Scopus

Documents

Saleem, N.[a b] , Gunawan, T.S.[b c] , Dhahbi, S.[d] , Bourouis, S.[e]

Time domain speech enhancement with CNN and time-attention transformer
(2024) *Digital Signal Processing: A Review Journal*, 147, art. no. 104408, .

DOI: 10.1016/j.dsp.2024.104408

[a] Department of Electrical Engineering, Faculty of Engineering and Technology, Gomal University, D.I.Khan, 29050, Pakistan
[b] Electrical and Computer Engineering Department, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia
[c] Department of Electrical Engineering, Chulalongkorn University, Bangkok, 10330, Thailand
[d] Department of Computer Science, College of Science and Art at Mahayil, King Khalid University, Muhayil Aseer, 62529, Saudi Arabia
[e] Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia

Abstract
Speech enhancement in the time domain involves improving the quality and intelligibility of noisy speech by processing the waveform directly without the need for explicit feature extraction or domain transformation. Deep learning is a powerful approach for time domain speech enhancement, offering significant improvements over traditional techniques. Formulating a resource-efficient deep neural model in the time domain without ignoring the contextual information and detailed features of input speech is still a vital challenge. To address this challenge, this study proposes a speech enhancement model using 1D-time domain dilated residual blocks in the convolutional encoder-decoder framework. Further, this study integrates a time-attention transformer (TAT) bottleneck between the encoder-decoder. The TAT model extends the transformer architecture by incorporating a time-attention mechanism, which enables the model to selectively attend to different segments of the speech signal over time. This allows the model to effectively capture long-term dependencies in the speech and learn to recognize important features. The experimental results indicate that the proposed speech enhancement outperforms the recent deep neural networks (DNNs) and substantially improves the intelligibility and quality of noisy speech. With the WSJ0 SI-84 database, the proposed SE improves the STOI and PESQ by 21.51% and 1.14 over noisy speech. © 2024

Author Keywords
Convolutional encoder-decoder; Time attention; Time-domain speech enhancement; Transformer

Index Keywords
Convolution, Decoding, Deep neural networks, Signal encoding, Speech intelligibility, Time domain analysis; Convolutional encoder-decoder, Convolutional encoders, Encoder-decoder, Features extraction, Noisy speech, Time attention, Time domain, Time-domain speech enhancement, Transformer, Waveforms; Speech enhancement

References

- Gupta, M., Singh, R.K., Singh, S.
Analysis of optimized spectral subtraction method for single channel speech enhancement
(2023) *Wirel. Pers. Commun.*, 128 (3), pp. 2203-2215.
- Chen, J., Benesty, J., Huang, Y., Doclo, S.
New insights into the noise reduction Wiener filter
(2006) *IEEE Trans. Audio Speech Lang. Process.*, 14 (4), pp. 1218-1234.
- Saleem, N., Khattak, M.I., Nawaz, A., Umer, F., Ochani, M.K.
Perceptually weighted β-order spectral amplitude Bayesian estimator for phase compensated speech enhancement
(2021) *Appl. Acoust.*, 178.
- Wang, D., Chen, J.
Supervised speech separation based on deep learning: an overview
(2018) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 26 (10), pp. 1702-1726.

- Jiang, W., Yu, K.
  Speech enhancement with integration of neural homomorphic synthesis and spectral masking
  (2023) *IEEE/ACM Trans. Audio Speech Lang. Process.*,

- Li, Y., Sun, M., Zhang, X.
  Perception-guided generative adversarial network for end-to-end speech enhancement
  (2022) *Appl. Soft Comput.*, 128.

- Saleem, N., Khattak, M.I., Al-Hasan, M., Qazi, A.B.
  On learning spectral masking for single channel speech enhancement using feedforward and recurrent neural networks
  (2020) *IEEE Access*, 8, pp. 160581-160595.

- Khattak, M.I., Saleem, N., Gao, J., Verdu, E., Fuente, J.P.
  Regularized sparse features for noisy speech enhancement using deep neural networks
  (2022) *Comput. Electr. Eng.*, 100.

- Qiu, Y., Wang, R., Hou, F., Singh, S., Ma, Z., Jia, X.
  Adversarial multi-task learning with inverse mapping for speech enhancement
  (2022) *Appl. Soft Comput.*, 120.

- Wang, Z.Q., Wang, P., Wang, D.
  Complex spectral mapping for single- and multi-channel speech enhancement and robust ASR
  (2020) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 28, pp. 1778-1787.

- Tan, K., Wang, D.
  Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement
  (2019) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 28, pp. 380-390.

- Li, A., Zheng, C., Zhang, L., Li, X.
  Glance and gaze: a collaborative learning framework for single-channel speech enhancement
  (2022) *Appl. Acoust.*, 187.

