


RESEARCH

Open Access



Time-Domain Anti-Interference Method for Ship Radiated Noise Signal

Yichen Duan^{1*} , Xiaohong Shen¹ and Haiyan Wang²

*Correspondence:
876123775@qq.com

¹ School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China

² School of Electronic and Information, Xi'an Polytechnic University, Xi'an, China

Abstract

Ship radiated noise signal is one of the important ways to detect and identify ships, and emission of interference noise to shield its own radiated noise signal is a common countermeasure. In this paper, we try to use the idea of signal enhancement to enhance the ship radiated noise signal with extremely low signal-to-noise ratio, so as to achieve anti-explosive signal interference. We propose a signal enhancement deep learning model to enhance the ship radiated noise signal by learning a mask in the temporal domain. Our approach is an encoder–decoder structure with U-net. U-net consists of 1d-conv with skip connection. In order to improve the learning ability of the model, we directly connect the U-net in series. In order to improve the learning ability of the model's time series information. The Transformer attention mechanism is adopted to make the model have the ability to learn temporal information. We propose a combine Loss function for scale-invariant source-to-noise ratio and mean squared error in time-domain. Finally, we use the actual collected data to conduct experiments. It is verified that our algorithm can effectively improve the signal-to-noise ratio of the ship radiated noise signal to 2 dB under the extremely low signal-to-noise ratio of -20 dB to -25 dB.

Keywords: Ship radiated noise signal, Anti-interference, Time domain, Signal enhancement, Deep learning, 1D-convolutional neural network

1 Introduction

Ship radiated noise is one of the important ways of detecting and identifying ships. In order to prevent detection by the enemy's perception system, it will take the interference signal for the purpose of confrontation to cover its own radiated noise signal. The explosion signal is a typical interference signal. The explosion signal has the characteristics of high energy and wide coverage frequency. If an effective anti-interference method can be studied, it will greatly improve the identification efficiency of enemy ships in wartime, improving the efficiency and success rate of tactical operations. This paper will mainly study the ship radiation anti-interference method under the interference of explosion signal.

The ship radiated noise signal is regarded as the target signal, and the interference signal is regarded as noise. The existence of the interference signal will make the ship radiated noise signal with a very low signal-to-noise ratio. To resist the interference of

interfering signals, a very heuristic idea is that if we can effectively enhance the target signal, we can eliminate the influence of interfering signals, and finally achieve anti-interference. Therefore, the problem of anti-interference of ship radiated noise signal can be transformed into a signal enhancement problem. This paper will focus on the signal enhancement research under the low signal-to-noise ratio of single-channel ship radiated noise signal.

The traditional signal enhancement methods mainly include: spectral subtraction, Wiener filtering, mask method and so on [1–3]. According to the data in our hands and after testing and simulation, we found that the traditional signal enhancement method has a very poor effect on the enhancement of ship radiated noise signal in the case of extremely low signal-to-noise ratio. The specific simulation results are shown in the experiment. This paper will focus on signal enhancement methods based on data-driven deep learning methods. Due to the dependence of deep learning on massive data, current signal enhancement methods focus on the field of speech enhancement research [4–13]. Among them, the single-channel time domain signal enhancement method based on deep learning is to establish the mapping relationship between the enhanced signal and the mixture signal through the deep learning model, directly realize the signal enhancement in the time domain [14–23]. The method of using the time domain has certain advantages, avoiding the calculation of the mutual conversion of the time domain to the frequency domain. Moreover, through training, the model can have the ability to extract better signal features of the signal, and the features obtained through data-driven training may be more suitable for the current signal enhancement task. And the current time-frequency domain representation of the signal has the requirement of time resolution. The time domain method can avoid this problem. In theory, the signal enhancement method in the time domain can use any resolution. Recently, end-to-end signal enhancement in time-domain has shown state-of-the-art results in a variety of signal enhancement tasks [24–32]. WaveNet is used to deal with the text-to-speech problem through the CNN, which is also applied to time-domain signal enhancement [33, 34]. Conv-TasNet verifies and proposes the superiority of signal enhancement in the time domain [35]. Dual-Path RNN can build a bidirectional RNN deep learning model. Realize the modeling and learning of long-sequence signals [36]. In order to improve the learning ability of the model and effectively obtain sequence and local signal features, WaveCRN establishes a flexible signal enhancement strategy by combining CNN and LSTM [37]. In addition to adopting different deep learning paradigms, there are some other deep Learning techniques are applied to deal with the time-domain signal enhancement problem. For example, various attention mechanisms are combined in the model, such as self-attention, transformer, etc [38, 39].

