REFERENCES

[1.] Mosleh, M., Setayeshi, S., Barekatain, B. and Mosleh, M., 2021. A novel audio watermarking scheme based on fuzzy inference system in DCT domain. Multimedia Tools and Applications, 80, pp.20423-20447.

[2.] Kim, W. and Lee, K., 2020, May. Digital watermarking for protecting audio classification datasets. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2842-2846). IEEE.

[3.] Pourhashemi, S.M., Mosleh, M. and Erfani, Y., 2021. A novel audio watermarking scheme using ensemble-based watermark detector and discrete wavelet transform. Neural Computing and Applications, 33, pp.6161-6181.

[4.] Masmoudi, S., Charfeddine, M. and Ben Amar, C., 2020. A semi-fragile digital audio watermarking scheme for MP3-encoded signals using Huffman data. Circuits, Systems, and Signal Processing, 39, pp.3019-3034.

[5.] Karajeh, H., Khatib, T., Rajab, L. and Maqableh, M., 2019. A robust digital audio watermarking scheme based on DWT and Schur decomposition. Multimedia Tools and Applications, 78, pp.18395-18418.

[6.] Olanrewaju, R.F. and Khalifa, O., 2012, July. Digital audio watermarking; techniques and applications. In 2012 International Conference on Computer and Communication Engineering (ICCCE) (pp. 830-835). IEEE.

[7.] Fkirin, A., Attiya, G., El-Sayed, A. and Shouman, M.A., 2022. Copyright protection of deep neural network models using digital watermarking: a comparative study. Multimedia Tools and Applications, 81(11), pp.15961-15975.

[8.] Pavlović, K., Kovačević, S. and Đurović, I., 2020, November. Speech watermarking using deep neural networks. In 2020 28th Telecommunications Forum (TELFOR) (pp. 1-4). IEEE.

[9.] Abdelwahab, K.M., Abd El-atty, S.M., El-Shafai, W., El-Rabaie, S. and Abd El-Samie, F.E., 2020. Efficient SVD-based audio watermarking technique in FRT domain. Multimedia Tools and Applications, 79, pp.5617-5648.

[10.] Salah, E., Amine, K., Redouane, K. and Fares, K., 2021. A Fourier transform based audio watermarking algorithm. Applied Acoustics, 172, p.107652.

[11.] Sinhal, R., Jain, D.K. and Ansari, I.A., 2021. Machine learning based blind color image watermarking scheme for copyright protection. Pattern Recognition Letters, 145, pp.171-177.

[12.] Shelke, R.D. and Nemade, M.U., 2016, December. Audio watermarking techniques for copyright protection: A review. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC) (pp. 634-640). IEEE.

[13.] Wu, S., Huang, Y., Guan, H., Zhang, S. and Liu, J., 2022. ECSS: High-Embedding-Capacity Audio Watermarking with Diversity Reception. Entropy, 24(12), p.1843.

[14.] Charfeddine, M., Mezghani, E., Masmoudi, S., Amar, C.B. and Alhumyani, H., 2022. Audio watermarking for security and Non-security applications. IEEE Access, 10, pp.12654-12677.

[15.] Yamni, M., Karmouni, H., Sayyouri, M. and Qjidaa, H., 2022. Efficient watermarking algorithm for digital audio/speech signal. Digital Signal Processing, 120, p.103251.

[16.] Kim, W. and Lee, K., 2020, May. Digital watermarking for protecting audio classification datasets. In ICASSP 2020-2020 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2842-2846). IEEE.

[17.] Chen, C.J., Huang, H.N., Tu, S.Y., Lin, C.H. and Chen, S.T., 2021. Digital audio watermarking using minimum-amplitude scaling on optimized DWT low-frequency coefficients. Multimedia Tools and Applications, 80, pp.2413-2439.

[18.] Mosleh, M., Setayeshi, S., Barekatain, B. and Mosleh, M., 2021. A novel audio watermarking scheme based on fuzzy inference system in DCT domain. Multimedia Tools and Applications, 80, pp.20423-20447.

[19.] Abdelwahab, K.M., Abd El-atty, S.M., El-Shafai, W., El-Rabaie, S. and Abd El-Samie, F.E., 2020. Efficient SVD-based audio watermarking technique in FRT domain. Multimedia Tools and Applications, 79, pp.5617-5648.

[20.] Suresh, G., Narla, V.L., Gangwar, D.P. and Sahu, A.K., 2022. False-positive-free SVD based audio watermarking with integer wavelet transform. Circuits, Systems, and Signal Processing, 41(9), pp.5108-5133.

[21.] Pourhashemi, S.M., Mosleh, M. and Erfani, Y., 2021. A novel audio watermarking scheme using ensemble-based watermark detector and discrete wavelet transform. Neural Computing and Applications, 33, pp.6161-6181.

[22.] Hu, H.T. and Lee, T.T., 2019. Hybrid blind audio watermarking for proprietary protection, tamper proofing, and self-recovery. IEEE Access, 7, pp.180395-180408.

[23.] Pourhashemi, S.M., Mosleh, M. and Erfani, Y., 2019. Audio watermarking based on synergy between Lucas regular sequence and Fast Fourier Transform. Multimedia Tools and Applications, 78(16), pp.22883-22908.

[24.] Narla, V.L., Gulivindala, S., Chanamallu, S.R. and Gangwar, D.P., 2021. BCH encoded robust and blind audio watermarking with tamper detection using hash. Multimedia Tools and Applications, 80(21-23), pp.32925-32945.

[25.] Islam, M.S., Naqvi, N., Abbasi, A.T., Hossain, M.I., Ullah, R., Khan, R., Islam, M.S. and Ye, Z., 2021. Robust dual domain twofold encrypted image-in-audio watermarking based on SVD. Circuits, Systems, and Signal Processing, 40, pp.4651-4685.

[26.] Lu, W., Li, L., He, Y., Wei, J. and Xiong, N.N., 2020. RFPS: a robust feature points detection of audio watermarking for against desynchronization attacks in cyber security. IEEE Access, 8, pp.63643-63653.

[27.] Nair, U. and Birajdar, G.K., 2020. Compressed domain secure, robust and high-capacity audio watermarking. Iran Journal of Computer Science, 3, pp.217-232.

[28.] Hu, H.T. and Lee, T.T., 2019. High-performance self-synchronous blind audio watermarking in a unified FFT framework. IEEE Access, 7, pp.19063-19076.

[29.] Nejad, M.Y., Mosleh, M. and Heikalabad, S.R., 2020. An enhanced LSB-based quantum audio watermarking scheme for nano communication networks. Multimedia Tools and Applications, 79(35-36), pp.26489-26515.

[30.] Mosleh, M., Setayeshi, S., Barekatain, B. and Mosleh, M., 2019. High-capacity, transparent and robust audio watermarking based on synergy between DCT transform and LU decomposition using genetic algorithm. Analog Integrated Circuits and Signal Processing, 100, pp.513-525.

[31.] Nejad, M.Y., Mosleh, M. and Heikalabad, S.R., 2020. A blind quantum audio watermarking based on quantum discrete cosine transform. Journal of Information Security and Applications, 55, p.102495.

[32.] Wu, S., Huang, Y., Guan, H., Zhang, S. and Liu, J., 2022. ECSS: High-Embedding-Capacity Audio Watermarking with Diversity Reception. Entropy, 24(12), p.1843.

[33.] Bhat, K.V., Das, A.K. and Lee, J.H., 2019. A mean quantization watermarking scheme for audio signals using singular-value decomposition. IEEE Access, 7, pp.157480-157488.

[34.] Qin, J., Lyu, S., Deng, J., Liang, X., Xiang, S. and Chen, H., 2023. A lattice-based embedding method for reversible audio watermarking. IEEE Transactions on Dependable and Secure Computing.

[35.] Dhar, P.K., Chowdhury, A.H. and Koshiba, T., 2020. Blind audio watermarking based on parametric slant-hadamard transform and hessenberg decomposition. Symmetry, 12(3), p.333.

[36.] Chen, K., Yan, F., Iliyasu, A.M. and Zhao, J., 2019. Dual quantum audio watermarking schemes based on quantum discrete cosine transform. International Journal of Theoretical Physics, 58, pp.502-521.

[37.] Deeba, F., Kun, S., Dharejo, F.A., Langah, H. and Memon, H., "Digital watermarking using deep neural network", International Journal of Machine Learning and Computing, vol.10, no.2, pp.277-282, 2020.

[38.] Pourhashemi, S.M., Mosleh, M. and Erfani, Y., "A novel audio watermarking scheme using ensemble-based watermark detector and DWT", Neural Computing and Applications, vol.33, no.11, pp.6161-6181, 2021.

[39.] Galajit, K., Karnjana, J., Aimmanee, P. and Unoki, M., "Digital audio watermarking method based on singular spectrum analysis with automatic parameter estimation using a convolutional neural network", In proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 63-73, 2018.

[40.] Kaur, A., Dutta, M.K. and Prinosil, J., "General regression neural network based audio watermarking algorithm using torus automorphism", In proceedings of 41st International Conference on Telecommunications and Signal Processing (TSP),pp. 1-4, 2018.

[41.] Mousavirad, S.J. and Ebrahimpour-Komleh, H., “Human mental search: a new population-based metaheuristic optimization algorithm,” Applied Intelligence, 47(3), pp.850-887, 2017.

[42.] He, S., Wu, Q.H. and Saunders, J.R., "Group search optimizer: an optimization algorithm inspired by animal searching behavior", IEEE transactions on evolutionary computation, vol.13, no.5, pp.973-990, 2009.

[43.] BraTS dataset, “https://www.med.upenn.edu/cbica/brats2020/data.html”.

[44.] Le Merrer, E., Perez, P. and Trédan, G., 2020. Adversarial frontier stitching for remote neural network watermarking. Neural Computing and Applications, 32(13), pp.9233-9244.

[45.] Ferdowsi, A. and Saad, W., 2018, May. Deep learning-based dynamic watermarking for secure signal authentication in the Internet of Things. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

[46.] Ingaleshwar, S. and Dharwadkar, N.V., 2021. Water chaotic fruit fly optimization-based deep convolutional neural network for image watermarking using wavelet transform. Multimedia Tools and Applications, pp.1-25.

[47.] Sy, N.C., Kha, H.H. and Hoang, N.M., 2020. An efficient robust blind watermarking method based on convolution neural networks in wavelet transform domain. Int. J. Mach. Learn. Comput, 10, pp.675-684.

[48.] Das, Arjon, and Xin Zhong. "A deep learning-based audio-in-image watermarking scheme." In 2021 International Conference on Visual Communications and Image Processing (VCIP), pp. 1-5. IEEE, 2021.

[49.] Patil, Abhijit J., and Ramesh Shelke. "An effective digital audio watermarking using a deep convolutional neural network with a search location optimization algorithm for improvement in Robustness and Impercepetibility." High-Confidence Computing (2023): 100153.

[50.] Digital Watermarking: 4th International Workshop, IWDW 2005, Siena, Italy, September 15-17, 2005, ProceedingsM. Barni, I. J. Cox, and T. Kalker, Eds.. New York: Springer, 2005, pp. 260–274, Lecture Notes in Computer Science 3710.

[51.] M.D. Swanson, B. Zhu, A. H. Twefik, and L. Boney, “Robust audio watermarking using perceptual masking,” Signal Process., vol. 66, no. 3, pp. 337–355, May 1998.

[52.] C. Cveji, A. Keskinarkaus, and T. Seppanen, “Audio watermarking using m-sequence and temporal masking[A],” in Proc. 7th IEEE Workshop Applicat. Signal Process. Audio Acoust., New York, 2001, pp. 227–230.

[53.] J. Seok, J. Hong, and J. Kim, “A novel audio watermarking algorithm for copyright protection of digital audio,” ETRI J., vol. 24, no. 3, pp. 181–189,

[54.] D. Kir and H. S. Malvar, “Spread spectrum watermarking of audio signals,” IEEE Trans. Signal Process., vol. 51, no. 4, pp. 1020–1033, Apr. 2003.

[55.] H. J. Kim, T. Kim, and I. K. Yeo, “A robust audio watermarking scheme,” in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS’04), 2004, pp. 696–699.

[56.] L. Girin and S. Marchand, “Watermarking of speech signals using the sinusoidal model and frequency modulation of the partials,” in IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP’04), 2004, pp. 633–636.

[57.] H. S. Lee and W. S. Lee, “Audio watermarking through modification of tonal maskers,” ETRI J., vol. 27, no. 5, pp. 608–61, 2005.

[58.] W. Jian and L. Fu-zhong, “Digital audio watermarking based on support vector machine (SVM),” J. Comput. Res. Devel., vol. 42, no. 9, pp. 1605–1611, 2005.

[59.] Y. G. Fu, R. Shen, and H. Lu, “Watermarking scheme based on support vector machine for color images,” Ints. Electron. Eng. Electron. Lett., vol. 40, no. 16, pp. 986–987, 2004.

[60.] C.-H. Li, Z.-D. Lu, and K. Zhou, “SVR-parameters selection for image watermarking,” in Proc. 17th IEEE Int. Conf. Tools with Artif. Intell. (ICTAI’05), Hong Kong, China, Nov. 14–16, 2005, pp. 466–470.

[61.] B. Chen and G. W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” IEEE Trans. Inf. Theory, vol. 47, no. 4, pp. 1423–1443, 2001.

[62.] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[63.] Y. Wang, “A new watermarking method of digital audio content for copyright protection,” in Proc. Int. Conf. Signal Process., Beijing, China, 1998, pp. 75–78.

[64.] I.J. Cox, M.L. Miller, The first 50 years of electronic watermarking, J. Appl. Signal Process. 56 (2) (2002) 225–230.

[65.] M. Barni, I.J. Cox, T. Kalker, Digital watermarking, in: Fourth International Workshop, International Workshop on Digital Watermarking 2005 (IWDW 2005), Siena, Italy, September 15–17, 2005, Lecture Notes in Computer Science, vol. 3710, Springer, Berlin, 2005, pp. 260–274.

[66.] L. Wen-Nung, C. Li-Chun, Robust and high-quality time-domain audio watermarking subject to psycho acoustic masking, in: Proceedings of the IEEE International Symposium on Circuits and Systems, vol. 2, AZ, 2002, pp. 45–48.

[67.] D. Megías, J. Herrera-joancomartí, J. Minguillón, A robust audio watermarking scheme based on MPEG 1 layer III compression, Communications and Multimedia Security—CMS 2003, Lecture Notes in Computer Science, vol. 963, Springer, New York, 2003, pp. 226–238.

[68.] H.J. Kim, Audio watermarking techniques, in: Pacific Rim Workshop on Digital Steganography, Kyushu Institute of Technology, Kitakyushu, Japan, July 3–4, 2003.

[69.] L. Wei, Y. Yi-Qun, L. Xiao-Qiang, Overview of digital audio watermarking, J. Commun. 26 (2) (2005) 100–111.

[70.] S. Sheng-He, L. Zhe-Ming, N. Xia-Mu, Digital Watermarking Technique, Science, Beijing, China, 2004.

[71.] J. Seok, J. Hong, J. Kim, A novel audio watermarking algorithm for copyright protection of digital audio, ETRI J. 24 (3) (2002) 181–189.

[72.] C.-P. Wu, P.-C. Su, C.-C. Jay Kuo, Robust audio watermarking for copyright protection, in: Proceedings of the SPIE, vol. 3807, July 1999, pp. 387–397.

[73.] W. Li, X.Y. Xue, Audio watermarking based on music content analysis: robust against time scale modification, in: Proceedings of the Second International Workshop on Digital Watermarking, Korea, 2003, pp. 289–300.

[74.] L. Girin, S. Marchand, Watermarking of speech signals using the sinusoidal model and frequency modulation of the partials, in: IEEE International Conference on Acoustics, Speech, and Signal processing (ICASSP 2004), 2004, pp. 633–636.

[75.] H.S. Lee, W.S. Lee, Audio watermarking through modification of tonal maskers, ETRI J. 27 (5) (2005) 608–661.

[76.] S.Q. Wu, J.W. Huang, Y.Q. Shi, Efficiently self-synchronized audio watermarking for assured audio data transmission, IEEE Trans. Broadcast. 51 (1) (2005) 69–76.

[77.] J.W. Huang, Y. Wang, Y.Q. Shi, A blind audio watermarking algorithm with selfsynchronization, in: Proceedings of the IEEE International Symposium Circuits and Systems, vol. 3, AZ, 2002, pp. 627–630.

[78.] C.T. Du, R.D. Wang, An audio watermarking detecting improved algorithm based on HAS, J. Eng. Grap. 26 (1) (2005) 74–79.

[79.] W. Li, X.Y. Xue, An audio watermarking technique that is robust against random cropping, J. Comput. Music 27 (4) (2003) 58–68.

