CommIT
Communications and Information Theory Chair

DEMUCS for data-driven RF signal denoising

ICASSP 2024, Seoul, South Korea

Çağkan Yapar (TUB), Fabian Jaensch (TUB), Jan C. Hauffen (TUB), Francesco Pezone (SAP), Peter Jung (TUB, DLR), Saeid K. Dehkordi (TUB), Giuseppe Caire (TUB)

DLR

Technische Universität Berlin

Introduction

- Traditional radio frequency (RF) signal separation (denoising/interference rejection) methods typically rely on prior knowledge of the interfering signal model, the assumption of non-overlapping time/frequency bands of the signal-of-interest (SOI) and the interfering signal, or the availability of multiple antennas.
- These requirements are not met in many realistic wireless communication scenarios.
- A reasonable choice for problems with a lack of accurate modeling and an abundance of data is to employ data driven solutions, specifically deep neural networks (DNN).
- DNNs for many signal denoising tasks, e.g. for seismic signals, gravitational waves and audio signals.

Technische Universität Berlin

Introduction (Time-frequency and and time domain methods)

Existing single-channel signal denoising/separation methods can be divided into two categories: time-frequency (TF) domain methods and time domain methods.

Time-frequency domain solutions either learn a spectral mapping from a TF representation of a noisy signal to the SOI, or estimate a mask approximating the SOI's position in each TF bin of a noisy spectrum. However, most TF domain methods recover only the amplitude and combine it with the noise phase to recover the waveform.

Moreover, as we see in the following figures, the spectrogram of communication signals is much more dense than that of speech signals due to their effective use of the frequency spectrum, making TF masking-based approaches less efficient.

The time-domain models, such as DEMUCS [1] and WaveNet [2], operate directly on the raw waveform, naturally preserving the phase information in the signal during processing.

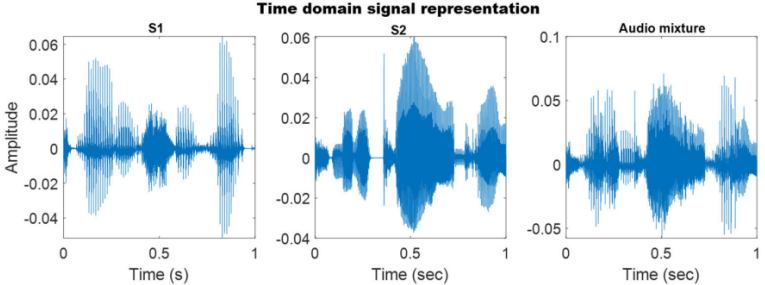

Figure: Time domain (top row) and time frequency domain (bottom row) representation of speech sources and their resulting audio mixtures. The red boxes show the portions of S1, which are still identifiable in the spectrogram of audio mixture. Figures and caption taken from **S. Gul and M. S. Khan, "A Survey of Audio Enhancement Algorithms for Music, Speech, Bioacoustics, Biomedical, Industrial, and Environmental Sounds by Image U-Net," IEEE Access, vol. 11, pp. 144456-144483, 2023.**

