A must-read paper and tutorial list for speech separation based on neural networks

This repository contains papers for pure speech separation and multimodal speech separation.

By Kai Li (if you have any suggestions, please contact me! Email: tsinghua.kaili@gmail.com).

Tip: For speech separation beginners, I recommend you to read "deep clustering" & "PIT&uPIT" works which will help understand the problem.

If you have found the code for some of the articles below, welcome to add links.

‼ New board: New papers are introduced every week ! Weekly_Report.md

Pure Speech Separation

✔ [Joint Optimization of Masks and Deep Recurrent Neural Networks for Monaural Source Separation, Po-Sen Huang, TASLP 2015] [Paper] [Code (posenhuang)]

✔ [Complex Ratio Masking for Monaural Speech Separation, DS Williamson, TASLP 2015] [Paper]

✔ [Deep clustering: Discriminative embeddings for segmentation and separation, JR Hershey, ICASSP 2016] [Paper] [Code (Kai Li)] [Code (funcwj)] [Code (asteroid)]

✔ [Single-channel multi-speaker separation using deep clustering, Y Isik, Interspeech 2016] [Paper] [Code (Kai Li)] [Code (funcwj)]

✔ [Permutation invariant training of deep models for speaker-independent multi-talker speech separation, Dong Yu, ICASSP 2017] [Paper] [Code (Kai Li)]

✔ [Recognizing Multi-talker Speech with Permutation Invariant Training, Dong Yu, ICASSP 2017] [Paper]

✔ [Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks, M Kolbæk, TASLP 2017] [Paper] [Code (Kai Li)]

✔ [Deep attractor network for single-microphone speaker separation, Zhuo Chen, ICASSP 2017] [Paper] [Code (Kai Li)]

✔ [A consolidated perspective on multi-microphone speech enhancement and source separation, Sharon Gannot, TASLP 2017] [Paper]

✔ [Alternative Objective Functions for Deep Clustering, Zhong-Qiu Wang, ICASSP 2018] [Paper]

✔ [End-to-End Speech Separation with Unfolded Iterative Phase Reconstruction Zhong-Qiu Wang et al. 2018] [Paper]

✔ [Speaker-independent Speech Separation with Deep Attractor Network, Luo Yi, TASLP 2018] [Paper] [Code (Kai Li)]

✔ [Tasnet: time-domain audio separation network for real-time, single-channel speech separation, Luo Yi, ICASSP 2018] [Paper] [Code (Kai Li)] [Code (asteroid)]

✔ [Supervised Speech Separation Based on Deep Learning An Overview, DeLiang Wang, Arxiv 2018] [Paper]

✔ [An Overview of Lead and Accompaniment Separation in Music, Zafar Rafi, TASLP 2018] [Paper]

✔ [Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation, Luo Yi, TASLP 2019] [Paper] [Code (Kai Li)] [Code (asteroid)]

✔ [Divide and Conquer: A Deep CASA Approach to Talker-independent Monaural Speaker Separation, Yuzhou Liu, TASLP 2019] [Paper] [Code]

✔ [Dual-path RNN: efficient long sequence modeling for time-domain single-channel speech separation, Luo Yi, Arxiv 2019] [Paper] [Code (Kai Li)]

✔ [End-to-end Microphone Permutation and Number Invariant Multi-channel Speech Separation, Luo Yi, Arxiv 2019] [Paper] [Code]

✔ [FaSNet: Low-latency Adaptive Beamforming for Multi-microphone Audio Processing, Yi Luo , Arxiv 2019] [Paper]

✔ [A comprehensive study of speech separation: spectrogram vs waveform separation, Fahimeh Bahmaninezhad, Interspeech 2019] [Paper]

✔ [Discriminative Learning for Monaural Speech Separation Using Deep Embedding Features, Cunhang Fan, Interspeech 2019] [Paper]

✔ [FaSNet: Low-latency Adaptive Beamforming for Multi-microphone Audio Processing, Yi Luo, Arxiv 2019] [Paper]

✔ [Interrupted and cascaded permutation invariant training for speech separation, Gene-Ping Yang, ICASSP, 2020][[Paper]](https://arxiv.org/abs/1910.12706)

✔ [FurcaNeXt: End-to-end monaural speech separation with dynamic gated dilated temporal convolutional networks, Liwen Zhang, MMM 2020] [Paper]

✔ [Filterbank design for end-to-end speech separation, Manuel Pariente et al., ICASSP 2020] [Paper]

✔ [Voice Separation with an Unknown Number of Multiple Speakers, Eliya Nachmani, Arxiv 2020] [Paper] [Demo]

✔ [AN EMPIRICAL STUDY OF CONV-TASNET, Berkan Kadıoglu , Arxiv 2020] [Paper] [Code]

✔ [Voice Separation with an Unknown Number of Multiple Speakers, Eliya Nachmani, Arxiv 2020] [Paper]

✔ [Wavesplit: End-to-End Speech Separation by Speaker Clustering, Neil Zeghidour et al. Arxiv 2020] [Paper]

✔ [La Furca: Iterative Context-Aware End-to-End Monaural Speech Separation Based on Dual-Path Deep Parallel Inter-Intra Bi-LSTM with Attention, Ziqiang Shi, Arxiv 2020] [Paper]

✔ [Enhancing End-to-End Multi-channel Speech Separation via Spatial Feature Learning, Rongzhi Gu, Arxiv 2020] [Paper]