- Wang, T., Pan, Z., Ge, M., Yang, Z., Li, H.
  Time-domain speech separation networks with graph encoding auxiliary
  (2023) *IEEE Signal Process. Lett.*, 30, pp. 110-114.

- Kolbæk, M., Tan, Z.H., Jensen, S.H., Jensen, J.
  On loss functions for supervised monaural time-domain speech enhancement
  (2020) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 28, pp. 825-838.

- Pascual, S., Serra, J., Bonafonte, A.
  Time-domain speech enhancement using generative adversarial networks
  (2019) *Speech Commun.*, 114, pp. 10-21.

- Yu, C., Hung, K.H., Wang, S.S., Tsao, Y., Hung, J.W.
  Time-domain multi-modal bone/air conducted speech enhancement
  (2020) *IEEE Signal Process. Lett.*, 27, pp. 1035-1039.

- Mowlaee, P., Kulmer, J.
  Phase estimation in single-channel speech enhancement: limits-potential
  (2015) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 23 (8), pp. 1283-1294.

- Yu, R., Chen, W., Ye, Z.
  A novel target decoupling framework based on waveform-spectrum fusion network

for monaural speech enhancement
(2023) *Digit. Signal Process.*, 141.

- Dang, F., Chen, H., Hu, Q., Zhang, P., Yan, Y.
First coarse, fine afterward: a lightweight two-stage complex approach for monaural speech enhancement
(2023) *Speech Commun.*, 146, pp. 32-44.

- Saleem, N., Khattak, M.I.
Multi-scale decomposition based supervised single channel deep speech enhancement
(2020) *Appl. Soft Comput.*, 95.

- Lee, J., Kang, H.G.
Real-time neural speech enhancement based on temporal refinement network and channel-wise gating methods
(2023) *Digit. Signal Process.*, 133.

- Yu, G., Li, A., Wang, H., Wang, Y., Ke, Y., Zheng, C.
DBT-net: dual-branch federative magnitude and phase estimation with attention-in-attention transformer for monaural speech enhancement
(2022) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 30, pp. 2629-2644.

- Hasannezhad, M., Ouyang, Z., Zhu, W.P., Champagne, B.
An integrated CNN-GRU framework for complex ratio mask estimation in speech enhancement
(2020) *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 764-768.
IEEE

- Luo, Y., Mesgarani, N.
Conv-tasnet: surpassing ideal time–frequency magnitude masking for speech separation
(2019) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 27 (8), pp. 1256-1266.

- Hsieh, T.A., Wang, H.M., Lu, X., Tsao, Y.
Wavecrn: an efficient convolutional recurrent neural network for end-to-end speech enhancement
(2020) *IEEE Signal Process. Lett.*, 27, pp. 2149-2153.

- Sahu, S.K., Mokhade, A., Bokde, N.D.
An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: recent progress and challenges
(2023) *Appl. Sci.*, 13 (3), p. 1956.

- Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M., Zhong, J.
Attention is all you need in speech separation
(2021) *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 21-25.
IEEE

- Evrard, M.
Transformers in automatic speech recognition
(2023) *Human-Centered Artificial Intelligence: Advanced Lectures*, pp. 123-139.
Springer International Publishing Cham

- Almadhor, A., Irfan, R., Gao, J., Saleem, N., Rauf, H.T., Kadry, S.
E2E-DASR: end-to-end deep learning-based dysarthric automatic speech recognition
(2023) *Expert Syst. Appl.*, 222.

- Guo, H., Jian, H., Wang, Y., Wang, H., Zhao, X., Zhu, W., Cheng, Q.
MAMGAN: multiscale attention metric GAN for monaural speech enhancement in the time domain
(2023) *Appl. Acoust.*, 209.

- Yu, W., Zhou, J., Wang, H., Tao, L.
SETransformer: speech enhancement transformer
(2022) *Cogn. Comput.*, pp. 1-7.

- Li, Y., Sun, Y., Wang, W., Naqvi, S.M.
U-shaped transformer with frequency-band aware attention for speech enhancement
(2023) *IEEE/ACM Trans. Audio Speech Lang. Process.*,

- Kim, J., El-Khamy, M., Lee, J.
T-gsa: transformer with Gaussian-weighted self-attention for speech enhancement
(2020) *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6649-6653.
IEEE

- Lin, J., van Wijngaarden, A.J.D.L., Wang, K.C., Smith, M.C.
Speech enhancement using multi-stage self-attentive temporal convolutional networks
(2021) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 29, pp. 3440-3450.

- Pandey, A., Wang, D.
Dense CNN with self-attention for time-domain speech enhancement
(2021) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 29, pp. 1270-1279.