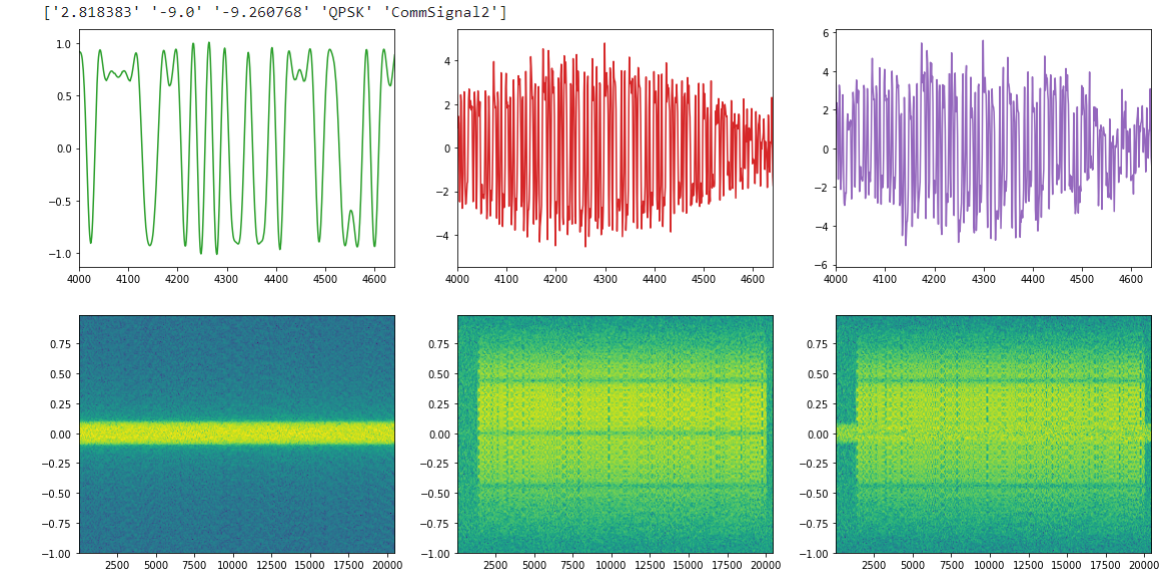

Figure: Time domain (top row) and time frequency domain (bottom row) representation of SOI signal and example noise type and their resulting RF signal mixtures. The SOI signal is not easily identified in the spectrogram. Figures taken from the challenge repository `https://github.com/RFChallenge/icassp2024rfchallenge/blob/0.2.0/notebook/RFC_Demo.ipynb`

Technische Universität Berlin

Introduction (Learning long-range relations)

- CNNs cannot well model the long-range dependencies in signals due to the limitations of the local receptive field of the convolution filters. Receptive fields of CNNs can be enlarged by increasing the network depth or enlarging the kernel size, however, these are not effective.

- Recurrent neural networks (RNN) and long-short-term memory (LSTM) can capture the features of sequences well, however, they are often unsuited for long sequences as they lack the ability to effectively capture the long-range relationships.

- The self-attention mechanism processes the entire waveform and leads to the consideration of the long-term dependencies in the sequences. However, the self-attention network may not be able to capture the subtle local details in an effective way.

Technische Universität Berlin

Prior failed attempts

- **CleanUNet:** Z. Kong, W. Ping, A. Dantrey and B. Catanzaro, "Speech Denoising in the Waveform Domain With Self-Attention," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7867-7871. Also adopted by one of the challenge participants: F. Damara, Z. Utkovski, S. Stanczak, "Signal separation in radio spectrum using self-attention mechanism," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, South Korea, 2024.

- **MANNER:** H. J. Park, B. H. Kang, W. Shin, J. S. Kim and S. W. Han, "MANNER: Multi-View Attention Network For Noise Erasure," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7842-7846.

Technische Universität Berlin

DEMUCS

- U-Net-variant with LSTM bottleneck.
- Strided convolutions allow large receptive fields.
- Gated Linear Unit (GLU) activation.
- Unidirectional or bidirectional LSTM (we considered the latter) layer between the encoder and decoder promotes improved sequential modeling.
- U-Net skip connections enable high resolution information preservation during network propagation.
- Characterized by its number of layers L in encoder/decoder, initial number of hidden channels H, layer kernel size K, stride S and resampling factor U.