[80.] Wu, Q., Ding, R. and Wei, J., 2022. Audio watermarking algorithm with a synchronization mechanism based on spectrum distribution. Security and Communication Networks, 2022.

[81.] Bajpai, J. and Kaur, A., 2016, January. A literature survey-Various audio watermarking techniques and their challenges. In 2016 6th International Conference-Cloud System and Big Data Engineering (pp. 451-457). IEEE.

[82.] Favorskaya, M. N., & Pakhirka, A. I. (2021). Adaptive HVS objectivity-based watermarking scheme for copyright protection. Procedia Computer Science, 192, 1441–1450

[83.] Ben Jabra, S., Ben Farah, M. "Deep Learning-Based Watermarking Techniques Challenges: A Review of Current and Future Trends.", Circuits Syst Signal Process (2024). https://doi.org/10.1007/s00034-024-02651-z

[84.] Z. Wang et al., "Data Hiding With Deep Learning: A Survey Unifying Digital Watermarking and Steganography," in IEEE Transactions on Computational Social Systems, vol. 10, no. 6, pp. 2985-2999, Dec. 2023,

[85.] Zear, A., Singh, A.K. and Kumar, P., "A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine", Multimedia tools and applications, vol.77, no.4, pp.4863-4882, 2018.

[86.] A. Das and X. Zhong, "A Deep Learning-based Audio-in-Image Watermarking Scheme," *2021* International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 2021, pp. 1-5, doi: 10.1109/VCIP53242.2021.9675375.

[87.] Sun, L., Xu, J., Liu, S., Zhang, S., Li, Y. and Shen, C.A., "A robust image watermarking scheme using Arnold transform and BP neural network", Neural Computing and Applications, vol.30, no.8, pp.2425-2440, 2018..

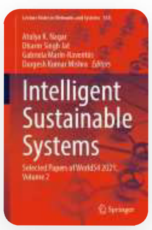
Appendix – I

Research Publication

Digital Audio Watermarking: Techniques, Applications, and Challenges

| Conference paper | First Online: 17 December 2021

| pp 679–689 | Cite this conference paper



Intelligent Sustainable Systems

Abhijit Patil & Ramesh Shelke

Part of the book series: Lecture Notes in Networks and Systems ((LNNS,volume 334))

Abstract

Increased use of Internet has led to increased sharing, storing, and distribution of the digital media across the globe. In the current era, huge amount of digital media, i.e., text documents, image, audio, and video is being populated and distributed at very fast rate. Therefore, it is very important to maintain the authenticity and copyright information of the digital media. Also, security of the digital media during transmission and receipt is an important issue. It is observed that the digital media is prone to malicious attack and being pirated. Watermarking is one of the solutions for providing the security, authenticity, and copyright to the digital media. Watermarking has been used in many applications in the area of image processing, speech processing, broadcasting of digital media, etc. The basic entities on which watermarking is applied are text documents, images, audio, and video. It becomes important to achieve robustness, security, and imperceptibility while transmitting the data with watermark. Lot of research is already been done in text and image watermarking as compared to audio and video. In this paper, we discussed about important concepts of digital audio watermarking and explain few important applications through the literature survey of various techniques used for it.

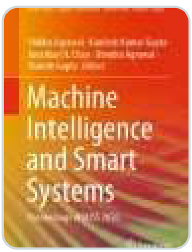
Softcover Book EUR 219.99

Tax calculation will be finalised at checkout

Purchases are for personal use only

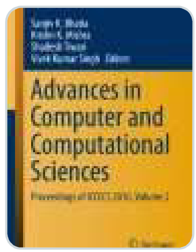
Institutional subscriptions →

Similar content being viewed by others



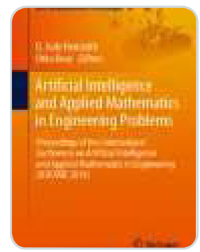
A Survey on Watermarking and Its Techniques

Chapter | © 2021



Digital Audio Watermarking: A Survey

Chapter | © 2018



A Review on Watermarking Techniques for Multimedia Security

Chapter | © 2020

References

1. Saini, L.K., Shrivastava, V.: A survey of digital watermarking techniques and its applications. Int. J. Comput. Sci. Trends Technol. (IJCST) 2(3) (2014)

Google Scholar

2. Jain, R., Trivedi, M.C., Tiwari, S.: Digital audio watermarking: a survey. In: Bhatia, S., Mishra, K., Tiwari, S., Singh, V. (eds.) Advances in Computer and Computational

Sciences. Advances in Intelligent Systems and Computing, vol. 554. Springer, Singapore (2018)

Google Scholar

3. Thanki, R.M.: Advanced Techniques for Audio Watermarking. Signals and Communication Technology. Springer (2020)

Google Scholar

4. Garg, P., Kishore, R.R.: Performance comparison of various watermarking techniques. Multimed. Tools Appl. **79**, 25921–25967 (2020)

Google Scholar

5. Hua, G., Huang, J., Shi, Y.Q., Goh, J., Thing, V.L.L.: Twenty years of digital audio watermarking—a comprehensive review. Sig. Process. **128**, 222–242 (2016). https://doi.org/10.1016/j.sigpro.2016.04.005

6. Bajpai, J., Kaur, A.: A literature survey—various audio watermarking techniques and their challenges. In: Proceedings of the 2016 6th International Conference—Cloud System and Big Data Engineering, Confluence 2016, pp. 451–457 (2016)

Google Scholar

7. Steinebach, M., Dittmann, J.: Watermarking-based digital audio data authentication. EURASIP J. Appl. Sig. Process. **10**, 1001–1015 (2003)

Google Scholar

8. Xiang, S., Yang, L., Wang, Y.: Robust and reversible audio watermarking by modifying statistical features in time domain. In: Advances in Multimedia (2017)

Google Scholar

9. Shelke, R.D., Nemade, M.U.: Audio watermarking techniques for copyright protection: a review. In: Proceedings—International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016, pp.634–640 (2017)

Google Scholar

10. Arnold, M.: Audio watermarking: features, applications and algorithms. In: IEEE International Conference on Multi-Media and Expo, pp. 1013–1016 (2000)

Google Scholar

11. Chauhan, S.P.S., Rizvi, S.A.M.: A survey: digital audio watermarking techniques and applications. In: Proceedings—4th IEEE International Conference on Computer and Communication Technology, ICCCT 2013. IEEE Computer Society, pp. 185–192 (2013)

Google Scholar

12. Podilchuk, C.I., Delp, E.J.: Digital watermarking: algorithms and applications. IEEE Sig. Process. Mag. **18**(4), 33–46 (2001). https://doi.org/10.1109/79.939835

13. Thanki, R.M.: Audio watermarking with encryption. In: Signals and Communication Technology, pp. 59–81. Springer (2020)

Google Scholar

14. Xiang, Y., Hua, G., Yan, B.: Digital Audio Watermarking Fundamentals, Techniques and Challenges. Springer Briefs in Electrical and Computer Engineering (2017)

Google Scholar

15. Olanrewaju, R.F., Khalifa, O.: Digital audio watermarking; techniques and applications. In: International Conference on Computer and Communication Engineering (ICCCE 2012), pp 3–5 (2012)

Google Scholar

16. Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J.J., Su, J.K.: Attacks on digital watermarks: classification, estimation-based attacks, and benchmarks. IEEE Commun. Mag. **39**(8), 118–125 (2001)

Google Scholar

17. Lalitha, N.V., Rao, C.S., Jaya Sree, P.V.Y.: A review of digital audio watermarking schemes. J. Crit. Rev. **7**(7) (2020)

Google Scholar

18. Bamatraf, A., Ibrahim, R., Mohd Salleh, M.N.B.: Digital watermarking algorithm using LSB. In: 2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE 2010) (2010)

Google Scholar

19. Hemis, M., Boudraa, B.: Digital watermarking in audio for copyright protection. In: Proceedings—ICACSIS 2014: 2014 International Conference on Advanced Computer Science and Information Systems, pp. 189–193 (2014)

Google Scholar

Author information

Authors and Affiliations

Pacific University, Udaipur, India

Abhijit Patil & Ramesh Shelke

Editor information

Editors and Affiliations

School of Mathematics, Computer Science, and Engineering, Liverpool Hope University, Liverpool, UK

Atulya K. Nagar

Namibia University of Science and Technology, Windhoek, Namibia

Dharm Singh Jat

University of Costa Rica, Curridabat, San Jose, Costa Rica

Gabriela Marín-Raventós

Sri Aurobindo Institute of Technology, Indore, Madhya Pradesh, India

Durgesh Kumar Mishra

Rights and permissions

Copyright information

About this paper

Cite this paper

Patil, A., Shelke, R. (2022). Digital Audio Watermarking: Techniques, Applications, and Challenges. In: Nagar, A.K., Jat, D.S., Marín-Raventós, G., Mishra, D.K. (eds) Intelligent Sustainable Systems. Lecture Notes in Networks and Systems, vol 334. Springer, Singapore. https://doi.org/10.1007/978-981-16-6369-7_62

.RIS .ENW .BIB

DOI
https://doi.org/10.1007/978-981-16-6369-7_62

Published
17 December 2021

Publisher Name
Springer, Singapore

Print ISBN
978-981-16-6368-0

Online ISBN
978-981-16-6369-7

eBook Packages
Intelligent Technologies and Robotics
Intelligent Technologies and Robotics (R0)

Publish with us

Policies and ethics

Digital Audio Watermarking : Techniques, Applications and Challenges.

Mr. Abhijit Patil
PhD Research Scholar,
Pacific University, Udaipur, India

Dr. Ramesh Shelake
Research Co-Supervisor,
Pacific University, Udaipur, India

i. Abstract

Increased use of internet has led to increased sharing, storing and distribution of the digital media across the globe. In the current era, huge amount of digital media i.e. text documents, image, audio and video is being populated and distributed at very fast rate. Therefore, it is very important to maintain the authenticity and copyright information of the digital media. Also, security of the digital media during transmission and receipt is an important issue. It is observed that the digital media is prone to malicious attack and being pirated. Watermarking is one of the solutions for providing the security, authenticity and copyright to the digital media. Watermarking has been used in many applications in the area of image processing, speech processing, broadcasting of digital media etc.

The basic entities on which watermarking is applied are text documents, images, audio and video. It becomes important to achieve robustness, security and imperceptibility while transmitting the data with watermark. Lot of research is already been done in text and image watermarking as compared to audio and video. In this paper we discussed about important concepts of digital audio watermarking and explain few important applications through the literature survey of various techniques used for it.

Keywords : Digital audio watermarking, Robustness, Watermark embedding and extraction, DCT, DWT, SVD, LSB, Watermark security and challenges.

ii. Introduction

The term Watermark was initially used at Bologna, Italy in 1282 in paper mills as paper mark of the company. From long back watermarks are being used in currency notes and postage stamps of many countries. The term Digital watermark was discovered in the year 1992, by Andrew Trickle and Charles Osborne.

Steganography is similar technique to watermarking but the basic difference here is that in steganography the digital data is hidden within other data for secure transmission. It doesn't provide copyright and can't handle the attacks. Its purpose is to just hide the information inside the digital media.

Digital audio watermarking is a process of hiding secure information (called as watermark) within a digital audio file, without affecting the quality of the original audio signal. [3,9]. As compared to image watermarking, Audio watermarking is more challenging because it is observed that the hearing capacity of humans is more powerful than the visual field. It makes use of HAS (Human Auditory System) properties such as frequency range, frequency discrimination, amplitude discrimination, Amplitude range.

iii. Important Properties of Audio Watermarking [1, 4, 8, 9]

Following are some important properties of audio watermarking.

Perceptual transparency.

It deals with the quality of original signal. Simply it is can be defined as the quality of watermarking procedure in which the quality of original signal is not degraded while inserting the watermark into original audio signal. In another words Perceptual transparency means that the algorithm used for watermarking should insert watermark data without degrading the quality of the host audio signal. This property also tells that the watermark is visible through human sensible organ or not.

Fidelity

Fidelity of a watermarking algorithm is viewed as the perceptual similarity between the watermarked audio and the original host audio signal when they are presented to a consumer. It is possible that the quality of watermarked audio signal may degrade while transmission because of adversary. The watermarking system should be able to minimize the changes in the host signal. It should keep the changes minimum as much as possible, and unnoticeable. Also the watermark should be invisible.

Watermark Bit rate

The total number of bits from the watermark signal which are embedded in the host audio signal in the duration of one second is called as Bit Rate. And it is given in terms of bits per second (bps).

Robustness

Robustness is the most important property of all watermarking techniques. It means that the watermark should be robust to withstand maximum possible attacks and threats. It can also be defined as the ability of watermarking algorithm to withstand against easy extraction of the watermark on application of common signal processing manipulations. Which means that the watermark should not be easily removed and detected only through authorized person.

Security

The host signal should be secure against the attacks after embedding the watermark into it. The attacker must not be able to easily detect the presence of embedded watermark. Anyone should not be able to remove the embedded data. Any unauthorized user should not be able to extract the data within a specific amount of time. Also, even if the attacker knows the exact watermark embedding algorithm, he should not be able to extract the watermark. Security can be increased with the help of encryption and decryption keys at the time of embedding and extraction process.

iv. Watermarking Procedure [1, 2, 3,5,7]

It is a procedure through which watermark signal is carried along with original signal. To achieve robustness, watermark is embedded without affecting the quality of the original audio signal. It consists of two steps:

1. Watermark Embedding
2. Watermark Extraction

The secret signal which is embedded into host signal is called as watermark. The original signal into which watermark is embedded is called as cover object. After embedding the original signal gets modified, which is called as embedded signal or watermarked data. In some applications we may have to use encryption key to improve the security and robustness of watermark.

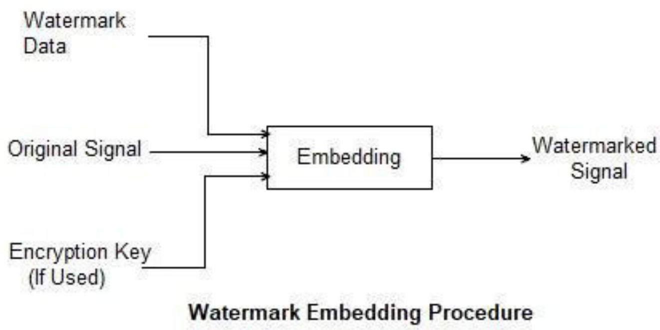


Fig. 1 Watermark Embedding

The embedding block, shown in Figure 1, Embedding procedure takes three inputs: watermark, original signal, and watermarking key. It produces watermarked signal as the output.

Whereas the inputs to the extraction block are embedded object, watermark key (decryption key if encryption is used while embedding). It produces the watermark as output as shown in Figure 2.

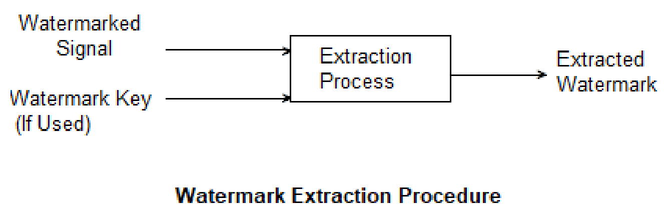


Fig. 2 Watermark Extraction

v. Applications of Audio Watermarking [1, 2, 7, 9, 10]

1. Copyright protection

Copyright protection is one of the important applications of watermarking. The objective here is to add copyright owner's identification information into the signal through watermark. It prevents others from claiming the copyright for same signal. This application needs to satisfy the high level of robustness and security.

2. Fingerprinting

High consumption of digital audio and video has led to many possible attacks. Multimedia piracy is a serious issue nowadays. Many companies want to add legal information along with the multimedia data to prove its authenticity and ownership.

Digital fingerprinting is the procedure in which a serial number is added to multimedia data and is used to trace the original legal buyer of the data. It deals with embedding the watermark into different distributed copies. It is basically used to carry information about the legal recipient of digital media. It enables to identify single distributed copies of digital work. It is very much similar to the concept of serial number of software product. It also requires high degree of robustness.

3. Content Authentication

There is a possibility of attacks during the transmission of digital media. The original signal may get modified. The purpose of this application is to detect any modifications to the original data. This application aims in finding the authenticity of original signal.

4. Copy Protection

Protecting the digital data from unauthorized copy. This application is used to protect the original content from unauthorized copy. It has been observed that copy protection is little bit difficult to achieve in open systems. However, in closed systems it is easily possible. In closed systems we can use watermarks to indicate the copy status of the digital media. The status can be like copy once or never copy. Also, we can use the copy software or a device which is able to detect the watermark. The system should be able to grant the requested operation based on the copy status of the digital media being copied.

5. *Broadcast Monitoring*

The process of tracking the digital media broadcasting channels for checking whether they are following the broadcasting laws and protecting the intellectual property rights.

While broadcasting the audio and video data such as advertisements it becomes important that they should be broadcasted on the time after they are purchased from the broadcasters. And it should be monitored.