✔ [Deep Attention Fusion Feature for Speech Separation with End-to-End Post-filter Method, Cunhang Fan, Arxiv 2020] [Paper]

✔ [Enhancing End-to-End Multi-channel Speech Separation via Spatial Feature Learning, Rongzhi Guo, ICASSP 2020] [Paper]

✔ [A Multi-Phase Gammatone Filterbank for Speech Separation Via Tasnet, David Ditter, ICASSP 2020] [Paper] [Code]

✔ [Lightweight U-Net Based Monaural Speech Source Separation for Edge Computing Device, Kwang Myung Jeon, ICCE 2020] [Paper]

✔ [LibriMix: An Open-Source Dataset for Generalizable Speech Separation, Joris Cosentino, Arxiv 2020] [Paper] [Code]

✔ [An End-to-end Architecture of Online Multi-channel Speech Separation, Jian Wu, Interspeech 2020] [Paper]

✔ [SAGRNN: Self-Attentive Gated RNN for Binaural Speaker Separation with Interaural Cue Preservation, Ke Tan, IEEE Signal Processing Letters] [Paper]

✔ [A convolutional recurrent neural network with attention framework for speech separation in monaural recordings, Chao Sun, Scientific Reports] [Paper]

✔ [Unsupervised Sound Separation Using Mixture Invariant Training, Scott Wisdom, NeurIPS 2020] [Paper]

✔ [Causal Deep CASA for Monaural Talker-Independent Speaker Separation, Yuzhou Liu, TASLP 2019] [Paper]

✔ [Sparse, Efficient, and Semantic Mixture Invariant Training: Taming In-the-Wild Unsupervised Sound Separation, Scott Wisdom, Arxiv 2021] [Paper]

✔ [Tune-In: Training Under Negative Environments with Interference for Attention Networks Simulating Cocktail Party Effect, Jun Wang, Arxiv 2021] [Paper]

✔ [Speech Separation Using an Asynchronous Fully Recurrent Convolutional Neural Network, Kai Li, NeuralPS 2021] [Paper] [Code]

Multi-Model Speech Separation

✔ [Audio-Visual Speech Enhancement Using Multimodal Deep Convolutional Neural Networks, Jen-Cheng Hou, TETCI 2017] [Paper] [Code]

✔ [The Conversation: Deep Audio-Visual Speech Enhancement, Triantafyllos Afouras, Interspeech 2018] [Paper]

✔ [End-to-end audiovisual speech recognition, Stavros Petridis, ICASSP 2018] [Paper] [Code]

✔ [The Sound of Pixels, Hang Zhao, ECCV 2018] [Paper] [Code]

✔ [Looking to Listen at the Cocktail Party: A Speaker-Independent Audio-Visual Model for Speech Separation, ARIEL EPHRAT, ACM Transactions on Graphics 2018] [Paper] [Code]

✔ [Learning to Separate Object Sounds by Watching Unlabeled Video, Ruohan Gao, ECCV 2018] [Paper]

✔ [Time domain audio visual speech separation, Jian Wu, Arxiv 2019] [Paper]

✔ [Audio-Visual Speech Separation and Dereverberation with a Two-Stage Multimodal Network, Ke Tan, Arxiv 2019] [Paper]

✓ [Co-Separating Sounds of Visual Objects, Ruohan Gao, ICCV 2019] [Paper] [Code]

✓ [AudioVisual Deep Clustering for Speech Separation, Rui Lu, TASLP 2019] [Paper]

✓ [Multi-modal Multi-channel Target Speech Separation, Rongzhi Guo, J-STSP 2020] [Paper]

✓ [An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation, Daniel Michelsanti, Arxiv 2020] [Paper]

✓ [Deep Audio-Visual Speech Separation with Attention Mechanism, Chenda Li, ICASSP 2020] [Paper]

✓ [Looking into Your Speech: Learning Cross-modal Affinity for Audio-visual Speech Separation, Jiyoung Lee, CVPR 2021] [Paper] [Demo Page]

✓ [Audio-Visual Speech Separation Using Cross-Modal Correspondence Loss, Naoki Makishima, ICASSP 2021] [Paper]

✓ [VisualVoice: Audio-Visual Speech Separation with Cross-Modal Consistency, Ruohan Gao, CVPR 2021] [Paper] [Demo Page] [Code]

Evaluation index

✓ [Performance measurement in blind audio sourceseparation, Emmanuel Vincent et al., TASLP 2004] [Paper]

✓ [SDR – Half-baked or Well Done?, Jonathan Le Roux, ICASSP 2019] [Paper]

Dataset

✓ [WSJ0] [Dataset Link]

✓ [WHAM & WHAMR] [Dataset Link]

✓ [Microsoft DNS Challenge] [Dataset Link]

✓ [AVSpeech] [Dataset Link]

✓ [LRW] [Dataset Link]

✓ [LRS2] [Dataset Link]

✓ [LRS3] [Dataset Link] [Multi Model Data Processing Script]

✓ [VoxCeleb] [Dataset Link]

✓ [LibriMix] [Dataset Link]

✓ [LibriSS] [Dataset Link]

Video Tutorial

✓ [Speech Separation, Hung-yi Lee, 2020] [Video] [Slide]

> I may not be able to get all the articles completely. So if you have an excellent essay or tutorial, you can update it in my format. At the same time, if you think the repository meets your needs, please give a star or fork, thank you.