DEMUCS Architecture

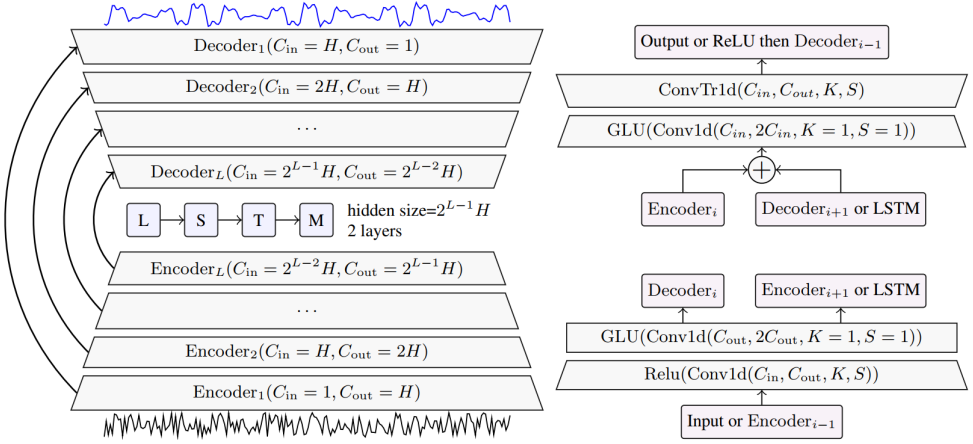

Figure: **Left:** Causal DEMUCS with the noisy speech as input on the bottom and the clean speech as output on the top. Arrows represents UNet skip connections. H controls the number of channels in the model and L its depth. **Right:** View of each encoder (bottom) and decoder layer (top). Arrows are connections to other parts of the model. C_{in} (resp. C_{out}) is the number of input channels (resp. output), K the kernel size and S the stride. Figures and caption taken from **A. Défossez, G. Synnaeve, and Y. Adi, "Real time speech enhancement in the waveform domain," Proc. Interspeech, 2020. For higher accuracy, we have employed the non-causal DEMUCS, i.e., used bi-directional LSTM in the bottleneck instead of the uni-directional LSTM shown above. Note also that $C_{\text{in}} = C_{\text{out}} = 2$ (instead of 1) in the considered task of RF signal denoising.**

DEMUCS for RF Signal Denoising and Bit Regression/End-to-end Modulation

- For QPSK + CommSignal3: $H = 80$ and $S = U = 4$, for the rest: $H = 64$, $S = U = 2$. $K = 8$ and $L = 5$ for all.

- Unlike [1], we did not normalize the input by its standard deviation, as it has been our experience that this is detrimental to performance.

- An extension of the DEMUCS for direct bit regression by appending the DEMUCS architecture with a fully connected layer that is applied to consecutive disjoint blocks of an appropriately chosen number of output samples (e.g., 64 for QPSK SOI) of DEMUCS and outputs the bits (e.g., 8 bits for QPSK) during the inference phase after applying a “hard decision” on the threshold value of 0.5.

- This idea of extending DEMUCS was inspired by the “Bit Regression” baseline method from the Single-Channel RF Challenge.

- We also later noticed a similar approach in the doctoral thesis "Machine Learning for Data-Driven Signal Separation and Interference Mitigation in Radio-Frequency Communication Systems" by Gary Lee, calling such approach as **end-to-end modulator**.

Technische Universität Berlin

Training

- For SOI estimation and bit regression, we use the MSE loss for training.
- We also considered training the bit regression DNN with cross entropy loss.
- For the QPSK+CommSignal2 and QPSK+CommSignal3 cases, we used a learning rate of $3 \cdot 10^{-4}$, otherwise $3 \cdot 10^{-5}$.
- We used ReduceLROnPlateau scheduler with a patience of 3 and EarlyStopping with a patience of 12 and Adam optimizer.

Technische Universität Berlin

Results

- DEMUCS closely follows WaveNet in most scenarios, yielding an overall MSE score of DEMUCS (WaveNet) -118.71 (-119.35) and a BER score of DEMUCS (WaveNet) -81 (-78) in the challenge test set.

- Almost (except for QPSK + CommSignal5G1) in all considered settings, the estimation of the SOI followed by the matched filtering baseline method yielded slightly better BER performance than the bit regression DEMUCS.

- Another observation is that for bit regression, using the MSE loss instead of the cross-entropy loss lead to slightly higher accuracy.

- Separate training for each SINR value resulted in worse results.

