDB18

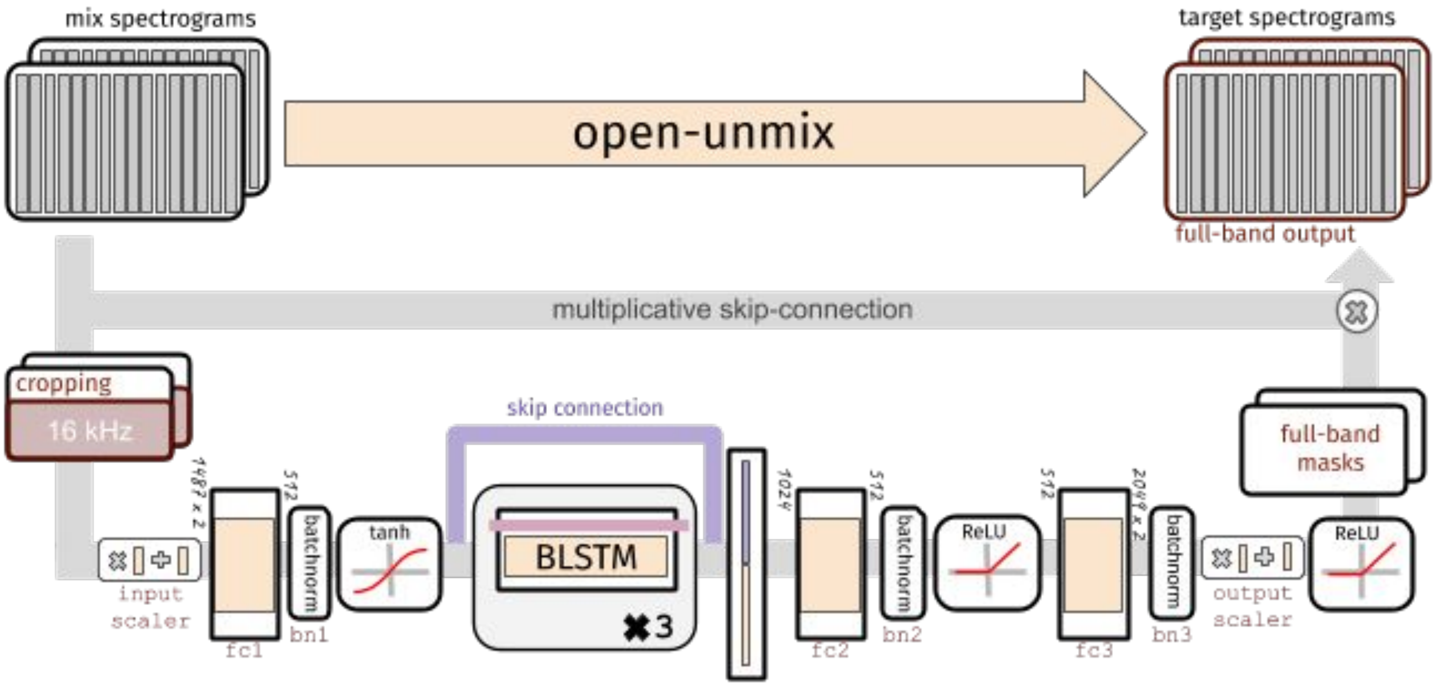
- Music audio recorded as a mixture of drums, bass, vocals, and other stems
- 150 full-length music tracks, total duration ~ 10 hrs
- More challenging than the removal of unrelated background noise



Methods - Data Preprocessing

- Data Augmentation
  - Chunking
  - Source augmentation
  - Random Track Mixing
- Dataset Robustness
  - Input/Output Standard Scaler
  - Fixed Validation Split

Methods: Model Architecture



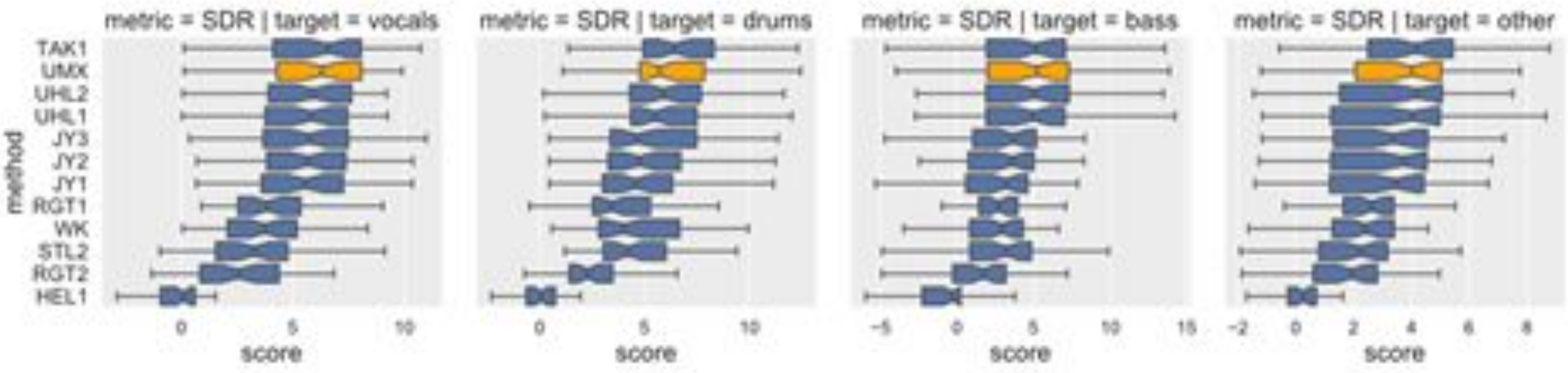
Methods: Metrics

- Precision / Recall: calculated on the binary masks (require ad-hoc threshold)

- SDR: Source-to-distortion ratio
- SIR: Source-to-interface ratio
- SAR: Source-to-artifact ratio
- ISR: Source-to-Spatial-Distortion Image

Results

- Source-to-distortion ratio (SDR): how well does the estimate match the ground truth source?



Conclusion & Future Work

- Open-unmix works effectively in music audio separation
- The performance decreases greatly when the length of the audio becomes longer
- Spectrogram-based / Waveform-based