- O'Malley, T., Ding, S., Narayanan, A., Wang, Q., Rikhye, R., Liang, Q.
Conditional conformer: improving speaker modulation for single and multi-user speech enhancement
(2023) *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
IEEE

- Wang, M., Chen, J., Zhang, X., Huang, Z., Rahardja, S.
Multi-modal speech enhancement with bone-conducted speech in time domain
(2022) *Appl. Acoust.*, 200.

- Pandey, A., Wang, D.
Self-attending RNN for speech enhancement to improve cross-corpus generalization
(2022) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 30, pp. 1374-1385.

- Fan, W., Li, D., Lu, W., Tsao, Y.
Time domain attention convolutional neural network for speech denoising
(2019) *Proc. ICASSP*,

- Jin, Y., Tang, C., Liu, Q., Wang, Y.
Multi-head self-attention-based deep clustering for single-channel speech separation
(2020) *IEEE Access*, 8, pp. 100013-100021.

- Li, L., Kang, Y., Shi, Y., Kürzinger, L., Watzel, T., Rigoll, G.
Adversarial joint training with self-attention mechanism for robust end-to-end speech recognition
(2021) *EURASIP J. Audio Speech Music Process.*, 2021, pp. 1-16.

- Pandey, A., Wang, D.
TCNN: temporal convolutional neural network for real-time speech enhancement in

the time domain
(2019) *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6875-6879.
IEEE

- Song, Z., Ma, Y., Tan, F., Feng, X.
Hybrid dilated and recursive recurrent convolution network for time-domain speech enhancement
(2022) *Appl. Sci.*, 12 (7), p. 3461.

- Wang, K., He, B., Zhu, W.P.
TSTNN: two-stage transformer based neural network for speech enhancement in the time domain
(2021) *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7098-7102.
IEEE

- Pandey, A., Wang, D.
Densely connected neural network with dilated convolutions for real-time speech enhancement in the time domain
(2020) *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6629-6633.
IEEE

- Chen, C., Hou, N., Ma, D., Chng, E.S.
Time domain speech enhancement with attentive multi-scale approach
(2021) *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 679-683.
IEEE

- Kishore, V., Tiwari, N., Paramasivam, P.
Improved speech enhancement using TCN with multiple encoder-decoder layers
(2020) *Interspeech*, pp. 4531-4535.

- Wang, K., He, B., Zhu, W.P.
Cptnn: cross-parallel transformer neural network for time-domain speech enhancement
(2022) *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1-5.
IEEE

- Kong, Z., Ping, W., Dantrey, A., Catanzaro, B.
Speech denoising in the waveform domain with self-attention
(2022) *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7867-7871.
IEEE

- Macartney, C., Weyde, T.
Improved speech enhancement with the wave-u-net
(2018), preprint

- Défossez, A., Synnaeve, G., Adi, Y.
Real time speech enhancement in the waveform domain
(2020) *Proc. Interspeech*, 2020, pp. 3291-3295.

- Pascual, S., Bonafonte, A., Serra, J.
SEGAN: speech enhancement generative adversarial network
(2017), preprint

- Pascual, S., Serra, J., Bonafonte, A.
Time-domain speech enhancement using generative adversarial networks
(2019) *Speech Commun.*, 114, pp. 10-21.

- Phan, H., McLoughlin, I.V., Pham, L., Chén, O.Y., Koch, P., De Vos, M., Mertins, A.
Improving GANs for speech enhancement
(2020) *IEEE Signal Process. Lett.*, 27, pp. 1700-1704.

- Phan, H., Le Nguyen, H., Chén, O.Y., Koch, P., Duong, N.Q., McLoughlin, I., Mertins, A.
Self-attention generative adversarial network for speech enhancement
(2021) *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7103-7107.
IEEE

- Li, L., Kürzinger, L., Watzel, T., Rigoll, G.
Lightweight end-to-end speech enhancement generative adversarial network using sinc convolutions
(2021) *Appl. Sci.*, 11 (16), p. 7564.

- Stoller, D., Ewert, S., Dixon, S.
Wave-u-net: a multi-scale neural network for end-to-end audio source separation
(2018), preprint

- Xiang, X., Zhang, X., Chen, H.
Two-stage learning and fusion network with noise aware for time-domain monaural speech enhancement
(2021) *IEEE Signal Process. Lett.*, 28, pp. 1754-1758.

- Paul, D.B., Baker, J.
The design for the wall street journal-based CSR corpus
(1992) *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992*,

- Saleem, N., Khattak, M.I., AlQahtani, S.A., Jan, A., Hussain, I., Khan, M.N., Dahshan, M.
U-shaped low-complexity type-2 fuzzy LSTM neural network for speech enhancement
(2023) *IEEE Access*, 11, pp. 20814-20826.

- Saleem, N., Khattak, M.I.
Deep neural networks for speech enhancement in complex-noisy environments
(2020) *Int. J. Interact. Multimed. Artif. Intell.*, 6 (1), pp. 84-91.