Technische Universität Berlin

Results

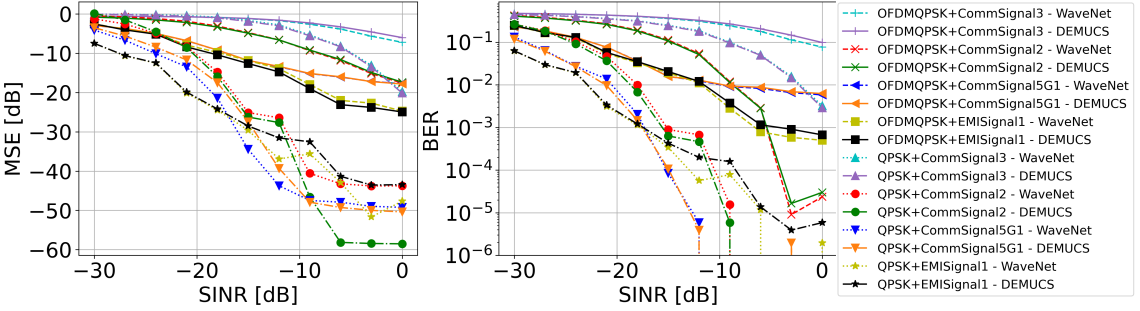

Figure: Comparison of the test MSE (Mean Squared Error) and BER (Bit Error Rate) accuracies of the baseline WaveNet [2] provided by the challenge organizers with those of the DEMUCS [1] architecture we adopted. **Left:** MSE, **Right:** BER.

[1] A. Défossez, G. Synnaeve, and Y. Adi, "Real time speech enhancement in the waveform domain", Proc. Interspeech, 2020.

[2] D. Rethage, J. Pons, and X. Serra, "A wavenet for speech denoising", Proc. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5069–5073.

Model size, training time

	QPSK				OFDMQPSK			
	Comm2	**Comm3**	**EMI**	**5G**	**Comm2**	**Comm3**	**EMI**	**5G**
# parameters	60.81 M	95.01 M	60.81 M	60.82 M	60.81 M	60.81 M	60.81 M	60.81 M
# macs	194.1 G	37.52 G	194.1 G	194.1 G	194.1 G	194.1 G	194.1 G	194.1 G
model size	696 MB	1 GB	696 MB	696 MB	696 MB	696 MB	696 MB	696 MB
# GPUs	1	1	1	1	1	1	1	1
GPU	RTX A6000	Quadro RTX 6000	Quadro RTX 6000	Quadro RTX 6000	Quadro RTX 6000	RTX A6000	Quadro RTX 6000	Quadro RTX 6000
# epoch w/ best val. loss	19	7	22	52	43	24	52	42
time per epoch (hr:min)	02:18	01:48	03:10	03:05	03:10	02:20	03:06	03:10

The main drawback of DEMUCS compared to most other architectures is its size.

Technische Universität Berlin

Remark: An unexpected overfitting case

- The common practice of early stopping based on validation loss failed in the QPSK + CommSignal3 setting for the DEMUCS architecture, as the validation loss continued to decrease along with the training loss.
- The training/validation dataset generation script provided by the challenge organizers is based on extracting random frames from a “global” dataset, INTERFERENCESET, and thus the training and validation datasets are not guaranteed to be disjoint, which may have played a role in the overfitting in this exceptional case.
- To remedy the overfitting, in this setting we used the TESTSET1EXAMPLE dataset for validation, which is not part of the INTERFERENCESET.

Technische Universität Berlin

Resources

- **Paper** *"DEMUCS for data-driven RF signal denoising"* to appear in ICASSPW 2024 proceedings, already available on the **challenge home page.**
- **Datasets** available on the **challenge home page:** `https://rfchallenge.mit.edu/icassp24-single-channel/`
- **Code** of DEMUCS adaptation to RF denoising and its modification into end-to-end demodulator on **GitHub:** `https://GitHub.com/CagkanYapar/RFDemucs`
- **Original implementation** of DEMUCS by **Facebook Research** on **GitHub:** `https://GitHub.com/FacebookResearch/Denoiser/`

Thank you!