In this paper, we propose a deep learning model for signal enhancement in the time domain, which is mainly aimed at the enhancement of ship radiated noise signal. Our method is inspired by Conv-TasNet. Our method is an encoder and decoder-based architecture with U-net. The U-net is established through 1d-conv, and combined with the transformer attention mechanism. In this paper, our enhanced target is ship radiated noise signal, and the interference signal is mainly explosion signal. The final experiments show that our method works well. The contributions of this paper mainly include two aspects. First, we propose the use of deep learning method for anti-interference of ship

radiated noise signal for the first time, which verifies that it is feasible to use deep learning paradigm for signal enhancement of ship radiated noise signal. Secondly, our experiment sets the signal-to-noise ratio at extremely low level, our final experimental results show that the ship radiation noise signal can still be effectively enhanced.

The rest of this paper is organized as follows: Sect. 2 defines the problem of signal enhancement of ship radiated noise. Section 3 proposed method and related technical details. Section 4 outlines experimental results and analysis. Section 5 represent Conclusion.

2 Problem definition

In this paper, we mainly focus on ship radiated noise signal enhancement, its application scenarios are different from general speech enhancement. Here, we need to clarify our main problems. The ship radiated noise signal is s and the interference signal is n , then the interfered signal can be expressed as:

$$y = s + n \quad (1)$$

where $\{y, s, n\} \in R^{l*m}$, m is the number of sampling points of the signal. We hope to use the signal enhancement method to enhance the received signal with a very low signal-to-noise ratio to reduce the influence of interference and obtain a signal \hat{y} with a higher signal-to-noise ratio from y .

$$\hat{y} = f_{\theta}(y) \quad (2)$$

where f_{θ} is the parameters θ of the deep learning model.

3 Proposed method

The method proposed in this paper is inspired by Conv-TasNet, and adopts an end-to-end signal enhancement strategy to build a 1dconv-Unet model based on the encoder decoder structure. Our idea is to frame and reduce the input signal through the input block, and then train the 1dconv-Unet model to obtain the mask of the signal [40], and multiply the mask by the time-domain signal vector to obtain the enhanced signal. After that, the signal is restored to the signal with the same time resolution as the input through the output block. The structure of the entire model is shown in Fig. 1. We will elaborate on it below.

3.1 Input block

The input signal $x \in R^{1*m}$, m is the length of the signal. Generally, the input signal is chunked into overlapping frames which is processed as frame-level enhancement: $s \in R^{T*L}$. Where T is the number of frames. L is the length of frames. After a 1d-conv layer, the frames are transformed into a vector representation of C channels. After passing through a nonlinear activation function:

$$I = H(\text{Conv1}(\text{seg_with_overlap}(s))) \in R^{T*C*L} \quad (3)$$

where $H()$ is nonlinear activation function, here we use the $PReLU()$ function.