There are many techniques available for broadcast monitoring. Some methods are costly and involves lot of errors. Hence there is a need of automatic monitoring. One solution to this is to use the watermarking techniques. With watermarking we can embed the watermark in the original work being broadcasted. And then a computer-based monitoring system can detect the embedded watermark, to make sure that they receive all of the digital data they purchase from the broadcasters. Hence this application requires a good watermarking technique and monitoring software.

vi. Possible attacks on Audio Watermarking. [1, 7, 9, 20]

As mentioned earlier, robustness is the most important property of any watermarking. It can be verified by subjecting it to various attacks. Resampling, noise, filter, and cropping are common signal processing attacks.

***Resampling*:** Resampling is nothing but the change in digital audio signal by altering the sample rate. In this type of attack, the sampling rate of digital audio signal is converted from one sampling rate to different one. Changes are made to sample values of the watermarked audio signal. For example, if watermarked audio signal sampled at 16 kHz, then it is resampled stepwise such as initially it is resampled at 8 kHz and then up sampled again to 16 kHz.

Noise: Noise is an unwanted signal added to original signal during the transmission or processing. Noise makes it difficult for the original signal to reach its destination. Noise affects the quality and usefulness of digital media. In this attack, most of the times an attacker can add Additive white Gaussian noise (AWGN) to degrade the quality of original audio signal as well as the watermark.

Filtering Attack: A filter is basically used to remove the unwanted features from digital signal. However, attackers may use some types of filters to extract the watermark from watermarked signal. In this type of attack, the watermarked audio signal is passed through a filter. Commonly used filters in signal processing are Low pass and high pass, and band pass filters. In watermarking attacks, generally a low-pass filter is used to remove the low-frequency component from it.

Cropping Attack: In this, some part of the watermarked audio signal is removed. And then it may be replaced by noise. For example, If the original watermarked signal contains 50,000 samples, then the attacker may remove 20000 out of 50,000 samples and replaced them by noise. Hence the overall quality of watermarked signal is degraded and some part of watermark may get removed, which makes the signal useless.

Cryptographic Attack: Encryption and decryption keys may be used to enhance the security of watermarked signal. But there are many types of cryptographic attacks available to steal the keys such as brute-force attack, replay attack. Attackers makes use of these cryptographic attacks break the security of watermark by stealing the key used for watermarking procedure. Once the attacker obtains the key, he can rewrite the watermark or he can guess the watermark data.

vii. Different Techniques used for Audio Watermarking. [1, 9,14]

Watermarking techniques are classified based on variety of criteria such as perceptual quality, accessibility, type of data used in watermarking (text, image, audio, video), type of applications etc. However, implementation of Audio watermarking technique can be done into two major categories,

1) Time domain: These are simple techniques which makes use of modification to the bit values of the original signal.
   The advantage of these techniques is that the original signal is not required while extracting the watermark. As well as they provide good payload capacity.
   The drawback here is that they are not robust and get distorted by attacks. It has sub types such as,
   a) Additive Watermarking
   b) Watermarking based on substitution.

2) Transform domain: These are the techniques which makes use of different transforms like DFT, DCT, DWT. The audio signal is converted into its transform coefficients. The watermark signal is embedded into these coefficients. Then inverse transform is applied to obtain watermarked audio signal back. It has subtypes such as,
   a) Substitution based Audio Watermarking
   b) Additive Watermarking
   c) Multiplicative Watermarking.

3) Hybrid Techniques: These techniques make use of combination of two or more transforms for watermarking process. The most widely used combinations are DWT+DCT, DWT+SVD, DWT+DCT+SVD.

LSB Method [1,2,7]: It is mostly used in image watermarking for applications such as copyright protection. It is very easy and simple technique. In this method the bit value of least significant bit is replaced by the watermark signal. Changing the LSB of a pixel do not degrade the quality of original image up-to a large extent. These changes in original image are not visible to human visual system because the intensity of pixels doesn't change much. Up to 4 bits can be used to hide secret information. However, this technique has a drawback that it is not much secure, hackers can easily find out the hidden information if they come to the technique. Also, there can be undesirable noise and data can be lost if lossy compression techniques are used.

DFT [1]

DFT is a very basic and important technique in transformation domain. It is used in many applications for performing Fourier analysis of the input signal. It decomposes the signal into its fundamental sinusoidal frequency coefficients. These coefficients are then replaced by watermark signal. A DFT for sequence of N complex numbers $X\{n\} = x_0, x_1, \ldots x_{N-1}$ is converted into another sequence of complex numbers by using,

$$S(k) = \sum_{n=0}^{n=N-1} Xn \, . \, e^{\frac{\sqrt{2\pi}\,Kn}{N}}$$

DCT [1,2]

Proposed by Nasir Ahmed in 1972, is the technique of representing the data points as sum of cosine functions. It is most widely used transformation domain technique in signal processing as well as compression. It is used in digital images, audio and video data. Cosine functions are critical for compression. DCT is very much similar to DFT but it considers only real valued cosine functions.

There are two ways to represent the audio signal: 1) analog audio: representing different levels of electrical voltage. 2) digital audio: representing the sequence of binary numbers. In case of audio signals DCT is used to convert the signal into its frequency version. In audio watermarking DCT coefficients of the signal are used for embedding. It is represented by following equation:

$$S(k) = c(k) \sum_{n=0}^{N-1} S(n) Cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right)$$

Where k = 0 to N-1,

S(n): input audio signal,

$$c(k) = 1/N$$

DWT [1,6,9]

It is a transform domain technique in which original signal is decomposed into a number of sets where each set of coefficients describes time information of the signal in given frequency band.

It is based on subband coding. It gives time scale representation of audio signal. Here we use the concept of wavelets. First DWT was initially invented by Alfred Harr called as Harr Wavelet. The family of wavelets contain Harr, Daubichies, Coiflets, Symlets, Discrete Meyer, biorthogonal wavelet transforms. The wavelets are discretely sampled.

In DWT the wavelets are defined as the function integrated to zero waving above and below X-axis. DWT of a signal is obtained by passing the signal through filters. The original signal is developed by using linear combination of wavelets. Some data operations are applied onto it.

SVD [6]

It uses linear algebra operations. When it is used in image data, images are represented as matrix and then linear algebra operations are performed using SVD technique. A matrix is decomposed into eigenvectors and eigenvalues. In SVD based image watermarking the watermark signal is embedded into host image by replacing singular values. The SVD technique can be described as:

$(U, S, V) = SVD(F)$

U, V: two orthogonal matrices.

S: a singular matrix,

Singular Value Decomposition of these matrices is represented by equation $F = U \times S \times V^{T}$,

Matrices U and V contain either real or complex values,

S: a diagonal matrix with nonnegative numbers.

viii. Performance Evaluation Parameters: [1, 8, 10, 14]

1. ***PSNR (Peak Signal to Noise Ratio):*** It is defined as the ratio between the most feasible power of the original signal to the power of corrupting noise. This noise may affect the reliability. Here, the original data is used as a signal, and the noise is the error introduced by compression. The unit used to represent it is logarithmic decibel scale.

It is mostly used in for image compression to calculate the quality of rebuilding of lossy compression codecs. It is easily defined with the help of mean squared error (MSE).

PSNR is the most standard metric for imperceptibility verification.

Mathematical Representation of PSNR:

$$PSNR = 20.\log_{10}(\text{MAX f}) - 10.\log_{10}(MSE)$$

Where MAX f is the maximum signal value,

MSE is the Mean squared error.

Considering the data as an image, the MSE is calculated as:

$$\text{MSE} = \frac{1}{m*n}\sum_{i=0}^{m-1}\sum_{j=0}^{m-1}(f(i,j)-g(i,j))^2$$

Where f is the original image and g is degraded image.

i, j are the row and column number for image data

2. ***BER (Bit Error Rate):*** It is the ratio of erroneous bits received to the total number of bits in the original signal. It finds number of erroneous bits per unit amount of time. Bit errors occur because of distortion, noise or bit synchronization error.

$$BER = \frac{Number\ of\ errorneous\ bits}{Total\ number\ of\ bits}$$

3. ***MSE (Mean Squared Error):*** It is used to compare two signals based on some quantitative measure. It gives us the degree of similarity or the percentage of error / distortion between them.

   Let X= (X_i | I = 1,2,3 …N) and Y = (Y_i | I = 1,2,3, ….. N) are two finite length discrete signals then,

$$\text{MSE}(X, Y) = 1/N \left(\sum_{i=1}^{N} (Xi - Yi)^2 \right.$$

ix. Conclusion

In this paper we discussed some important aspects of audio watermarking. Through the literature survey we found that lot of research work is already done but there is a scope of research in providing robustness and security to the watermarked signal. Time domain and frequency domain techniques (LSB, DFT, DCT, DWT) are most commonly used for watermarking. Some hybrid techniques are designed by many researchers giving better results for providing robustness and security. Although there are many hybrid techniques designed for securing audio watermark data, attackers are finding the new ways to hack the signal and break its security. We plan to find a best hybrid technique which will provide robustness and security to the audio watermarked signal at highest level.

References

[1] Garg, P., Kishore, R.R. Performance comparison of various watermarking techniques. Multimed Tools Appl 79, 25921–25967 (2020).

[2] Shelke, R. D., & Nemade, M. U. (2017). Audio watermarking techniques for copyright protection: A review. In Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016 (pp. 634–640).

[3] Jain R., Trivedi M.C., Tiwari S. (2018) Digital Audio Watermarking: A Survey. In: Bhatia S., Mishra K., Tiwari S., Singh V. (eds) Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, vol 554. Springer, Singapore.

[5] Arnold, M. (2000). Audio watermarking: Features, applications and algorithms. In IEEE International Conference on Multi-Media and Expo (pp. 1013–1016).

[6] Hemis, M., & Boudraa, B. (2014). Digital watermarking in audio for copyright protection. In Proceedings - ICACSIS 2014: 2014 International Conference on Advanced Computer Science and Information Systems (pp. 189–193).

[7] Chauhan, S. P. S., & Rizvi, S. A. M. (2013). A survey: Digital audio watermarking techniques and applications. In Proceedings - 4th IEEE International Conference on Computer and Communication Technology, ICCCT 2013 (pp. 185–192). IEEE Computer Society.

[8] Bajpai, J., & Kaur, A. (2016). A literature survey - Various audio watermarking techniques and their challenges. In Proceedings of the 2016 6th International Conference - Cloud System and Big Data Engineering, Confluence 2016 (pp. 451–457).

[9] Rohit M Thanki. Advanced Techniques for Audio Watermarking. Signals and Communication Technology. Springer. 2020

[10] Yong Xiang., Guang Hua., Bin Yan. Digital Audio Watermarking Fundamentals, Techniques and Challenges. Springer Briefs in Electrical and Computer Engineering. 2017.

[11] Martin Steinebach., Jana Dittmann. Watermarking-Based Digital Audio Data Authentication, EURASIP Journal on Applied Signal Processing 2003:10, 1001–1015

[12] Abdullah Bamatraf., Rosziati Ibrahim., Mohd. Najib B. Mohd Salleh Digital Watermarking Algorithm Using LSB. 2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE 2010), December 5-7, 2010

[13] Lalit Kumar Saini , Vishal Shrivastava. A Survey of Digital Watermarking Techniques and its Applications. International Journal of Computer Science Trends and Technology (IJCST) Volume 2 Issue 3, May-Jun 2014.

[14] N.V. Lalitha , Ch. Srinivasa Rao, P.V.Y. Jaya Sree. A REVIEW OF DIGITAL AUDIO WATERMARKING SCHEMES. Journal of Critical Reviews. Vol 7, Issue 7, 2020

[16] Xiang, S., Yang, L., & Wang, Y. (2017). Robust and Reversible Audio Watermarking by Modifying Statistical Features in Time Domain. Advances in Multimedia, 2017.

[17] Ingemar J. Cox, Joe Kilian, F., Thomson Leighton, Talal Shamoon. Secure Spread Spectrum Watermarking for Multimedia. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 6, NO. 12, DECEMBER 1997.

[18] R. F.Olanrewaju, Othman Khalifa: Digital Audio Watermarking; Techniques and Applications: International Conference on Computer and Communication Engineering (ICCCE 2012), pp 3–5 J (2012).

[19] Thanki, R. M. (2020). Audio watermarking with encryption. In Signals and Communication Technology (pp. 59–81). Springer.

[20] Voloshynovskiy, S., Pereira, S., Pun, T., Eggers, J. J., & Su, J. K. (2001). Attacks on digital watermarks: Classification, estimation-based attacks, and benchmarks. IEEE Communications Magazine, 39(8), 118–125.

[21] Lalit kumar Saini, Vishal Shrivastava, A Survey of Digital Watermarking Techniques and its Applications, IJCST – Volume 2 Issue 3, May-Jun 2014.

[22] Digital Watermarking: https://en.wikipedia.org/wiki/Digital_watermarking.

Effective Digital Audio Watermarking Using Dwt And Neural Networks

Abhijit Patil[1*], Dr. Ramesh Shelke[2]

Abstract

Watermarking is the process in which a digital signal is added with another secret digital signal. Digital audio watermarking has been widely used in many applications such as copyright protection, tamper detection, piracy prevention, content authentication, etc. The audio watermarking process has to satisfy many properties such as robustness, imperceptibility, and security. There are many classical as well as hybrid techniques available in the literature to achieve these properties. However, it's difficult to achieve all the properties at the highest level using a single technique. Echo hiding and pitch shifting approaches are being used from beginning. Some new techniques are being designed that make use of machine learning and deep learning algorithms, "bio-inspired algorithms" such as swarm intelligence algorithms [18], and genetic algorithms, AI-based techniques such as simulated annealing can also be used for optimization in watermarking process. Arnold scrambling and use of cryptographic algorithms can be used for increasing the security of watermarks. However, it is observed that many modern approaches are not giving efficient results in embedding and extraction of watermark with minimum bit error. Finding optimal locations for the watermark bits embedding into host signal is a challenge. In this paper, we are discussing some of the important hybrid and novel techniques used for digital audio watermarking. We proposed and demonstrated the work using a custom-designed simple backpropagation neural network. Our focus is to demonstrate the usefulness of neural network architectures in the subject of study.

Keywords: Audio Watermarking, Neural Network, DWT, Robustness, Imperceptibility, Deep Learning.

DOI Number: 10.14704/Nq.2022.20.17.Nq88005 **Neuroquantology 2022; 20(17):23-30**

1. Introduction

The main reason for Digital audio watermarking technology being more challenging is that "the human auditory system is extremely more sensitive than the human visual system" [28]. Audio watermarking algorithms are broadly classified into time and frequency domain techniques. Frequency domain techniques are more popular in the research community. Conventional techniques used for watermarking procedure are LSB technique, DCT, DWT, SVD, and or the combination of two or more among these [27-28]. Major group of schemes in the topic of study use spread spectrum techniques [32-34]. There is another group of audio schemes which use patchwork algorithms [14, 35-38]. Few techniques use the synthesized echoes of original signal as a secret message to be embedded [29-31]. However, it is observed that neural network-based techniques are also being used for making the watermarking procedure more robust and secure against the attacks. In recent year many researchers had used deep learning algorithms for watermarking [1-14]. The characteristics that determine the effectiveness of watermarking algorithms are Robustness, Imperceptibility and Security. Through the extensive literature survey it is observed that the neural network and deep learning-based techniques have a great impact on increasing the robustness of the watermark. It has been observed that the neural network-based techniques had shown better results in providing robustness and imperceptibility [1-3, 7-13]. "Artificial neural networks" (ANN), "Convolutional Neural Networks" (CNN), Encoders and decoders, are few of them [1-5]. Bit error rate, Mean Square

***Corresponding Author:-** Abhijit Patil
Address: [1*]Ph.D. Scholar, Department of Computer Engineering, PAHER University Udaipur, Rajasthan, India
[2]Research Supervisor, Department of Computer Engineering, PAHER University, Udaipur, Rajasthan, India
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Error (MSE), and Peak Signal Noise Ratio PSNR) are the important parameters that are used for the comparison of different watermarking algorithms. In this paper, we are discussing the neural network-based techniques used for digital audio watermarking and their effectiveness. We are proposing a neural network-based approach for improving the robustness and imperceptibility of watermarked speech data. The aim of using deep learning techniques in audio watermarking is to make the watermark more robust against the attacks, to improve the strength of watermark [5]. As the audio watermarking is more challenging than image watermarking, deep learning techniques are found to be best suitable if used in combination with DWT, DCT. The traditional techniques are manual and provide the strength to watermark up to some limit. We can improve the robustness and imperceptibility of the watermark if a suitable deep learning technique is used. The rest of the paper is organized as: Section 2 involves a literature survey of different neural network-based approaches used previously. Section 3 contains the implementation of the general framework. Section 4 discusses Experimental results. And finally, section 5 concludes the paper.

