Project Synopsis for 02456 Deep Learning

Matej Hoffmann (s221913), Chuansheng Liu (s212661),
Sturla Njarðarson (s222676), Dominik Stiller (s221811)

November 1, 2022

Title Speech separation in the waveform domain

Supervisor Bjørn Sand Jensen (bjje@dtu.dk) and WS Audiology

Motivation Separating mixed signals into their respective sources is an open problem in machine learning. This includes the separation of music into stems, and conversations into speaker utterances. For the application of hearing aids, access to individual speaker signals allows the reduction of background noise and focus on speakers. In this project, we will adapt the Demucs model [1], originally created for stereo music, to speech separation. To the best of our knowledge, this has not been attempted before. We evaluate the performance on noisy mono two-speaker mixtures from the LibriMix dataset [2] using quantitative metrics and human opinions.

Background The Demucs deep learning model is described in [1]. It operates in the waveform domain instead of the spectrogram, taking the audio mixture as input and giving the separated signals as output. Based on the U-net model, the number of channels is increased by a series of 1D convolutions in the encoder, with two LSTM layers at the bottleneck. The skip connections to the decoder aid in generating the correct phase. Both the encoder and decoder contain GLUs for masking.

Milestones

- 31 October: initial meeting with supervisor
- 7 November: get familiar with data and model from paper
- 21 November: implement model and achieve useful results
- 5 December: improve and quantify model performance for data with more than 2 speakers
- 8 December: create poster for poster session
- 4 January: write report

References

[1] A. Défossez, N. Usunier, L. Bottou, and F. Bach, *Music source separation in the waveform domain*, 2019. DOI: 10.48550/ARXIV.1911.13254.

[2] J. Cosentino, M. Pariente, S. Cornell, A. Deleforge, and E. Vincent, *Librimix: An open-source dataset for generalizable speech separation*, 2020. DOI: 10.48550/ARXIV.2005.11262.