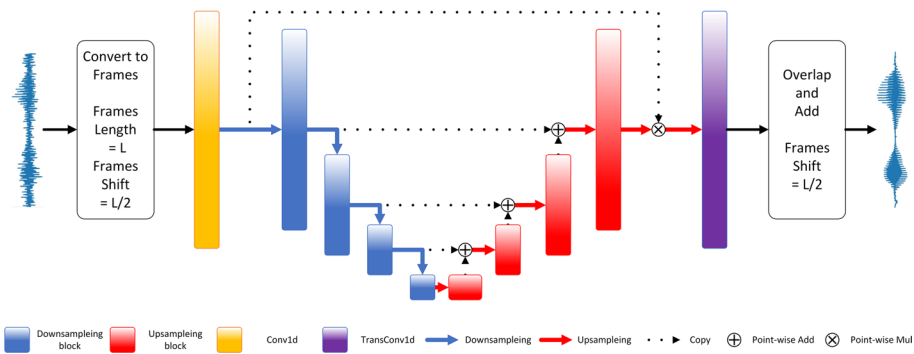


Fig. 1 Model

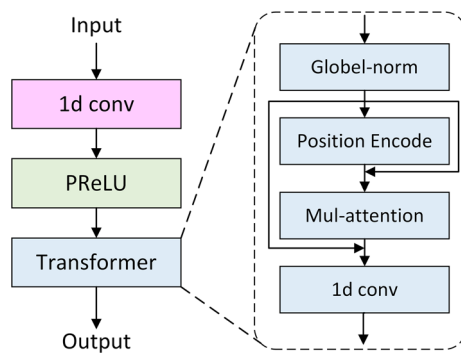


Fig. 2 Downsample

3.2 1d-Conv Unet block

This block is mainly composed of two modules: downsample block and upsample block. The multi-layer stacking of the downsample block realizes the dimension reduction transformation of the input vector. And upsample block realizes the up-dimension transformation of vector, and there will be skip connection in the conversion process of upsample. The interaction of information in different transformation processes at the same resolution is beneficial to improve model performance.

The main components of the Downsample block are shown in the Fig. 2. Among them, 1d-conv is mainly responsible for dimensionality reduction operations at different resolutions. In many other signal enhancement articles, in order to make the model have better temporal information learning ability, it is mostly through some deep learning paradigms such as LSTM and RNN. However, such the typical time series deep learning model will greatly increase the complexity and computational complexity of the model. Therefore, we choose the transformer as the attention mechanism here, so that the model has the ability to learn time series information [41].

In actual training, we found that the use of globe layer norm can greatly improve the performance of the model.

$$gLN(F) = \frac{F - E[F]}{\sqrt{\text{Var}[F] + \varepsilon}} \odot \gamma + \beta \tag{4}$$

$$E[F] = \frac{1}{NT} \sum_{NT} F \tag{5}$$

$$\text{Var}[F] = \frac{1}{NT} \sum_{NT} (F - E[F])^2 \tag{6}$$

where $F \in R^{C*T}$ is the vector of input, γ, β is learnable parameters.

The main components of the upsample block are shown in the Fig. 3. Among them, 1d-conv is mainly responsible for expanding the input vector at the channel level. It is then dilated at the temporal resolution by Pixel shuffle. Pixel shuffle is mainly used in the field of image super-resolution research [42]. It is usually applied to 2-dimensional data. Here, we apply it to a 1-dimensional signal. The input vector $v \in R^{T*C*K}$. For the r upsample rate, we first expand the vector at the channel level through 1d-conv $v' \in R^{T*rC*K}$, and then expand at the temporal resolution level $o \in R^{T*C*rK}$. The pixel shuffle operation process is shown in the Fig. 3.

In order to improve the learning ability of the model at different time resolutions. Here, we will use multiple Unet networks in series, and use different dilated 1d-conv convolutions. In actual training, it is found that setting different dilated parameters can significantly improve the performance of the model. We believe that under the condition that multiple Unet networks are connected in series, using different dilated parameters can extract time series information at different time resolutions. Greatly improves the performance of the model.

3.3 Output block

The Output block is to ensure that the time resolution of the signal is restored to the length of the original signal. The input $v \in R^{T*C*K}$ is converted to by 1d-transconv $m \in R^{T*L}$, and finally converted to the same size as the input signal by the overlap-and-add method.