2. Literature Review

In audio watermarking the commonly used watermarks are binary message [1], binary image [2],[3],[7],[15], or an audio signal[14]. Whenever an image is used as a watermark, it has to be converted into a one-dimensional binary signal so that it can be properly inserted at the desired positions in the host signal. Deep learning networks such as encoders and decoders are found suitable for inserting the watermark data (bits) at the best positions in the host signal so that they cannot be easily extracted or attacked. These techniques are used in combination with conventional techniques i.e. DWT or DCT. The choice of watermark data (Image or Audio) depends on the type of application it is used. We studied research work carried out in recent years in this area and found some good techniques. Here we are discussing few of them along with a table of comparison.

K. Pavlović et.al.[1] have performed speech watermarking using Encoder, Decoder, and STFT techniques. One Dimensional Binary Message was used as a watermark and the dataset was audio recordings from the parliament of Montenegro. The Audio files of 226 speakers of around 232 minutes per speaker were collected for training which were sampled at 44.1 kHz. They obtained good results with Decoder Accuracy 99.82% and PSNR 57.5dB. G. Wu et.al.[2] have used discrete wavelet transform(DWT) on the audio signal, for selecting the important coefficients which are ready to be trained in the neural network. The concept of watermark memorization in the nerve cells of CPN Counter propagation network was used. The network was designed to be fault tolerant. The architecture is designed with an adaptive number of parallel Counter propagation networks. These parallel networks were able to treat each audio frame separately and the corresponding watermark bits. They found that CPN improves the efficiency for watermark embedding process was improved. Also more correctness in extracting the watermark data. The method was robust and as well as improved the inaudibility of audio watermark. C. Maha et. al. [3] in this paper a blind audio watermarking scheme based on neural network is described. The basis of the method is human psychoacoustic model (HPM) with error correcting code. They used DWT, HPM as a reference model, and BPNN. Along with this they used Hamming Code for increasing the security of the watermark. A binary image of size 32*32 is used as a watermark and the audio file (.wav file) with 44.1 Khz Sampling rate, 16 bits per sample was used for testing the results. They found that the HPM provides better robustness and imperceptibility.

Das et.al.[4] presented the deep learning scheme for embedding an audio watermark into an image. An unsupervised learning approach is used. In this paper “a robust and blind audio-in-image watermarking scheme” is described. They designed a network called as similarity network that is able to recognize the audio watermarks under distortions. The deep network is modelled as “Encoder, Decoder, Embedder, and Extractor Networks”. J. Hu et.al. [8] Used triple forward Neural Network and Wavelet transform for embedding the watermark. Important coefficients the audio signal are found to embed the watermark into it. Later on corresponding algorithms of watermark generation, embedding, and extraction are used. It is a "non-blind audio watermarking scheme" in which original signal is needed for while extracting the watermark. The watermark is simulated for common signal processing attacks and shown that the algorithm is robust against the attacks. Chuan-Yu Chang et. al. [9] proposed a novel "Fully Connected Counter Propagation Network." (FCNN) for image watermarking. Along with this they performed: Imperceptibility Testing, Robustness testing(for grayscale watermark, for binary watermark), Authenticity testing. Rather than the cover images, the watermark is embedded in synapses of FCNN. Their experimental results shown that the quality of watermarked image was not degraded. As mentioned in the paper, the watermark become robust, because the watermark was stored in the synapses. Jiang Jing et.al. [10] In this paper "a patchwork method for digital watermarking based on Radial Basis Neural Network (RBNN)" is discussed. The method randomly selects two patches using a key, then some constant value is added in one patch and at the same time subtracted from other. M pairs of patches and 2M sets of pseudorandom numbers are used for embedding watermarking information. The watermark is embedded into the sample audio signal. The RBNN is trained using a randomly selected sample from embedded audio signal. The method is based on the wavelet domain. The watermark signals are embedded in approximation coefficients [10]. Quality of watermark is verified using PSNR and ER(Extraction Ratio) parameters. In works of Sarreshtedari et.al. [22] they have used source-channel coding approach for digital self-embedding speech signals. It generates a tamper-proof signal. Hash generation algorithm is used for preserving MSB of speech signal frames. In the works of Huiqin Wang et. al. [12] Neural Network based Controller was used for checking and ensuring the strength of embedded watermark data. Signal to Mask Ratio from psychoacoustic model (SMR), and DCT coefficients are used as the input to the model. [13] H. Yang et. al, here they had used the concept of WSF(Watermark Scaling Factor) and MMF(Minimum Masking Threshold). WSF is determined with the help of signal data and some statistical parameters. An artificial neural network was designed which not only determines watermark scaling factor (WSF), but uses the concept of MMT such that the power spectrum of watermark always remain below MMT. Embedding of the watermark is carried out in DCT domain. The embedded watermark remains robust against few attacks since it depends on secret key, ANN architecture and final weights.

Authors	Paper Title	Method Used	Watermark Used	Dataset Used	Experimental Results / Remarks
[1]K. Pavlović, S. et. al. 2020	"Speech watermarking using Deep Neural Networks"	Encoder, Decoder, STFT	One Dimensional Binary Message.	The audio recording data is taken from parliament of Montenegro and was sampled at 44.1 kHz. (Total 226 speakers and recordings of total 232 minutes duration per speaker)	Obtained Decoder Accuracy 99.82% PSNR 57.5dB
[2] G. Wu and X. Zhou 2008	"A Fast Audio Digital Watermark Method Based on CPNN"	DWT, Counter Propagation Networks	64*64 bit binary image.	16 bit mono audio signal files : i) music.wav and ii) speaker.wav.	Obtained SNR 61.32 and 40.15 for two samples respectively. BER for extracted data 6.2%
[3] C. Maha, E. Maher and B. A. Chokri 2008	"A blind audio watermarking scheme based on neural network and HPM with error correcting code in wavelet domain"	DWT, Human Psychoacoustic Model, BPNN, Hamming Code	A Binary image of size 32*32 pixels.	A Wav file with 44.1 Khz Sampling rate, 16 bits per sample.	Hamming Code provides security and avoids corruption of watermark. HPM gives better robustness and imperceptibility.
[4] Das, Arjon, and Xin Zhong. 2021	"A Deep Learning-based Audio-in-Image Watermarking Scheme."	"Audio-In-Image Watermarking" "WM Network (Encoder, Decoder, and Embedder, and Extractor Networks)", Similarity Network	"Speech Commands Dataset"	Image, "rescaled 128 × 128 Microsoft COCO Dataset"	Root Mean Squared Error (RMSE) and SSIM parameters are used for verification. High fidelity and robustness was obtained.
[7] C. Chang, W. Shen and H. Wang 2006	"Using Counter-propagation Neural Network for Robust Digital Audio Watermarking in DWT Domain"	DWT based Counter Propagation Network, Synchronization Code.	Binary image of 32 * 32 size.	16 bits mono-track audio music with sampling rate 44.1KHz Audio file of frame size 512.	

[8] J. Hu, X. Qiu and D. He 2008	"Digital Audio Watermarking Algorithm based on Neural Networks"	Multi Layer Feed Foreword network in DWT Domain	Black and white image.	Audio signal with 8 bits/sample, 44.1kHz sample rates. 3-grade Daubechies-4 wavelet to decompose the audio signal.	No distortion in original watermark, Robust against low pass filtering, resampling,and compression attacks
[9] C Yu Chang et. al.	"A neural-network-based robust watermarking scheme."	FCNN	(Used for Image Watermarking)	(Used for Image Watermarking)	(For Image Watermarking)
[12] H. Wang et. al. 2006	"New Audio Embedding Technique Based on Neural Network"	NN Controller, Masking Level, Dimensionality Reduction, Signal to ask ratio.	Binary Image of Size 256 * 256	Audio File of duration 4 Seconds and sampling frequency 22.05KHz.	Robustness is checked against noise, low pass filtering and resampling attacks. There is some distortion in extracted watermark.
[13] H. Yang, et. al. 2002	"An artificial neural network-based scheme for robust watermarking of audio signals."	Simple ANN with Discrete Cosine Transform, Concept of WSF(Watermark scaling Factor), and MMt(Minimum Masking Threshold are used.	Audio Signal	A3dsplash.wav (with sampling rate: 44.1KHz, and bit rate: 705kbps with mono) This signal is divided into 131 frames. Every frame containing with 1024 samples. Each frame is further segmented into 32 sub bands with 32 samples.	Produced Good Imperceptibility. MSE level obtained was -37dB

Table1: Literature Review of Neural Network Based Techniques for Audio Watermarking

3. Proposed Method

Audio Watermarking is carried out in two steps: i) Embedding and ii) Extraction. The host audio signal used is an audio file of 90 minutes duration, and the watermark is 180*180 binary image. We used backpropagation neural network for watermark embedding and extraction process along with DWT. Our framework takes two sources of input which are

i. The watermark image and,

ii. The audio file (Which has to be watermarked)

We used a sample lecture audio file as host signal for input. We had divided the 90 minutes audio files into small chunks whose size can be set from 10 seconds to 5 minutes to increase the processing speed. We used binary image of size 50 by 50 as a watermark. While processing, we take a 2D binary image (watermark).. Then this image is converted into binary using open cv2. For the second input i.e. audio signal we had taken a sample audio file of lecture recording. For pre-processing we used DWT as the standard and most efficient technique. DWT is applied to these chunks for obtaining the lower frequency components. In next step analysis of the chunks is done for each frame. Then, the custom built Neural Network embeds binary bits of watermark in the audio signal. For the decoding process the second neural network decodes back the watermark from the audio. DWT is applied again on the 10 second chunks and they are converged/merged back as well. We had used python programming and the required libraries for demonstrating the experimental results.

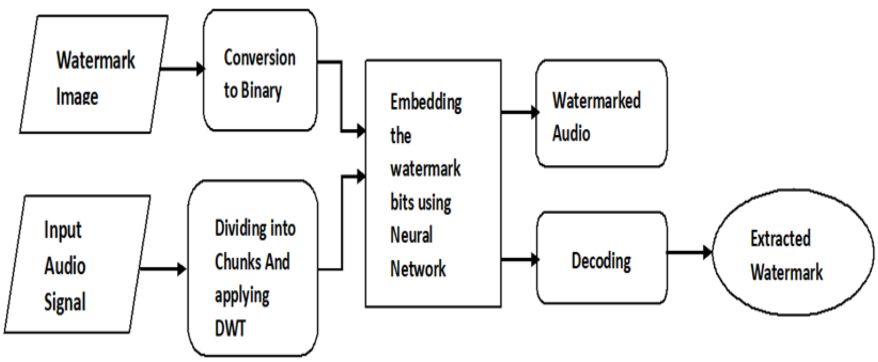


Fig 1 : Flow Diagram of Proposed Work

3.1. *Input Processing*

The first step in our system is to process the watermark image. For the proposed project we have taken the JPG image as sample watermark with the dimensions of 180 x 180 pixels. The JPG image is converted into binary format which is embedded in the audio signal.

First we convert it into Grayscale by setting a threshold of 128. Further, we had converted the image into binary of size 50 by 50 to use as the watermark. All these preprocessing steps are carried out using cv2 and numpy. Hence the watermark is hereby converted into a numpy array of 50 x 50 dimension. The audio file that has to be watermarked is first sampled at rate of 44.1 KHz. We down sample the given audio file to 16 KHz using an external library librosa. We divided the whole input audio file of 90 mins into chunks of length 10 seconds to 5 minutes for testing the experimental results. Hence we have the input watermark in the binary form and the audio down sampled into 16 KHz.

3.2. *Encoding using Neural Network*

In order for successful embedding of our watermark in the audio file, our Neural Network uses the lowest significant bits of the audio data for embedding the watermark inside it. In simple terms, the embedding process efficiently replaces the Lowest Significant bit of each byte in the audio file (in our case the binary watermark data). Logical AND and Logical OR operations are used here for hiding the secret data(watermark) inside audio file.

3.3. *Training the Network*

The neural network does some logical calculations and completes the watermarking process. The developed neural network has two input features and one output feature. The learning rate is set at 0.01. Sigmoid function is used as an Activation function. The model is trained at 100000 epochs.

3.4. *Decoding the Network*

After the encoding process, in order to decode the audio and extract the watermark from it the encoded audio is first converted into a byte array. The LSB is extracted from the byte array the extracted bytes is the matrix of pixel values of our watermark image. By plotting this matrix we can visualize the watermark image.

4. Results and Discussion

The proposed neural network does all the logical calculations and completes the watermarking process. The developed neural network has two input features and one output feature. The learning rate is set at 0.01. The Sigmoid function is used as an Activation function. The model is trained at 100000 epochs.

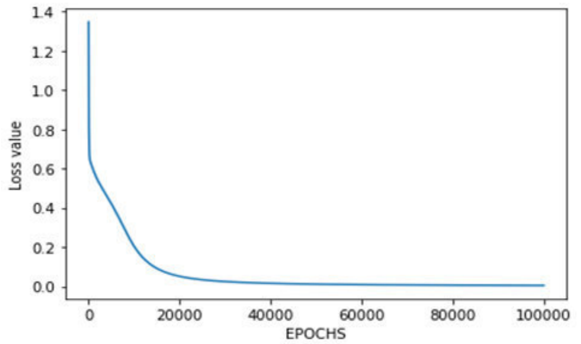


Fig 2 : Graph showing Training of the Model for 100000 epochs and reduction in loss

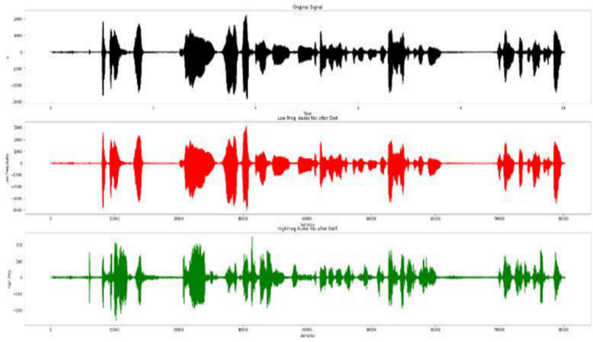


Fig 3: Audio Signal before applying DWT.(Color codes used : Black- Original Audio, Red-Low Frequency Component, Green-High Frequency Component.)

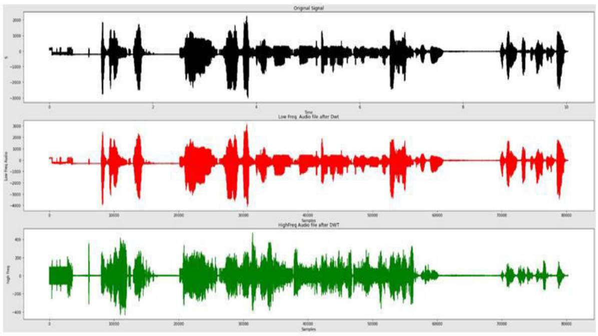


Fig 4: Audio Signal after applying DWT.(Color codes used : Black- Original Audio, Red-Low Frequency Component, Green-High Frequency Component.)

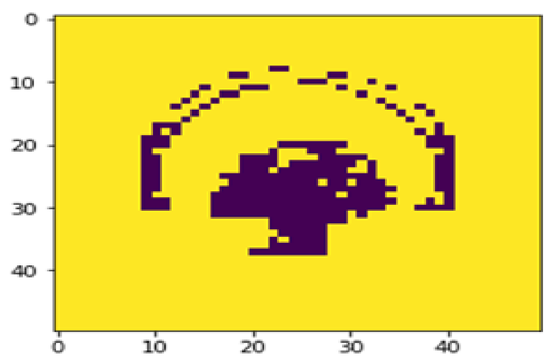


Fig 5: Sample Watermark Image

eISSN 1303-5150

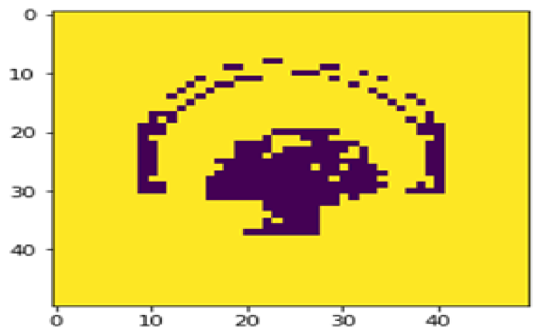


Fig 6: Extracted Watermark Image (After Decoding).

5. Conclusion

In this paper we discussed various neural network based approaches used for digital audio watermarking. Also, we designed a simple backpropagation neural network and used it along with DWT for embedding and extraction of watermark into audio signal. The experimental results shown that the watermark is successfully embedded into audio signal such that it is difficult for the attacker to remove it or tamper it. Bit Error Rate is found very low. There is a negligible amount of bit loss in the extracted watermark and we achieved robustness at higher level. We simulated the obtained output signal for robustness and imperceptibility and found that the watermark is robust against the common signal processing attacks such as noise attacks. The imperceptibility is also found to be good. We tested the results with 20 different persons to check the imperceptibility, none of the user was able to differentiate between original and watermarked audio. Thus, in this paper we discussed many algorithms based on neural networks and demonstrated a simple method using backpropagation network with DWT domain for digital audio watermarking. In future work, we will try to make the watermark more secure by using security algorithms. We will design a network such that, instead of using small chunks and then integrating them together at the output side, we will be able to collect a single audio file as the output.