- Pandey, A., Wang, D.
A new framework for CNN-based speech enhancement in the time domain
(2019) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 27 (7), pp. 1179-1188.

- Taal, C.H., Hendriks, R.C., Heusdens, R., Jensen, J.
A short-time objective intelligibility measure for time-frequency weighted noisy speech
(2010) *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4214-4217.
IEEE

- Beerends, J.G., Hekstra, A.P., Rix, A.W., Hollier, M.P.
Perceptual evaluation of speech quality (pesq) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model
(2002) *J. Audio Eng. Soc.*, 50 (10), pp. 765-778.

- Hu, Y., Loizou, P.C.
Evaluation of objective quality measures for speech enhancement
(2007) *IEEE Trans. Audio Speech Lang. Process.*, 16 (1), pp. 229-238.

- Chen, J., Wang, D.
Long short-term memory for speaker generalization in supervised speech

separation
(2017) *J. Acoust. Soc. Am.*, 141 (6), pp. 4705-4714.

- Tan, K., Chen, J., Wang, D.
Gated residual networks with dilated convolutions for monaural speech enhancement
(2018) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 27 (1), pp. 189-198.

- Giri, R., Isik, U., Krishnaswamy, A.
Attention wave-u-net for speech enhancement
(2019) *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 249-253.
IEEE

- Kim, E., Seo, H.
SE-conformer: time-domain speech enhancement using conformer
(2021) *Interspeech*, pp. 2736-2740.

- Fan, C., Yi, J., Tao, J., Tian, Z., Liu, B., Wen, Z.
Gated recurrent fusion with joint training framework for robust end-to-end speech recognition
(2020) *IEEE/ACM Trans. Audio Speech Lang. Process.*, 29, pp. 198-209.

- Zadorozhnyy, V., Ye, Q., Koishida, K.
SCP-GAN: self-correcting discriminator optimization for training consistency preserving metric GAN on speech enhancement tasks
(2022), preprint

- Baby, D., Verhulst, S.
Sergan: speech enhancement using relativistic generative adversarial networks with gradient penalty
(2019) *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 106-110.
IEEE

- Yu, G., Li, A., Zheng, C., Guo, Y., Wang, Y., Wang, H.
Dual-branch attention-in-attention transformer for single-channel speech enhancement
(2022) *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7847-7851.
IEEE

- Abdulatif, S., Cao, R., Yang, B.
Cmgan: conformer-based metric-gan for monaural speech enhancement
(2022), preprint

- Qiu, Z., Fu, M., Yu, Y., Yin, L., Sun, F., Huang, H.
Srtnet: time domain speech enhancement via stochastic refinement
(2023) *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
IEEE

- Shi, H., Mimura, M., Wang, L., Dang, J., Kawahara, T.
Time-domain speech enhancement assisted by multi-resolution frequency encoder and decoder
(2023) *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5.
IEEE

- Dang, F., Hu, Q., Zhang, P., Yan, Y.
ForkNet: simultaneous time and time-frequency domain modeling for speech

enhancement
(2023), preprint

- Lu, Y.X., Ai, Y., Ling, Z.H.
MP-SENet: a speech enhancement model with parallel denoising of magnitude and phase spectra
(2023), preprint

- Saleem, N., Gunawan, T.S., Shafi, M., Bourouis, S., Trigui, A.
Multi-attention bottleneck for gated convolutional encoder-decoder-based speech enhancement
(2023) *IEEE Access*,

- Saleem, N., Gunawan, T.S., Kartiwi, M., Nugroho, B.S., Wijayanto, I.
NSE-CATNet: deep neural speech enhancement using convolutional attention transformer network
(2023) *IEEE Access*,

- Hou, Z., Hu, Q., Chen, K., Lu, J.
Local spectral attention for full-band speech enhancement
(2023), preprint

- Nicolson, A., Paliwal, K.K.
Masked multi-head self-attention for causal speech enhancement
(2020) *Speech Commun.*, 125, pp. 80-96.

- Kadri, R., Bouaziz, B., Tmar, M., Gargouri, F.
Efficient multimodel method based on transformers and CoAtNet for Alzheimer's diagnosis
(2023) *Digit. Signal Process.*, 143.

- Zadorozhnyy, V., Ye, Q., Koishida, K.
SCP-GAN: self-correcting discriminator optimization for training consistency preserving metric GAN on speech enhancement tasks
(2022), preprint

Correspondence Address
Saleem N.; Department of Electrical Engineering, Pakistan; email: nasirsaleem@gu.edu.pk

Publisher: Elsevier Inc.

ISSN: 10512004
CODEN: DSPRE
Language of Original Document: English
Abbreviated Source Title: Digital Signal Process Rev J
2-s2.0-85183963155
Document Type: Article
Publication Stage: Final
Source: Scopus