$$O = \text{overlap_and_add}(\text{Conv1}(s)) \in R^{1*m} \tag{7}$$

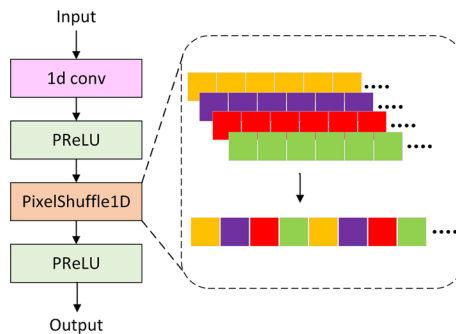


Fig. 3 Upsample

3.4 Loss function

The model proposed in this paper adopts an end-to-end training strategy, and the target of model training is the scale-invariant source-to-noise ratio. This metric is often used to evaluate the performance of signal enhancement methods. It is defined as:

$$s_{\text{target}} = \frac{\langle \hat{s}, s \rangle s}{\|s\|^2} \quad (8)$$

$$e_{\text{noise}} = \hat{s} - s_{\text{target}} \quad (9)$$

$$L_S = \text{SI-SNR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2} \quad (10)$$

where \hat{s} and s are the estimated and original clean sources. $\langle \cdot \rangle$ is inner product operation. A mean squared error (MSE) loss in the time domain is defined as:

$$L_T(s, \hat{s}) = \frac{1}{N} \sum_{k=0}^{N-1} (s[k] - \hat{s}[k])^2 \quad (11)$$

where \hat{s} and s are same as (8). Finally, We combine the two loss functions:

$$L = \alpha L_T + (1 - \alpha) L_S \quad (12)$$

where α is a hyperparameter. The combination loss function we designed is divided into two parts, the first part is that SI-SNR is our task objective. We hope that the SI-SNR of the enhanced signal will be significantly improved. The second part is to want the signal output by our model to be as consistent as possible in the time domain. We found that the combination loss function can improve the effect of signal enhancement.

4 Experiment

4.1 Data

Our ship radiated noise signal data comes from DeepShip [43]. Our interference signal data were collected in the South China Sea. The explosion signals are simulated by periodic air guns ranging from seconds to minutes for scientific research. The sampling locations are located in two different sea areas, with a total length of 8 hours. During this period, there will be short-term long-distance ships passing through. Moreover, some of the noise signals have oil exploration platform operating noise. The waveform of the interference signal is shown in Fig. 4. When we use these airsoft signals. we split it into ten-second segment. When making the dataset, we clean the data to make sure that there is an airsoft signal in the segment, and then fuse it with the ship's radiated noise signal at -20 dB to -25 dB. Finally, it is fused with the radiation noise signal of the ship. The data set we finally formed has a total of 32,312 signals, and the sampling frequency is 8 KHz.

Two samples in our dataset are shown in the Fig. 4. The raw_sig in the figure is the ship radiated noise signal, the noi_sig is the air gun signal, and the mix_sig is the mixed signal, and the legend on the mixed signal shows the signal-to-noise ratio of the mixed signal.