Conflicts of Interest

"We the authors/ co-authors of this manuscript declare that there is no conflict of interest regarding the publication of this paper."

Funding Statement

All Authors and co-authors have seen and agree with the contents of manuscript and we have NO affiliations with or involvement in any organization or entity with any financial interest. No funding is provided for the work by any organization. Hence there is no financial interest to report.

References

K. Pavlović, S. Kovačević and I. Đurović, "Speech watermarking using Deep Neural Networks," 2020 28th Telecommunications Forum (TELFOR), 2020, pp. 1-4, doi: 10.1109/TELFOR51502.2020.9306626.

G. Wu and X. Zhou, "A Fast Audio Digital Watermark Method Based on Counter-Propagation Neural Networks," 2008 International Conference on Computer Science and Software Engineering, 2008, pp. 583-586, doi: 10.1109/CSSE.2008.675.

C. Maha, E. Maher and B. A. Chokri, "A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain," 2008 3rd International Symposium on Communications, Control and Signal Processing, 2008, pp. 1138-1143, doi: 10.1109/ISCCSP.2008.4537396.

Das, Arjon, and Xin Zhong. "A Deep Learning-based Audio-in-Image Watermarking Scheme." 2021 International Conference on Visual Communications and Image Processing (VCIP). IEEE, 2021.

M. Gupta and R. Rama Kishore, "A Survey of Watermarking Technique using Deep Neural Network Architecture," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 630-635, doi: 10.1109/ICCCIS51004.2021.9397226.

Yue Li, Hongxia Wang, Mauro Barni, "A survey of Deep Neural Network watermarking techniques, Neuro computing", Volume 461,2021,Pages 171-193,ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.07.051.

C. Chang, W. Shen and H. Wang, "Using Counter-propagation Neural Network for Robust Digital Audio Watermarking in DWT Domain," 2006 IEEE International Conference on Systems, Man and Cybernetics, 2006, pp. 1214-1219, doi: 10.1109/ICSMC.2006.384880.

J. Hu, X. Qiu and D. He, "Digital Audio Watermarking Algorithm based on Neural Networks," 2008 International Conference on Apperceiving Computing and Intelligence Analysis, 2008, pp. 89-92, doi: 10.1109/ICACIA.2008.4769978.

Chuan-Yu Chang and Sheng-Jyun Su, "A neural-network-based robust watermarking scheme," 2005 IEEE International Conference on Systems, Man and Cybernetics, 2005, pp. 2482-2487 Vol. 3, doi: 10.1109/ICSMC.2005.1571521.

J. Jiang and C. Pun, "Digital Watermarking Based on Patchwork and Radial Basis Neural Network," 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks, 2011, pp. 242-246, doi: 10.1109/CICSyN.2011.59.

Liu, Y., Guo, M., Zhang, J., Zhu, Y., & Xie, X. (2019). A Novel Two-stage Separable Deep Learning Framework for Practical Blind Watermarking. Proceedings of the

27th ACM International Conference on Multimedia - MM '19. doi:10.1145/3343031.3351025

Huiqin Wang, Li Mao and Keshan Xiu, "New Audio Embedding Technique Based on Neural Network," First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), 2006, pp. 459-462, doi: 10.1109/ICICIC.2006.479.

H. Yang, J. C. Patra and C. W. Chan, "An artificial neural network-based scheme for robust watermarking of audio signals," 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002, pp. I-1029-I-1032, doi: 10.1109/ICASSP.2002.5743970.

J. Bajpai and A. Kaur, "A literature survey - various audio watermarking techniques and their challenges," 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), 2016, pp. 451-457, doi: 10.1109/CONFLUENCE.2016.7508162.

S. Kirbiz and B. Gunsel, "Robust Audio Watermark Decoding by Supervised Learning," 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 2006, pp. V-V, doi: 10.1109/ICASSP.2006.1661387

C. Maha, E. Maher, K. Mohamed and B. A. Chokri, "DCT Based blind audio watermarking scheme," 2010 International Conference on Signal Processing and Multimedia Applications (SIGMAP), 2010, pp. 139-144.

Mohammed, Aree. (2016). Audio Watermarking Based on Hybrid Low and High Wavelet Frequencies. International Journal of Informatics and Communication Technology (IJ-ICT). 5. 28. 10.11591/ijict.v5i1.pp28-35.

O. M. S. Hassan et al., "An Efficient Robust Color Watermarking Algorithm Based on DWT, DCT, BFO and Implementation," 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET), 2021, pp. 90-95, doi: 10.1109/ICSET53708.2021.9612547.

A. Kaur and S. Singh, "A hybrid technique of cryptography and watermarking for data encryption and decryption," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 351-356, doi: 10.1109/PDGC.2016.7913175.

D. Stanescu, D. Borca, V. Groza and M. Stratulat, "A hybrid watermarking technique using singular value decomposition," 2008 IEEE International Workshop on Haptic Audio visual Environments and Games, 2008, pp. 166-170, doi: 10.1109/HAVE.2008.4685318.

M. Charfeddine, E. Mezghani, S. Masmoudi, C. B. Amar and H. Alhumyani, "Audio Watermarking for Security and Non-Security Applications," in IEEE Access, vol. 10, pp. 12654-12677, 2022, doi: 10.1109/ACCESS.2022.3145950.

Sarreshtedari, S., Akhaee, M., & Abbasfar, A. (2015). A Watermarking Method for Digital Speech Self-Recovery. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 1–1. doi:10.1109/taslp.2015.2456431

Archive.org.2022.Statistics_134_002_Spring_2010_UC_Berkeley_Concepts_of_Probabiltydirectorylisting].

Gupta, M., & Rama Kishore, R. (2021). A Survey of Watermarking Technique using Deep Neural Network Architecture. 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). doi:10.1109/icccis51004.2021.939

Kaur, A., Dutta, M. K., & Prinosil, J. (2018). General Regression Neural Network Based Audio Watermarking Algorithm Using Torus Automorphism. 2018 41st International Conference on Telecommunications and Signal Processing (TSP). doi:10.1109/tsp.2018.8441174

Jin, W., Dai, H., & Zhang, Z. (2009). Audio Watermarking Algorithm Robust to TSM Based on Counter Propagation Neural Network. 2009 2nd International Congress on Image and Signal Processing. doi:10.1109/cisp.2009.5303745

Rohit M Thanki. Advanced Techniques for Audio Watermarking. Signals and Communication Technology. Springer. 2020

Yong Xiang., Guang Hua., Bin Yan. Digital Audio Watermarking Fundamentals, Techniques and Challenges. Springer Briefs in Electrical and Computer Engineering. 2017.

H. O. Oh, J. W. Seok, J. W. Hong, and D.-H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '01), 2001, vol. 3, pp. 1341–1344, vol.3.

B.-S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread echo method for digital audio watermarking," IEEE Trans. Multimedia, vol. 7, no. 2, pp. 212–221, Apr. 2005.

O.-C. Chen and W.-C. Wu, "Highly robust, secure, and perceptual-quality echo hiding scheme," IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 3, pp. 629–638, Mar. 2008.

Y. Xiang, D. Peng, I. Natgunanathan, and W. Zhou, "Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking," IEEE Trans. Multimedia, vol. 13, no. 1, pp. 2–13, Feb. 2011.

P. Zhang, S.-Z. Xu, and H.-Z. Yang, "Robust audio watermarking based on extended improved spread spectrum with perceptual masking," Int. J. Fuzzy Syst., vol. 14, no. 2, pp. 289–295, 2012.

A. Valizadeh and J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," IEEE Trans. Inf. Forensics Security, vol. 6, no. 2, pp. 267–282, Jun. 2011.

I.-K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," IEEE Trans. Speech Audio Process., vol. 11, no. 4, pp. 381–386, Jul. 2003.

N. Kalantari, M. Akhaee, S. Ahadi, and H. Amindavar, "Robust multiplicative patchwork method for audio watermarking," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 6, pp. 1133–1141, Aug. 2009.

I. Natgunanathan, Y. Xiang, Y. Rong, W. Zhou, and S. Guo, "Robust patchwork-based embedding and decoding scheme for digital audio watermarking," IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 8, pp. 2232–2239, Oct. 2012.

Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and S. Nahavandi, "Patchwork-based audio watermarking method robust to de-synchronization attacks," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 22, no. 9, pp. 1413–1423, Sep. 2014.

Tsai, H.-H., Cheng, J.-S., & Yu, P.-T. (2003). Audio Watermarking Based on HAS and Neural Networks in DCT Domain. EURASIP Journal on Advances in Signal Processing, 2003(3). doi:10.1155/s1110865703208027 Physical Layer (PHY) Specification, IEEE Std. 12(11) (1997) 260-280.

M. Charfeddine, E. Mezghani, S. Masmoudi, C. B. Amar and H. Alhumyani, "Audio Watermarking for Security and Non-Security Applications," in IEEE Access, vol. 10, pp. 12654-12677, 2022, Doi: 10.1109/ACCESS.2022.3145950.

High-Confidence Computing 3 (2023) 100153

Contents lists available at ScienceDirect

High-Confidence Computing

journal homepage: www.sciencedirect.com/journal/high-confidence-computing

Research Article

An effective digital audio watermarking using a deep convolutional neural network with a search location optimization algorithm for improvement in Robustness and Imperceptibility

Abhijit J. Patil [a,*], Ramesh Shelke [b]

[a] Computer Engineering, Pacific Academy of Higher Education and Research University, Udaipur 313003, India
[b] Electronics and Telecommunications, University of Mumbai, Mumbai 400032, India

ARTICLE INFO

Article history:
Received 4 April 2023
Revised 13 June 2023
Accepted 5 July 2023

Keywords:
Search location optimization algorithm
Deep convolutional neural network
DWT
Robustness
Imperceptibility

ABSTRACT

Watermarking is the advanced technology utilized to secure digital data by integrating ownership or copyright protection. Most of the traditional extracting processes in audio watermarking have some restrictions due to low reliability to various attacks. Hence, a deep learning-based audio watermarking system is proposed in this research to overcome the restriction in the traditional methods. The implication of the research relies on enhancing the performance of the watermarking system using the Discrete Wavelet Transform (DWT) and the optimized deep learning technique. The selection of optimal embedding location is the research contribution that is carried out by the deep convolutional neural network (DCNN). The hyperparameter tuning is performed by the so-called search location optimization, which minimizes the errors in the classifier. The experimental result reveals that the proposed digital audio watermarking system provides better robustness and performance in terms of Bit Error Rate (BER), Mean Square Error (MSE), and Signal-to-noise ratio. The BER, MSE, and SNR of the proposed audio watermarking model without the noise are 0.082, 0.099, and 45.363 respectively, which is found to be better performance than the existing watermarking models.

1. Introduction

The ever-increasing advancements in digital innovation lead to positive impacts on communities and societies [1]. The mysterious users will perform modifications in the digital content, which impacts the copyright properties due to the availability of internet data [2]. For instance, the speech signal of the individual will tamper with and be duplicated by criminals or robbers to do some illegal activities, which demonstrates the significance of ownership or copyright protection [1]. There are numerous attacks like cropping, geometric, and filtering utilized by the attackers that affect the ownership of the digital contents [2,3]. There are raising concerns regarding illegal digital content due to the availability of filters that tampers with digital content [1]. These raising concerns result in the exploration of information security, such as duplication, ownership protection, and authentication [4]. Watermarking is an advanced innovation that prevents the audio contents from the aforementioned attacks to recognize unwanted modifications [2,5]. The significance of digital watermarking has increased the various research studies. Digital watermarking is used in many different applications, including copy control, broadcast monitoring, transaction tracking, and owner recognition. In general, maximum payloads, little degradation, and a small false positive rate are requirements for successful and reliable digital watermarking [3].

Digital watermarking is characterized as the process that protects the delicate digital media information that conserves the ownership of the data. The watermark is acquired through a variety of processes and is resistant to a variety of assaults and removal methods [6]. Based on the characteristics of the carrier or cover signals, digital watermarking is divided into three categories, including, audio, and video watermarking [7]. Since people's listening power is more sensitive than human sight power, it is discovered that the invisibility in audio watermarking is a more challenging procedure than the other two watermarking [8]. Hence, more researchers focus on audio watermarking as it has a high capability of watermarking audio signals [9,10]. The four conditions are needed to be satisfied for tampering recognition in audio or speech signals [11]. The inaudibility is the first property to be satisfied in the audio watermarking, which states that the secret messages are not perceptible to normal human ears. The idea is that human ears should not be able to identify the difference between the initial version and the watermarked version. The second characteristic is blindness, which specifies that just the watermarked information is necessary for extracting

* Corresponding author.
E-mail address: abhijitjp774@gmail.com (A.J. Patil).

https://doi.org/10.1016/j.hcc.2023.100153

the watermarks and that the host signal is not necessary for the retrieval process. The robustness is the third feature that states that the embedded signal resists malicious attacks. The fragility is the last feature that states the embedded watermark needs to be sensitive to malicious or tampering attacks [1].

There exist various techniques to hide confidential data in carrier audio signals like quantization, masking, spread spectrum, and so on [7]. The low clarity and low extents are the main limitations of the aforementioned hiding techniques that limit real-time utilization. Further, the system robustness at high capacities is not evaluated under real-time execution of the watermarking system [9,12]. The traditional methods tend to determine the optimal inter-dependence between the attributes of audio watermarking to provide embedded in a specific domain, that determines certain regions or places and merging these locations, with computational overload and simplicity. Then higher dependence on the decision threshold and the pre-determined rules negatively impacts the dictating and diagnostic potential. Some intellectual methods based on machine learning classifiers are utilized [9]. The machine learning classifiers like neural networks are utilized in the last few decades to embed the watermark in the carrier signal [13,14]. These existing methods ensure high perpetual transparency and good robustness that mitigates the complexity of machine learning techniques [7]. However, these methods show some errors and inappropriate watermark extraction in certain data due to the achievements of high watermark capacity [9].

This research concentrates to develop robust deep learning-based optimization for the audio watermarking system. The encoding and extraction steps are carried out independently by the suggested watermarking models. The DCNN is used to identify the best embedding position that improves system performance without degrading the original data's quality. Utilizing the suggested search site optimization technique, the classifier's parameters are trained to their maximum potential. The DWT has also been used to encode secret data in digital sound signals. In the extraction phase, the inverse DWT extracts the hidden data from the carrier signal. The contribution of the research lies in

- **Search location optimization:** Search location optimization is the meta-heuristic intellectual algorithm, which integrates the characteristic features of the searching behavior of the creatures with the exploring behavior of human beings. The searching behavior of a creature is integrated with the locating behavior of humans by improving the speed of the searching process which consists of three phases initializing the position and head angles; the region, direction, and period are explored in this phase. In the promoter scanning phase, the three degrees are enhanced using the global best location strategy of humans that helps in obtaining the optimal solution. At last in the feeder selection phase, the random walk is performed to fetch the available resources randomly.
- **Search location-based DCNN for selection of optimal embedding location:** The watermarking is performed using the optimized DCNN for which the internal model parameters are tuned by proposed search location-based optimization to improve the system performance.

The organization of the research article: Section 2 illustrates the need for the digital watermarking model with a review of existing audio watermarking models. Section 3 demonstrates the proposed digital watermarking models, mathematical models, and algorithmic functions. The analysis of the results is presented in Section 4, and Section 5 concludes the paper.

2. Motivation

The advantages of the deep learning techniques in signal processing such as reducing computational and time complexity by extracting the significant features in the signal motivate to development of an audio deep learning-based watermarking model. This section provides a brief description of the existing watermarking system with its drawbacks.

2.1. Literature review

The existing deep learning-based techniques for watermarking the signals are reviewed as follows, Farah Deeb et al. [6] presented a deep neural network classifier for watermarking digital data and also performing effective ownership verification. The developed technique is more robust to the various attacks, but the model attains unreliable performance through the various standards with a low accuracy level. Seyed Mostafa Pourhashemi et al. [9] performed the embedding process using the combination of DWT and the ensembled intelligent extraction approach. The watermarking capacity of the developed model is high but the performance still needs to be improved by introducing the optimization algorithm, which identifies the optimal frames for embedding. Kasorngalajit et al. [1] employed a parameter estimation model using the CNN classifier, which accurately and rapidly chooses the spectrum to be modified. The computational time of the developed method is minimized and is mostly applicable to the real-time data, however, the additional training dataset is required for the CNN. Preeti Garg and R. Rama Kishore [2] ensure the security and embedding by the two-level DWT and cosine transform with the different cover images and the encrypted image. The attained PSNR value against the various attacks is more than 40 dB, although an optimization technique is absent for the location embedding. KasornGalajit et al. [4] incorporate a convolutional neural network with the singular spectrum analysis for measuring the parameters. The computational time for embedding the watermark into an audio signal by the modification of the singular spectrum is low, on the contrary, the developed technique is fragile to a few attacks. Arashdeep Kaur et al. [7] extract the watermark more effectively by developing the general regression neural network and the singular value decomposition along with audio signal extraction. Depending on the obviousness, payload, and sturdiness, the efficiency of the developed model is high, but the SNR value of the developed method is low than the other compared methods. Khaled M. Abdelwahab et al. [14] presented an SVD-based audio watermarking model in the FRT domain with segment-dependent execution, additionally the original and the obtained watermark have a good correlation coefficient. The detection correlation is high for the available severe attacks until now, there are slight variations between the original audio signal and the watermarked signal. Huda Karajeh et al. [3] introduced a hybridized model which integrates the Schur decomposition hybrid method with the DWT for embedding the watermark image foreground bits in the diagonal coefficients' least significant bit. Depending on the quietness, sturdiness, and payload capability, the performance is high, but this method's primary flaw is that it only embeds a small-sized watermark image.

Several layers are present in the proposed model with few parameters optimized using SLOA that performs as a significant method with consuming less time. The CNN with Re-layer eliminates every negative value to zero reducing the values of the parameter. The tampered signal was fetched using two-level DWT after the arrival of the watermark-embedded audio signal. The usage of DWT for decomposing safeguards the embedded signal from threads and increases reliability.

2.2. Challenges

- Most of the audio watermarking model relies on the Deep learning classifier in the embedding process. However, the training of the classifier is complex due to numerous layers and parameters [6].
- Some of the existing methods are delicate to some non-malignant signal processing, such as echo addition and pitch shifting with low degrees. Hence, an effective model is required for the effective reorganization of signal tampering [9].
- The existing SVD-based audio watermarking process is time-consuming and complex due to the presence of differential evolution (DE) optimizations [4]. Hence, an algorithm that consumes less time is required to train the classifier.
- The conventional audio watermarking model is found to be effective in embedding the secret message in the carrier signal. Yet, the low transparencies and low opacities are considered the major challenge in the conventional methods [7].
- The detecting and diagnostic ability of the existing digital watermarking models is found to be low as the embedded model depends on the predefined rules and decision threshold.

3. Proposed digital audio watermarking based on search location optimization

This section elucidates the significant steps involved in the digital audio watermarking system and Fig. 1 demonstrates the watermarking models. This research focuses on device watermarking techniques that enhance imperceptibility, security, and robustness. The audio watermarking is carried out in two different phases as embedding phase and the extracting phase. Initially, the audio signals that need to be watermarked are divided into different signal blocks. Then the optimal signal from the block is determined by using the proposed optimization known as search location optimization, which is devised by hybridizing the locating characteristics of humans [15] and the searching behavior of a creature [16]. After finding the smallest block in the audio signal to embed the watermark in, DWT decomposition embeds the watermark in the audio signals. Using 2-level Inverse DWT methods, the hidden message is acquired or retrieved via the watermarked audio file during the extraction phase.

3.1. Read the input data

The brain tumor image from the multi-modal brain tumor segmentation [17] data from the year 2020 is considered the secret message to embed in the audio signal. The multimodal MRI scans of lower-grade glioma (LGG) and glioblastoma with a pathologically confined diagnosis are obtained from these datasets. This dataset is made available after the pre-processing techniques, such as skull-stripping and pixel interpolation.

The audio signal with a Bit-rate of 1411 Kbps is utilized as the carrier signal to embed the secret message (brain tumor image). In this research, the audio signal with a duration of 00:02:02 s and stored as a WAV file is utilized as the carrier signal. The size of the carrier audio file is found to be 20.5 MB.

3.2. Embedding phase

The watermarked digital signal is generated in the embedding phase by considering both the host signal and the input watermarks. The steps involved in the embedding phases such as audio signal block formation and secret data embedding are widely elaborated in the following sections

Table 1
DCNN layer information.

Layer type	Output shape	Parameters
Conv2d	(None, 181, 1, 32)	544
Leaky_re_lu	(None, 181, 1, 32)	0
Max_pooling2d	(None, 91, 1, 32)	0
Conv2d_1	(None, 91, 1, 64)	18 496
Leaky_re_lu_1	(None, 91, 1, 64)	0
Max_pooling2d_1	(None, 91, 1, 64)	0
Conv2d_2	(None, 91, 1, 128)	73 856
Leaky_re_lu_2	(None, 91, 1, 128)	0
Max_pooling2d_2	(None, 91, 1, 128)	0
Flatten	(None, 11 648)	0
Dense	(None, 128)	1 491 072
Leaky_re_lu_3	(None, 128)	0
Dense_1	(None, 3)	387

3.2.1. Formation of audio signal blocks

Let us consider the audio signal as S_{audio} acts as the cover audio signal in which the secret medical image is embedded. To initialize the embedding process the carrier audio signal is exposed to a random interval split-up, where the cover signal is divided into i number of frames.

$$S_{audio} = S_{audio,i}\,(t)\,;\,(1 \le i \le S_{tot})\,(1 \le t \le \mathrm{T}_{tot}) \tag{1}$$

where S_{tot} denotes the total signal and T_{tot} represents the total time interval. The ith signal undergoes a random interval split-up for the block formation and it is represented as

$$S_{audio,i}\,(t) = \left\{block_1^i, block_2^i, \ldots\ldots\ldots block_{b_{tot}}^i\right\}(i \le j \le b_{tot}) \tag{2}$$

where b_{tot} represents the total number of the signal block and the Wavelet transform is utilized with two-level decomposition.

3.2.2. Optimal block selection using search location optimization

From the generated signal blocks the optimal blocks for embedding the secret medical message or image are selected through the search location optimization on DCNN. The DCNN is commonly used in signal processing that effectively selects the optimal block to embed the secret data. The architecture of the DCNN classifier is depicted in Fig. 2 and Table 1.

The proposed model comprised 3 convolutional layers, 4 leaky ReLu layers, 3 max-pooling layers, 1 flatten layer, and 2 dense layers utilized. The fundamental layers of the DCNN known as convolutional layers are utilized to generate the feature vector. The first convolutional layer generates the feature vector of the output of size (181×1) with a batch size of 32. The subsequent layer is the ReLu layer that utilizes the sigmoid function that scales the CNN and generates the output of (181×1) Then the max-pooling layer is used to reduce the spatial size of the input image and these three layers are continued until it obtains the batch size of 128 and the single long feature of size 11 648 is obtained by the flattening layer. The dense layer is utilized to categorize the input to 128 class output and it is scaled by the ReLu layer. Finally, the dense layer is utilized to obtain 3 class outputs with batch size 387.

3.3. Proposed Search Location Optimization Algorithm (SLOA)

A novel SLOA algorithm invokes the searching behavior of a creature [16] and is designed by improving the searching speed of a creature using the locating characteristics [15] obtained from the locator to promote the global optimal solution. Mostly the undetermined optima in the optimization problem can be considered as the open patch-up that is dispersed randomly over the exploration area. As a result, the followers in the group move through the exploration area in pursuit of the patch-up, and is believed that the important mutational characters of the

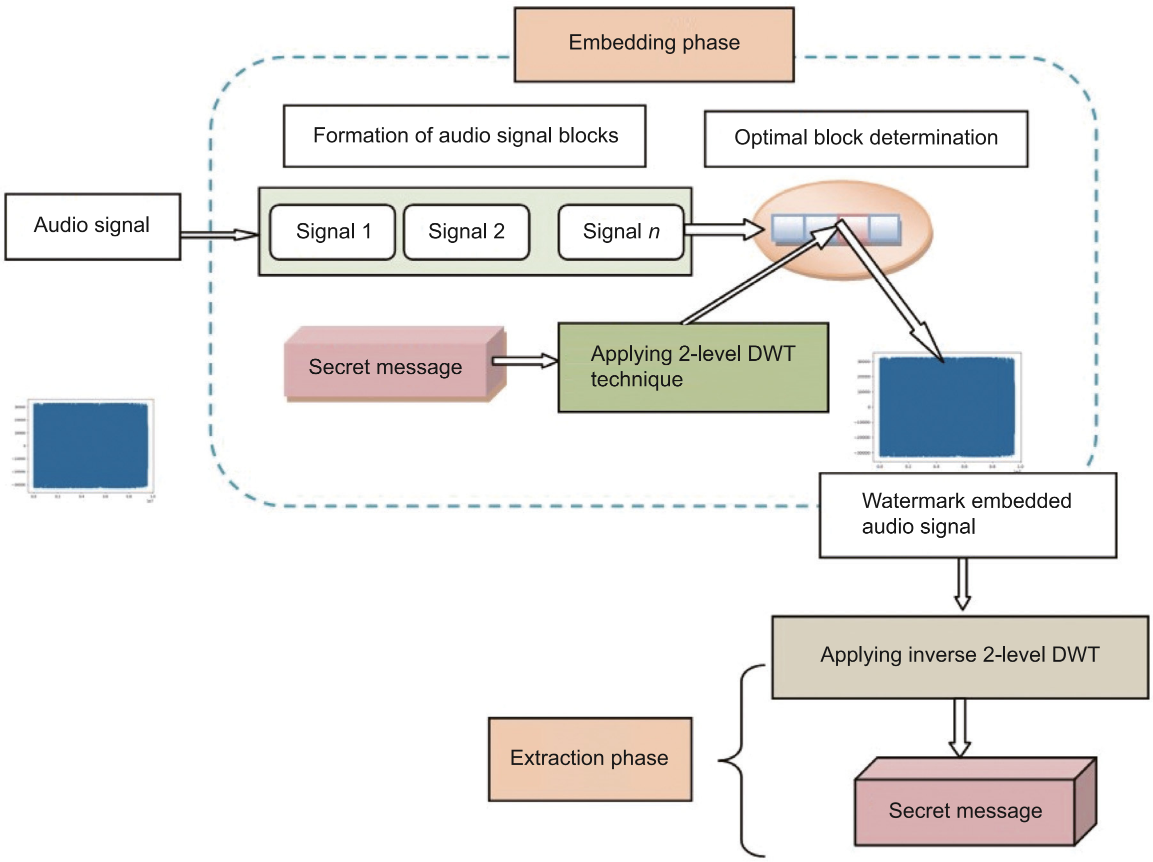


Fig. 1. Flow diagram of proposed digital audio watermarking based on search location optimization.

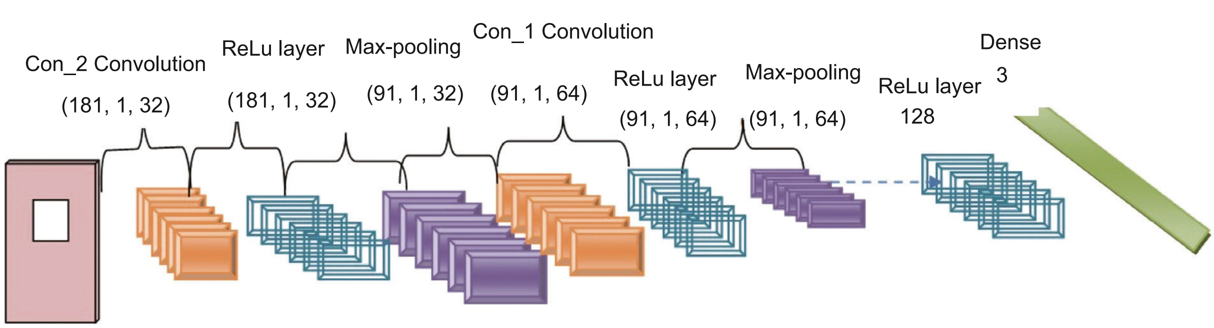


Fig. 2. DCNNLayer architecture.

promoters as well as the feeders are the same. Consequently, they are also able to move between the two characters. The follower in the group, who is positioned in the best promising location and provides the best fitness value is selected as the promoter at every iteration, then the promoter comes to a standstill position and searches the surroundings for optimum. An essential part of exploring location is scanning, which is a series of techniques used by animals to move their sense receptors and occasionally their followers to gather information from their surroundings. In the SLOA strategy, the promoter uses eyesight, which is the primary scanning technique preferred by most classes of the creature. Various species employ retinas with an adjustable spatial solution to encode a wide range of views for visual scanning, and they subsequently utilize quick eye movements to point the extreme resolution area in the field of view in the direction of the prospective focused position. Thus, the scanning strategy plays a major role in the part of survival, which is enhanced by integrating the locating characteristics of humans to maximize the fast scanning and minimize the various iterations, and computational time.

Inspiration

Creatures typically move towards hunting and exploring for food, depending on how they search for food, creatures are divided into promoters and feeders. Three types of followers make up a group in the SLOA strategy, that is promoters, feeders, and dispersed followers. Among these, the dispersed followers make the exploration space more random, during a particular period, only one promoter is considered for every exploration, and the remaining followers are assumed as the feeders, and dispersed to make the computation easier. The most interesting fact is the presuming that all source feeders may join the source discovered by the promoter. The common creature's exploration characteristics and the communal scanning technique serve as the basis for the SLOA strategy, which is more applicable to solving the optimization of continuous function problems. The most common creature's exploring behavior served as inspiration for the SLOA algorithm, while the promoter of the SLOA strategy and the best particle globally are relatively comparable, the promoters conduct the performance by remembering the head angle, which is quite different from the feeders and dispersed followers. The mathematical modeling for the SLOA algorithm based on the exploration direction angle is described in the following section.

3.3.1. Mathematical modeling of search location optimization algorithm

(i) Initialize the positions and head angles

For the u dimensional exploring region, the jth follower at the pth exploring period has a present location as $Y_j^p \in \mathbb{R}^u$, and the head angles are initiated as follows:

$$\theta_j^p = \left(\theta_{j_1}^p, \ldots, \theta_{j_{(u-1)}}^p\right) \in \mathbb{R}^{u-1} \tag{3}$$

The jth follower explores direction with the unit vector and is measured using the follower's head angle θ_j^p through the cartesian coordinate transformation from the polar form, which is expressed as follows:

$$\mathrm{M}_j^p\left(\theta_j^p\right) = \left(m_{j_1}^p, \ldots, m_{j_u}^p\right) \in \mathbb{R}^u \tag{4}$$

$$m_{j_1}^p = \prod_{k=1}^{u-1} \cos\left(\theta_{j_k}^p\right) \tag{5}$$

$$m_{j_l}^p = \sin\left(\theta_{i_{(l-1)}}^p\right) \cdot \prod_{k=l}^{u-1} \cos\left(\theta_{j_k}^p\right) \quad (l = 2, \ldots\ldots, u-1) \tag{6}$$

$$m_{j_u}^p = \sin\left(\theta_{j_{(u-1)}}^p\right) \tag{7}$$

Let us assume that the pth exploration period, the attained exploration direction M by the follower's head angle $\theta_j^p = (\pi/3, \pi/4)$ utilizing Eq. (7) is expressed as

$$\mathrm{M}_j^p = \left(1/2, \sqrt{6}/4, \sqrt{2}/2\right) \tag{8}$$

(ii) Promoter scanning

Initially, the promoter randomly set three points in the exploring field for the process of scanning, which includes zero-degree, right-side hypercube, and left-side hypercube. The three points of the promoter Y_v are expressed in Eqs. (9), (10), and (11).

$$Y_x = Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p\right) \tag{9}$$

$$Y_R = Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p + R_2\varphi_{\max}/2\right) \tag{10}$$

$$Y_h = Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p - R_2\varphi_{\max}/2\right) \tag{11}$$

where the promoter is represented as Y for the zero degrees as Y_x, the right side as Y_R, and the left side as Y_h. R_1 and R_2 is the representation of the random numbers, in which the R_1 is in the range of $\mathbb{R}^1$ with the standard deviation as 1 and mean as 0. The maximum pursuit angle is represented as $\varphi_{\max}$, R_2 is in the range of $\mathbb{R}^{u-1}$ with $(0, 1)$ and the maximum search space of the promoter and the target is represented as $h_{\max}$.