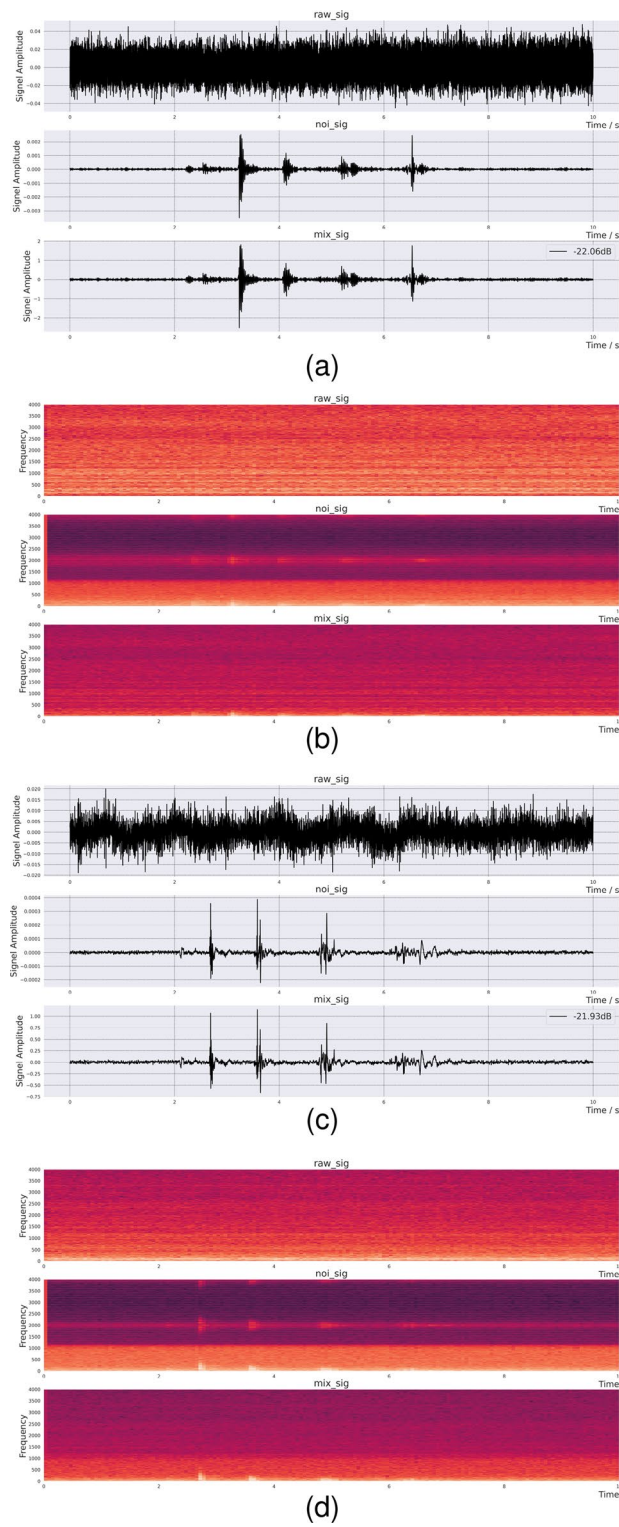


Fig. 4 Two samples of the dataset

(a) and (c) are the time-domain waveforms of the two samples, we can find that the two samples are seriously distorted after mixing with the noise signal. (b)(d) are the visualization images of the two samples after time-frequency transformation. We can find that the frequency range of ship radiated noise signal is mainly distributed 0–1200 Hz. Our airsoft noise signaling industry is mainly distributed in this frequency band too. After mixing, it can be seen that the ship's radiated noise signal has been submerged in the noise, which is very difficult to identify.

4.2 Experiment setting

Our model uses 10 layers of Unet-1dconv blocks, 5 layers of downsample blocks and 5 layers of upsample blocks. Cascading 5 Unet-1dconv blocks. The default batch is 1 during model training. The optimization algorithm is SGD, and the learning rate is 0.0001. We set 50 epochs, and if the loss function 5 epochs did not drop, the training would stop. Our dataset has a total of 32,312 signals. We divide it into three parts, 50% is used as training set, 20% is used as validation set, and 30% is used as test set. The original length of the data is 80,000, and the length of each frame through slices is 4000 with the 50% overlap. We use scale-invariant source-to-noise ratio and source-to-noise ratio (SNR) as objective evaluation. Our experimental comparison methods are: Ideal Amplitude Mask (IAM), Ideal Ratio Mask (IRM), Ideal Binary Mask (IBM), Winner-Filter, Conv-TasNet. We will conduct multiple experiments and show the best results. We deploy our model using Pytorch. Using two NVIDIA 2080TI for training. The definition of SNR is as follows:

$$\text{SNR (dB)} = 10\log_{10} \frac{|\hat{s}|^2}{|s_{\text{reference}} - \hat{s}|^2} \quad (13)$$

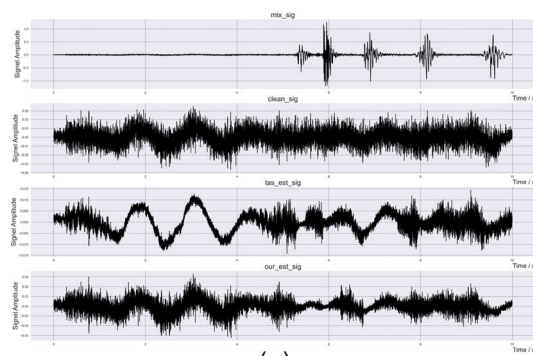
where \hat{s} and $s_{\text{reference}}$ are estimated and reference signals