The human moving towards the global best location by holding the optimal strategy of the promoters, which is expressed as follows:

$$Y_v^{p+1} = Y_v^p + C * \left(R * G - Y_v^p\right) \tag{12}$$

Utilizing Eq. (12), the promoter scanning phase is modified as the standardized form by integrating the excellent locating characteristics of the human in the Eqs. (9), (10), and (11), which are expressed as follows:

$$Y_x = \frac{1}{2}\left\{Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p\right) + Y_v^p + C * \left(R * G - Y_v^p\right)\right\} \tag{13}$$

$$Y_x = \frac{1}{2}\left\{Y_v^p\left(2 - C\right) + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p\right) + C * R * G\right\} \tag{14}$$

where the global best location is denoted as G with the constant term as C for the promoter scanning at zero degrees. Similarly, the promoter scanning phase depending on the right side of the hypercube is modified by the constant parameters as α and γ are as follows:

$$Y_R = \alpha\left\{Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p + R_2\varphi_{\max}/2\right)\right\} + \gamma\left\{Y_v^p + C * \left(R * G - Y_v^p\right)\right\} \tag{15}$$

$$Y_R = \alpha Y_v^p\left(\alpha + \gamma - C \cdot \gamma\right) + \alpha \cdot R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p + R_2\varphi_{\max}/2\right) + \gamma \cdot C * R * G \tag{16}$$

The left side hypercube of the promoter scanning process is enhanced by the global as well as the personal best location of the human, which is formulated as

$$Y_h = \frac{1}{2}\left\{Y_v^p + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p - R_2\varphi_{\max}/2\right) + Y_v^p + C * \left(R * G - Y_v^p\right) + D * \left(R * \mathrm{P} - Y_v^p\right)\right\} \tag{17}$$

$$Y_h = \frac{1}{2}\left\{Y_v^p\left(2 - C - D\right) + R_1 h_{\max} \mathrm{M}_v^p\left(\theta^p - R_2\varphi_{\max}/2\right) + \left(C * R * G\right) + \left(D * R * \mathrm{P}\right)\right\} \tag{18}$$

where the constant terms are represented as $C\,and\,D$ with the global best location as G and the personal best location of the promoter is represented as P. Thus, the global best location is identified more accurately in the promoter scanning phase along with the personal best solution at all angles.

(iii) Feeder selection

The promoter identifies the optimal point with the suitable fitness function and then migrates to any other point if they identify a better source than the previous one. Otherwise, the promoter remains in the existing point by rotating the head to a newly generated angle.

$$\theta^{p+1} = \theta^p + R_2\beta_{\max} \tag{19}$$

where the maximum rotating angle of the promoter is represented as $\beta_{\max} \in \mathbb{R}^1$. After the termination of i iterations, the promoters ever find the optimal region, they rotate their head back to an angle of 0 degree, which is expressed as

$$\theta^{p+1} = \theta^p \tag{20}$$

where the term $i \in \mathbb{R}^1$ is assumed as constant.

During every iteration, a set of followers are assumed as feeders, and they will keep tracking the promoters to reach the sources, which are determined by the promoters. One of the three significant phases involved in the SLOA strategy is mimicking the region to identify the exploring region of the promoter. The stalking phase involves stalking other creatures without revealing any other searching strategy, and in the stealing phase, the feeders acquired the sources directly from the promoters. At the pth iteration, the characteristic of mimicking region of the jth follower can be designed as a random movement towards the promoter, and is expressed as

$$Y_j^{p+1} = Y_j^p + R_3 \circ \left(Y_v^p - Y_j^p\right) \tag{21}$$

Random walks are considered the most effective exploration technique for the available sources in a random manner, at the pth iteration, the randomized head angle is generated as θ_j using $\theta^{p+1} = \theta^p + R_2\beta_{\max}$ by choosing a random space.

$$h_j = i \cdot R_1 h_{\max} \tag{22}$$

and change to a different point as

$$Y_j^{p+1} = Y_j^p + h_j \mathrm{M}_j^p\left(\theta^{p+1}\right) \tag{23}$$

The SLOA algorithm is described in Algorithm 1 as follows:

3.4. Secret data embedding using DWT

The DWT is utilized to embed the audio signals as secret data in the data transaction process as it inherits the fast-computational capability and it provides complicated information about the data. The DWT is used to accomplish the action of embedding and extracting information employing the wavelet coefficient once the optical location has been chosen through search location optimization.

DWT in the embedding process

The original audio signal with the size $a \times b$ is enclosed by the secret signal, which is represented as δec with the size $c \times d$. At the initial level of the embedding process, the audio signals band information is collected using the wavelet transform. The wavelet transforms are gathered based on two levels of band High and Low, which are subcategorized into High-High (HH), High-Low (HL), Low-High (LH), and Low-Low (LL). These sub-bands provide the edge information of the image.

The initial level sub-bands are denoted as

$$Cov_{sig} = \left\{ \mathrm{B}^{sub}_{\mathrm{HH}}, \mathrm{B}^{sub}_{\mathrm{H}L}, \mathrm{B}^{sub}_{L\mathrm{H}}, \mathrm{B}^{sub}_{LL} \right\} \tag{24}$$

where Cov_{sig} illustrates the cover signal and $\mathrm{B}^{sub}_{\mathrm{HH}}, \mathrm{B}^{sub}_{\mathrm{H}L}, \mathrm{B}^{sub}_{L\mathrm{H}}, \mathrm{B}^{sub}_{LL}$ denotes the sub-band in the image with $\mathrm{HH}, \mathrm{H}L, L\mathrm{H}, LL$co-efficient. The sub-band dimensions are represented as $\left[\frac{c}{2} \times \frac{d}{2}\right]$. These sub-bands are processed and it is subjected to the second level to generate 16 sub-bands represented by

$$Cov^{sig}_{\mathrm{HH}} = \left\{ \mathrm{B}_{\mathrm{HH}1}, \mathrm{B}_{\mathrm{H}L1}, \mathrm{B}_{L\mathrm{H}1}, \mathrm{B}_{LL1} \right\} \tag{25}$$

$$Cov^{sig}_{\mathrm{H}L} = \left\{ \mathrm{B}_{\mathrm{HH}2}, \mathrm{B}_{\mathrm{H}L2}, \mathrm{B}_{L\mathrm{H}2}, \mathrm{B}_{LL2} \right\} \tag{26}$$

$$Cov^{sig}_{L\mathrm{H}} = \left\{ \mathrm{B}_{\mathrm{HH}3}, \mathrm{B}_{\mathrm{H}L3}, \mathrm{B}_{L\mathrm{H}3}, \mathrm{B}_{LL3} \right\} \tag{27}$$

Algorithm 1: The proposed modified Douglas–Peucker (mDP) algorithm

Input
A list $\mathfrak{S}$ containing the digitized point of the obstacle.
Threshold value ε for maximum dissimilarity tolerance.
Output
A list of points $\mathfrak{V}$ represents the final result of polygonal approximation of the obstacle.
Step 1: Line
$\mathfrak{V}$={}; `// Create the polygonal approximation point list`
Find the two points $\{\mathscr{P}_i, \mathscr{P}_j\}$ in $\mathfrak{S}$ with the maximum distance from each other;
Separate the curvilinear obstacle into two curves, C_1 and C_2
Step 2: Maximum deviation
Find deviation of points in C_1 and C_2 from the line $(\mathscr{P}_i, \mathscr{P}_j)$;
Find maximum deviation d_{max} as d_m and d_n and points $\mathscr{P}_m$ and $\mathscr{P}_n$ corresponding to d_m and d_n;
Step 3: Termination/recursion condition
if $d_{max} \leq \varepsilon$ **then**
  $\mathfrak{V} = \{\mathfrak{V}, \mathscr{P}_i, \mathscr{P}_j\}$; `// Add new point in the list`
else
  $\mathfrak{V} = \{\mathfrak{V}, \text{DP\_max}(\mathscr{P}_i, \mathscr{P}_m)\}$; `// Add new point in the list`
  $\mathfrak{V} = \{\mathfrak{V}, \text{DP\_max}(\mathscr{P}_m, \mathscr{P}_j)\}$;
  $\mathfrak{V} = \{\mathfrak{V}, \text{DP\_max}(\mathscr{P}_i, \mathscr{P}_n)\}$;
  $\mathfrak{V} = \{\mathfrak{V}, \text{DP\_max}(\mathscr{P}_n, \mathscr{P}_j)\}$;
end

The sub-band dimension is obtained by $\left[\frac{a}{4} \times \frac{b}{4}\right]$ and using this wavelet co-efficient B_{HH} and B_{LL} the data embedding process is performed. The embedding process is represented as

$$\mathrm{E}_{tl}\,(\mathrm{A}, \mathrm{X}) = \omega_{tl}\,(\mathrm{A}, \mathrm{X}) + \mathrm{E}_{str} * \mathrm{Sec}^{sig}_{mbits}\,(\mathrm{A}, \mathrm{X}) \tag{28}$$

where $\mathrm{E}_{tl}\,(\mathrm{A}, \mathrm{X})$ represents the signal embedding process that illustrates the watermarked audio signals. *tl* denotes the total wavelet bands and *mbits* denotes the total message bits that range from 1 to 8. The Sec^{sig} denotes the secret message signals and its wavelet band is denoted as $\omega_{tl}\,(\mathrm{A}, \mathrm{X})$ and the variable that denotes the embedding strength is denoted as E_{str}. The embedding process is depicted in Fig. 3.

3.5. Extraction phase

In the extraction process, the embedded watermark is obtained from the watermark audio signal by the inverse wavelet transform. The process involved in the extraction phase is briefly described in this section.

3.5.1. Inverse wavelet transform

The original audio signal is recreated from the embedded audio data with the secret message using the inverse wavelet transform. Similarly to the wavelet transform, the inverse wavelet transform breaks the hidden message into two phases. The initial level decomposition is represented as

$$IDWT\,(\mathrm{E}_{tl}\,(\mathrm{A}, \mathrm{X})) = \left\{ \mathrm{B}^*_{LL}, \mathrm{B}^*_{\mathrm{H}L}, \mathrm{B}^*_{L\mathrm{H}}, \mathrm{B}^*_{\mathrm{HH}} \right\} \tag{29}$$

The inverse wavelet transforms comprised of the information about sub-bands and it is represented as $IDWT\,(\mathrm{E}_{tl}\,(\mathrm{A}, \mathrm{X}))$. The $\mathrm{E}_{tl}\,(\mathrm{A}, \mathrm{X})$ represents the second-level decomposition of the audio signals.

3.5.2. Extraction of data

The extraction phase is the contradictory process of the embedding process from which the original audio signal is recovered from the cover signals. The extraction process is carried out based on information like wavelet embedded image, optimal point location, and the cover signal. The secret data is thus extracted by the inverse DWT (IDWT). The secret data extraction is carried out in two different decomposition levels demonstrated as

$$DWT - 1\left(\mathrm{E}^*\right) = \mathrm{B}^*_{LL}, \mathrm{B}^*_{\mathrm{H}L}, \mathrm{B}^*_{L\mathrm{H}}, \mathrm{B}^*_{\mathrm{HH}} \tag{30}$$

where $DWT - 1$ denotes the first level decomposition and E illustrates the embedded signals

$$DWT\left(\mathrm{B}^{inv}_{\mathrm{HH}}\right) = \mathrm{B}^*_{\mathrm{HH}1}, \mathrm{B}^*_{\mathrm{H}L1}, \mathrm{B}^*_{L\mathrm{H}1}, \mathrm{B}^*_{LL1} \tag{31}$$

$$DWT\left(\mathrm{B}^{inv}_{\mathrm{H}L}\right) = \mathrm{B}^*_{\mathrm{HH}2}, \mathrm{B}^*_{\mathrm{H}L2}, \mathrm{B}^*_{L\mathrm{H}2}, \mathrm{B}^*_{LL2} \tag{32}$$

$$DWT\left(\mathrm{B}^{inv}_{L\mathrm{H}}\right) = \mathrm{B}^*_{\mathrm{HH}3}, \mathrm{B}^*_{\mathrm{H}L3}, \mathrm{B}^*_{L\mathrm{H}3}, \mathrm{B}^*_{LL3} \tag{33}$$

$$DWT\left(\mathrm{B}^{inv}_{LL}\right) = \mathrm{B}^*_{\mathrm{HH}4}, \mathrm{B}^*_{\mathrm{H}L4}, \mathrm{B}^*_{L\mathrm{H}4}, \mathrm{B}^*_{LL4} \tag{34}$$

The optimal location is determined and the secret data is obtained from the cover audio signal and the data extraction is mathematically determined as

$$E_{\mathrm{sec}}\,(\mathrm{A}, \mathrm{X}) = \omega^*_{tl}\,(\mathrm{A}, \mathrm{X}) - \omega^*_{LL} \tag{35}$$

4. Results and discussion

The performance of the proposed optimization-based digital audio watermarking is enumerated with the experimental setup, dataset description, and comparative analysis.

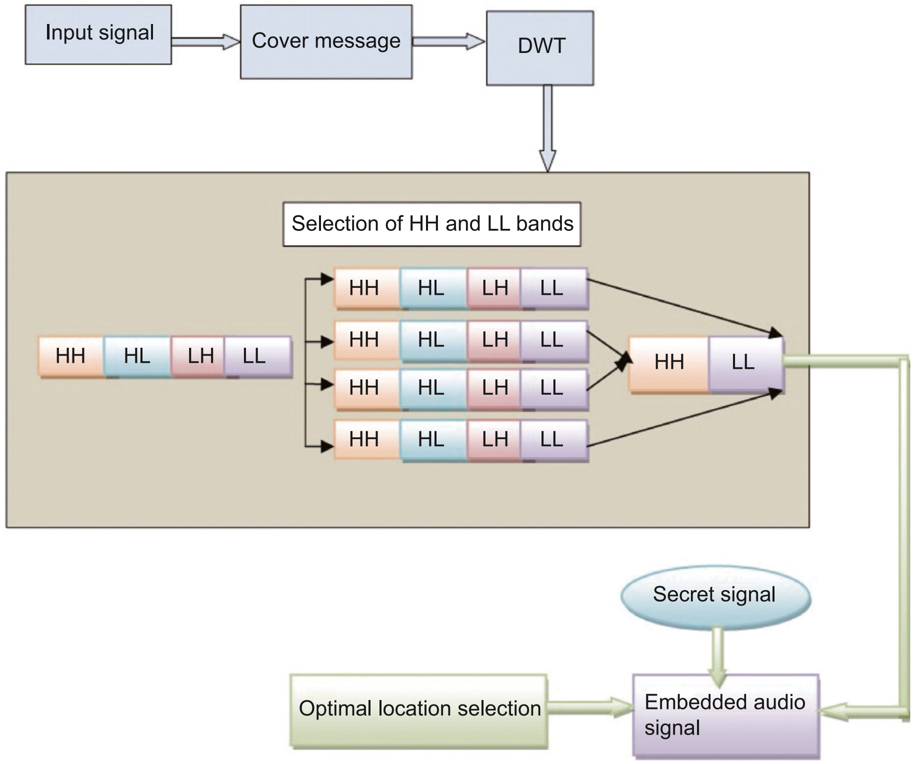


Fig. 3. The secret data embedding process.

4.1. Experimental setup

The proposed digital audio watermarking system is implemented in PYTHON and the system configuration of the implementation consists of the PyCharm software executed in Windows 10 system software.

4.1.1. Dataset description

The brain tumor image is utilized as an input image from the BraTS database, which is in the form of multimodal scans with various medical practices. Consequently, the available scans were interpreted by the experts for the different glioma sub-regions in the BraTS database.

4.1.2. Performance metrics

The effectiveness of the proposed digital audio watermarking system is evaluated by analyzing the metrics, such as Mean Square Error (MSE), Bit Error Rate (BER), and Signal-to-noise ratio (SNR).

Mean square error

The mean square error is one of the statistical models, which is estimated by taking the average squared difference between the estimated and the actual value.

$$\mu_{squared\ error} = \frac{1}{n_{data}} \sum_{i=1}^{n_{data}} \left(\Upsilon_i - \hat{\Upsilon}_i \right) \tag{36}$$

where n_{data} denotes the data available, Υ_i represents the original value, and $\hat{\Upsilon}$ represents the estimated value.

Bit error rate

The bit error rate is utilized to estimate the number of errors per unit of time and it is obtained by dividing the total number of errors by the total number of transmitted bits.

$$Bit_{error} = \frac{\sum_{i=1}^{w_{size}} \sum_{j=1}^{w_{size}} S_{org}(i,j) \oplus S_{ex}(i,j)}{w_{size} \times w_{size}} \tag{37}$$

The watermark size is denoted by $w_{size} \times w_{size}$ and $\oplus$ represents the exclusive OR operator.

Signal-to-noise ratio

The SNR is defined as the ratio between the signal power to the noise power and it is generally expressed in decibels. The mathematical representation of the peak-to-signal ratio is expressed as

$$SNR = 10 \log_{10} \left(\frac{S_{\rho ow}}{N_{\rho ow}} \right) \tag{38}$$

4.2. Experiment analysis

The experimental results for embedding the image into the audio signal are revealed in Fig. 4 with the input signal, corresponding embedded image, recovered image, the original image, and the embedded signal for the various noises. The analysis is done by including various types of noise such as salt and pepper noise, gaussian noise, and random noise, and without noise, the proposed method performed well even without noise.