4.3 Experiment results

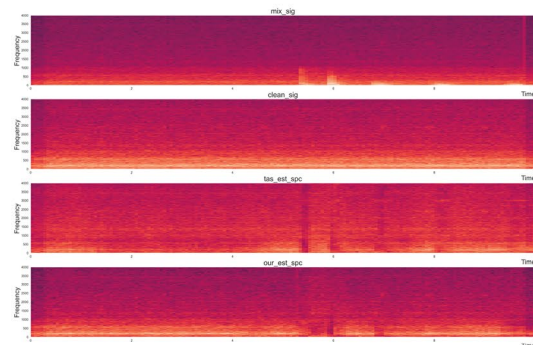
The objective evaluation results of our experiments are shown in the Table 1. We use SI-SNR and SNR for verification. We find that traditional methods such as IAM, IRM, IBM, Winner-Filter do not perform well in the task of ship radiated noise signal enhancement. And it can be found that the two evaluation are quite different. We believe that SI-SNR reduces the influence of signal strength on signal-to-noise through orthogonality. Part of the frequency components can be filtered out by winner filtering, but it has a lower signal gain, so it has the high SI-SNR and the low SNR. And vice versa for other mask

Table 1 Objective evaluation result

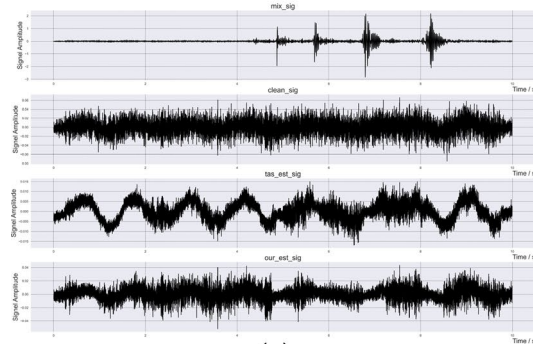
Modle	SI-SNR	SNR
IAM	− 55.2528	0.0639
IRM	− 98.7227	0.0058
IBM	− 56.6350	0.0737
Winner filter	− 0.0061	− 48.2690
Conv-TasNet	1.7453	1.2535
Ours	2.3459	2.9996



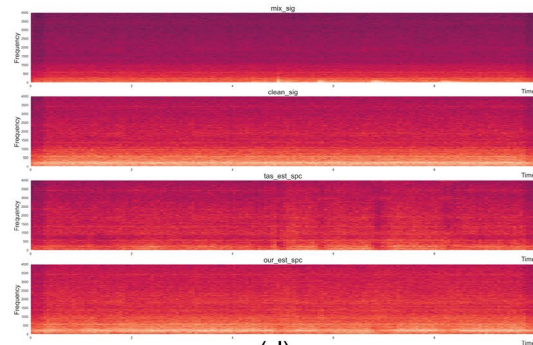
(a)



(b)



(c)



(d)

Fig. 5 The output of two models

methods. We can see that the data-driven deep learning method has crushing advantages in two objective evaluation. Our method outperforms the state-of-art Conv-TasNet on all of the objective evaluation.

Next we compare through the visualization of the signal. Since other traditional methods differ too much in objective evaluation criteria, we only compare Conv-TasNet. We visualize the signal after the enhancement of the two signal enhancement models. The signal time-domain waveform and time-frequency spectrum are drawn respectively. Its visualization results are shown in the Fig. 5. Where “mix_sig” is the input of the model and the signal to be enhanced. “clean_sig” is the reference signal, “tas_est_sig” is the output of Conv-TasNet, and “our_est_sig” is the output of our method. By comparing (a) and (c), we find that our enhancement effect in the time domain is better than that of Conv-TasNet. By comparing (b) (d), we find that the enhancement effect is better in the low frequency band of 0–1000 Hz, and the line spectrum is clearer.

5 Conclusion and discussion

We propose a data-driven deep learning method for ship radiated noise signal enhancement. We directly enhance the signal in the time domain, eliminating the need for conversion between different domains. We build the encoder–decoder structure of the Unet network at different temporal resolutions. We introduce a transformer attention mechanism to enable our model to learn temporal information. We conduct experiments with actual collected data and verify that our method can effectively enhance the signal time in the case of extremely low signal-to-noise ratios of -20 to -25 dB. In the future, we would like to propose suitable anti-reference methods for ship radiated noise signal for more interference types.

Acknowledgements

This research was supported by National Natural Science Foundation of China (Key projects) (Grant No. 62031021). This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Author contributions

YD: Methodology, Software, Investigation, Funding Acquisition, Formal Analysis, Funding Acquisition, Writing - Original Draft. XS: Conceptualization, Supervision. HW: Supervision, Funding Acquisition. All authors read and approved the final manuscript.

Author's information

Duan Yichen, pd.D candidate in Northwestern Polytechnical University. His research interests include signal processing, deep learning, and fault detection.

Declarations

Competing interests

The authors declare that there is no conflict of interests.

Received: 23 April 2022 Accepted: 4 July 2022

Published online: 23 July 2022

References

1. V. Botchev, Speech enhancement: theory and practice (2nd ed). *Comput. Rev.* **54**(10), 604–605 (2013)
2. M. Tayseer, A. Adeel, A. Hussain, A survey on techniques for enhancing speech. *Int. J. Comput. Appl.* **179**(17), 1–14 (2018)
3. D.L. Wang, J. Chen, Supervised speech separation based on deep learning: an overview. *IEEE/ACM Trans. Audio Speech Lang. Process.* **PP**(99), 1 (2017)
4. L. Chai, J. Du, Q.F. Liu et al., Using generalized gaussian distributions to improve regression error modeling for deep learning-based speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **PP**(99), 1 (2019)

5. C. Chermaz, D. Leuchtman, S. Tanner, et al., Compressed representation of cepstral coefficients via recurrent neural networks for informed speech enhancement, in *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2021)
6. C. Chermaz, D. Leuchtman, S. Tanner, et al., Compressed representation of cepstral coefficients via recurrent neural networks for informed speech enhancement, in *ICASSP 2021—2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE (2021)
7. M. Sadeghi, X. Alameda-Pineda, Mixture of inference networks for VAE-based audio-visual speech enhancement. *IEEE Trans. Signal Process.* **PP**(99), 1–1 (2021)
8. K. Zhao, Y. Yang, Y. Wang, et al., The method of disentangled and interpretable representations for speech enhancement, in *2021 IEEE 5th advanced information technology, electronic and automation control conference (IAEAC)*. IEEE (2021)
9. L. Zhou, Q. Zhong, T. Wang et al., Speech enhancement via residual dense generative adversarial network. *Comput. Syst. Sci. Eng.* **38**(3), 279–289 (2021)
10. A. Pandey, D.L. Wang, Exploring deep complex networks for complex spectrogram enhancement, in *ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2019)
11. S. Chowdhury, A. Chatterjee, Speech enhancement using k-sparse autoencoder techniques, in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (2021)
12. W. Zhou, M. Lu, R. Ji, Meta-SE: a meta-learning framework for few-shot speech enhancement. *IEEE Access* **PP**(99), 1 (2021)
13. M. Hasannezhad, W.P. Zhu, B. Champagne, A novel low-complexity attention-driven composite model for speech enhancement, in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE (2021)
14. J. Wu, H. Liu, L. Gan, et al., A Single Channel End-to-End Speech Enhancement Using Complex Operations (2021)
15. K. Wang, B. He, W.P. Zhu, CAUNet: context-aware U-Net for speech enhancement in time domain, In *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE (2021)
16. C. Chermaz, D. Leuchtman, S. Tanner, et al., Compressed representation of cepstral coefficients via recurrent neural networks for informed speech enhancement, in *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2021)
17. D. Ditter, T. Gerkman, A multi-phase gammatone filterbank for speech separation via Tasnet, in *ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2020)
18. X. Xiang, X. Zhang, H. Chen, A convolutional network with multi-scale and attention mechanisms for end-to-end single-channel speech enhancement. *IEEE Signal Process. Lett.* **28**, 1455–1459 (2021)
19. H. Kim, J.W. Shin, Target exaggeration for deep learning-based speech enhancement. *Digit. Signal Process.* **116**, 103109 (2021)
20. W.H. Heo, H. Kim, O.W. Kwon, Integrating Dilated Convolution into DenseLSTM for Audio Source Separation (2021)
21. C. Fan, J. Tao, B. Liu et al., End-to-end post-filter for speech separation with deep attention fusion features. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 1303–1314 (2020)
22. Y. Li, Y. Sun, S.M. Naqvi, Single-channel dereverberation and denoising based on lower band trained SA-LSTMs (2021)
23. C. Xu, W. Rao, E.S. Chng et al., SpEx: multi-scale time domain speaker extraction network. *IEEE/ACM Trans. Audio Speech Lang. Process.* **PP**(99), 1 (2020)
24. M. Kolbk, Z.H. Tan, S.H. Jensen et al., On loss functions for supervised monaural time-domain speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 825–838 (2020)
25. M. Kolbk, Z.H. Tan, S.H. Jensen et al., On loss functions for supervised monaural time-domain speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 825–838 (2020)
26. L. Chao, T. Jiang, W. Sheng, Single-channel speech enhancement based on adaptive low-rank matrix decomposition. *IEEE Access* **PP**(99), 37066–37076 (2020)
27. B. Sfa, M. Djendi, Hybrid PSO-NLMS (HPSO-NLMS) algorithm for blind speech quality enhancement in time domain. *Appl. Acoust.* **177**, 107936 (2021)
28. Z.Q. Wang, G. Wichern, J.L. Roux, On the compensation between magnitude and phase in speech separation. *IEEE Signal Process. Lett.* **28**, 2018–2022 (2021)
29. C. Yu, Y.T. Lin, K.H. Hung, et al., Time-domain multi-modal bone/air conducted speech enhancement (2019)
30. X. Xiang, X. Zhang, H. Chen, Two-stage learning and fusion network with noise aware for time-domain monaural speech enhancement. *Signal Process. Lett. IEEE* **PP**(99), 1 (2021)
31. J.R. Jensen, S. Karimian-Azari, M.G. Christensen, J. Benesty, Harmonic beamformers for speech enhancement and dereverberation in the time domain. *Speech Commun.* **116**, 1–11 (2020)
32. D. Baby, S. Verhulst, End-to-end raw speech waveform enhancement using conditional generative adversarial networks (2018)
33. A. Oord, S. Dieleman, H. Zen, et al., WaveNet: A Generative Model for Raw Audio (2016)
34. D. Rethage*, J. Pons*, X. Serra, A wavenet for speech denoising, in *Icassp IEEE International Conference on Acoustics*. IEEE (2018)
35. Y. Luo, N. Mesgarani, Conv-TasNet: surpassing ideal time-frequency magnitude masking for speech separation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **PP**(99), 1 (2019)
36. Y. Luo, Z. Chen, T. Yoshioka, Dual-path RNN: efficient long sequence modeling for time-domain single-channel speech separation, in *ICASSP 2020—2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE (2020)
37. T.A. Hsieh, H.M. Wang, X. Lu, et al., WaveCRN: an efficient convolutional recurrent neural network for end-to-end speech enhancement (2020)
38. A. Pandey, D.L. Wang, Dense CNN with self-attention for time-domain speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **29**, 1270–1279 (2020)
39. W. Yu, J. Zhou, H.B. Wang et al., SETransformer: speech enhancement transformer. *Cogn. Comput.* **14**, 1152–1158 (2021)
40. O. Ronneberger, P. Fischer, T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Springer, Berlin, 2015)
41. A. Vaswani, et al., Attention is all you need, in *Advances in Neural Information Processing Systems* (2017)

42. W. Shi, J. Caballero, F. Huszár, et al., Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in *IEEE* (2016)
43. M. Irfan, J. Zheng, S. Ali et al., DeepShip: an underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Syst. Appl.* **183**(5), 115270 (2021)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