4.3. Performance analysis

In this section, the MSE, BER, and SNR of the SLOA optimization utilizing the five different audio signals are analyzed by considering the salt and pepper noise, gaussian noise, random noise, and without noise.

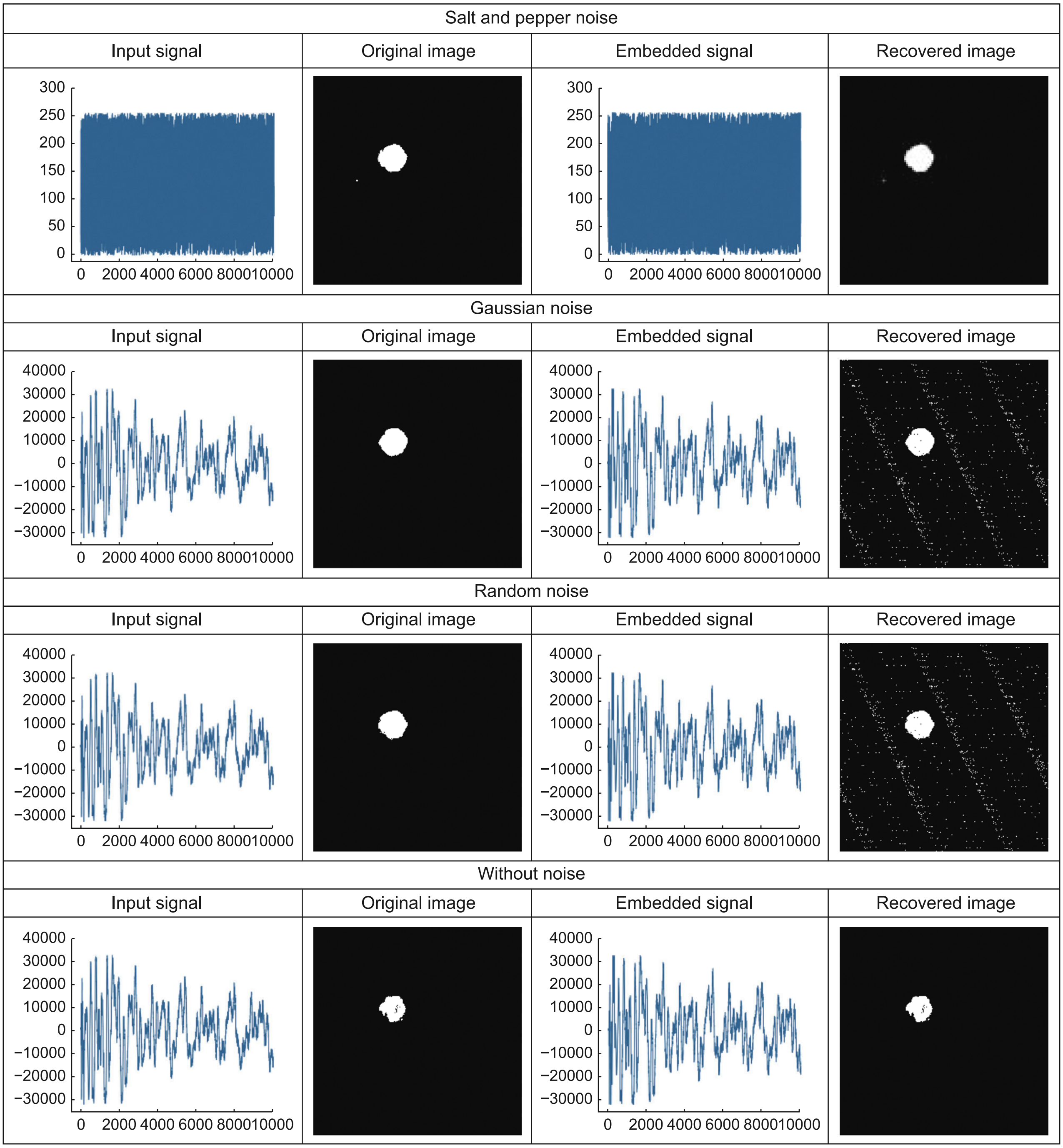


Fig. 4. Experimental results for Image 1.

4.3.1. Performance analysis for image 1 based on signals

The SLOA optimization performance for five different signals using the various noise in terms of the MSE, BER, and SNR are revealed in Fig. 5 and Table 2. Fig. 5(a) represents the MSE for both the signals and the various corresponding noises. The MSE for signal 5 based on the salt and pepper noise, gaussian noise, random noise, and without noise are 0.065, 0.062, 0.059, and 0.056.

Fig. 5(b) represents the BER for both the signals and the various corresponding noises. The BER for signal 5 based on the 3 different noise, and without noise are 0.081, 0.081, 0.081, and 0.035.

Fig. 5(c) represents the SNR for both the signals and the various corresponding noises. The SNR for signal 5 based on the 3 different noise, and without noise are, and without noise are 53.422 dB, 54.422 dB, 54.422 dB, and 56.422 dB respectively.

4.3.2. Performance analysis for image 1 based on noise intensity

The SLOA optimization performance for various noise intensities using the various noise in terms of the MSE, BER, and SNR are revealed in Fig. 6. Fig. 6(a) represents the MSE for both the noise intensity and their corresponding noises. The MSE for the 0.4 noise intensity based on the 3 different noise, and without noise are 0.093, 0.093, 0.090, and 0.078.

Fig. 6(b) represents the BER for both the noise intensity and their various corresponding noises. The BER for the 0.4 noise intensity based on the 3 different noise, and without noise are and without noise are 0.074, 0.065, 0.054, and 0.043.

Table 2
Performance for image 1 based on signals.

Signals	Salt and pepper noise			Gaussian noise			Random noise			Without noise		
	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)
Signal 1	0.111	0.095	42.363	0.104	0.090	43.363	0.101	0.089	43.363	0.099	0.082	45.363
Signal 2	0.107	0.085	43.378	0.094	0.084	44.378	0.091	0.084	44.378	0.082	0.071	46.378
Signal 3	0.097	0.084	45.041	0.083	0.083	46.041	0.081	0.083	46.041	0.080	0.059	48.041
Signal 4	0.086	0.083	50.756	0.075	0.083	51.756	0.073	0.083	51.756	0.070	0.047	53.756
Signal 5	0.065	0.081	53.422	0.062	0.081	54.422	0.059	0.081	54.422	0.056	0.035	56.422

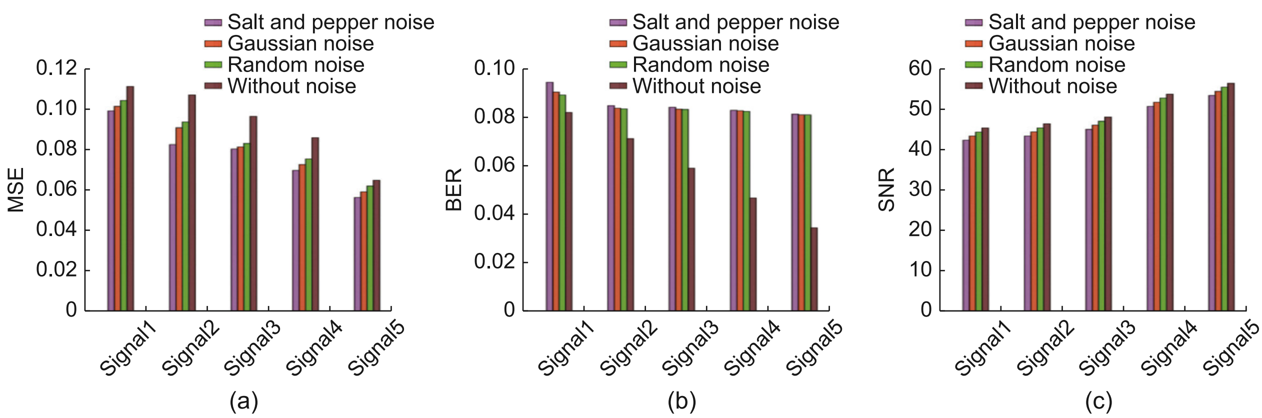


Fig. 5. Performance analysis for image 1 based on signals. (a) MSE. (b) BER. (c) SNR.

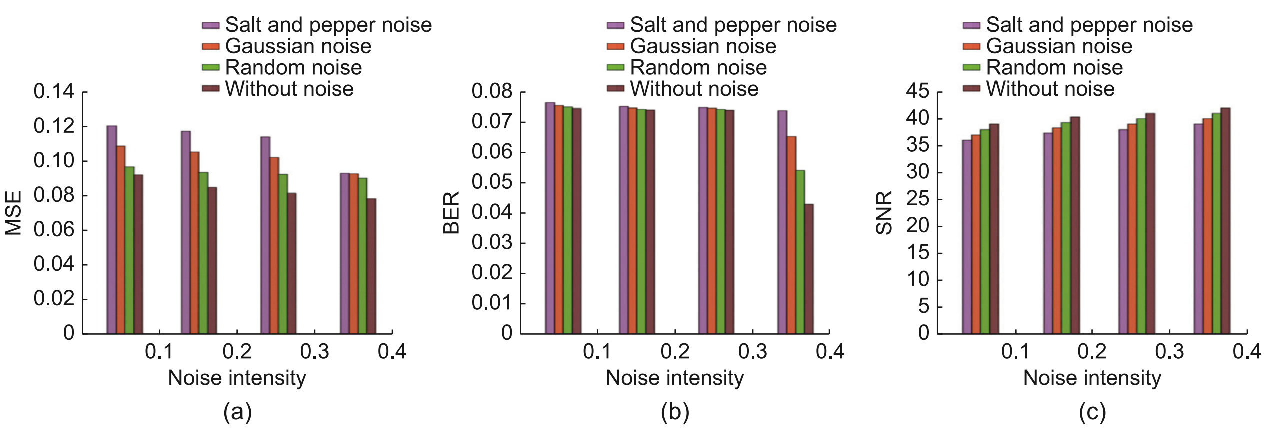


Fig. 6. Performance analysis for image 1 based on noise intensity. (a) MSE. (b) BER. (c) SNR.

Fig. 6(c) represents the SNR for both the noise intensity and their various corresponding noises. The SNR for the 0.4 noise intensity based on the 3 different noise, and without noise are 39.001 dB, 40.001 dB, 41.001 dB, and 42.001 dB respectively.

4.3.3. Performance analysis for image 2 based on signals

The SLOA optimization performance for five different signals using the various noise in terms of the MSE, BER, and SNR are revealed in Fig. 7 and Table 3. Fig. 7(a) represents the MSE for both the signals and the various corresponding noises. The MSE for signal 5 based on the 3 different noise, and without noise are 0.046, 0.042, 0.038, and 0.034.

Fig. 7(b) represents the BER for both the signals and the various corresponding noises. The BER for signal 5 based on the 3 different noises, and without noise are 0.079, 0.078, 0.078, and 0.051.

Fig. 7(c) represents the SNR for both the signals and the various corresponding noises. The SNR for signal 5 based on the 3 different noise, and without noise are 57.634 dB, 58.634 dB, 59.634 dB, and 60.634 dB respectively.

4.3.4. Performance analysis for image 2 based on noise intensity

The SLOA optimization performance for various noise intensities using the various noise in terms of the MSE, BER, and SNR are revealed in Fig. 8. Fig. 8(a) represents the MSE for both the noise intensity and their corresponding noises. The MSE for the 0.4 noise intensity based on the 3 different noise, and without noise are 0.110, 0.060, 0.058, and 0.055.

Fig. 8(b) represents the BER for both the noise intensity and their various corresponding noises. The BER for the 0.4 noise based on the 3 different noise, and without noise are 0.073, 0.070, 0.062, and 0.054.

Fig. 8(c) represents the SNR for both the noise intensity and their various corresponding noises. The SNR for the 0.4 noise based on the 3 different noise, and without noise are 42.001 dB, 43.002 dB, 44.004 dB, and 45.011 dB respectively.

4.4. Comparative methods

The methods considered for evaluating the performance of SLOA optimization in the optimal block selection by considering

Table 3
Performance for image 2 based on signals.

Signals	Salt and pepper noise			Gaussian noise			Random noise			Without noise		
	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)
Signal 1	0.131	0.086	43.045	0.102	0.084	44.045	0.098	0.083	45.045	0.098	0.079	46.045
Signal 2	0.106	0.085	44.067	0.087	0.081	45.067	0.083	0.080	46.067	0.053	0.077	47.067
Signal 3	0.091	0.080	46.062	0.072	0.080	47.062	0.068	0.080	48.062	0.052	0.068	49.062
Signal 4	0.076	0.079	54.134	0.061	0.079	55.134	0.057	0.079	56.134	0.049	0.059	57.134
Signal 5	0.046	0.079	57.634	0.042	0.078	58.634	0.038	0.078	59.634	0.034	0.051	60.634

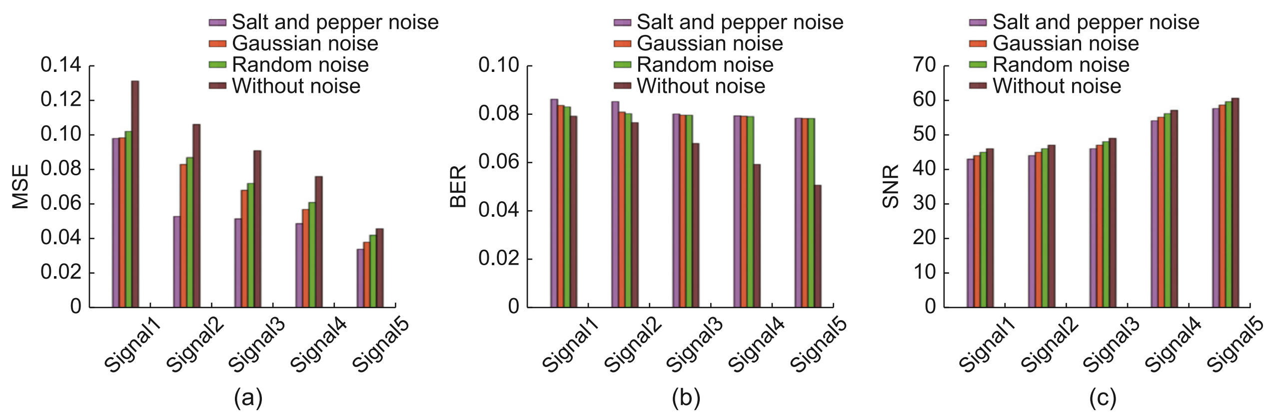


Fig. 7. Performance analysis for image 2 based on signals. (a) MSE. (b) BER. (c) SNR.

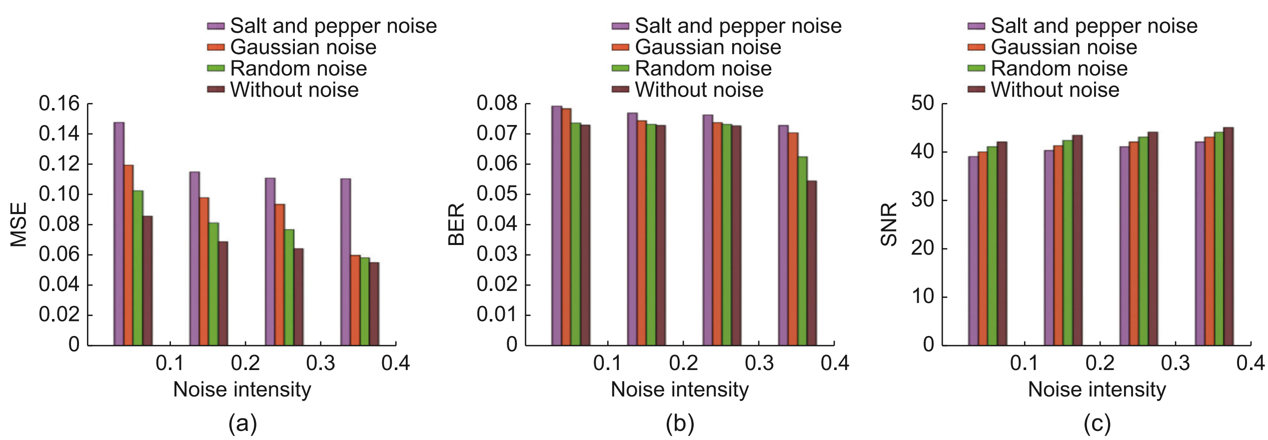


Fig. 8. Performance analysis for image 2 based on noise intensity. (a) MSE. (b) BER. (c) SNR.

the MSE, SNR, and BER are the [18–20], and DCNN with Hybrid Swarm optimization (HSO).

4.4.1. Comparative analysis for image 1 based on signals

The SLOA optimization performance for five different signals, when compared to the other existing methods in terms of the MSE, BER, and SNR, is revealed in Fig. 9. Fig. 9(a) represents the MSE for both the signals as well as the corresponding existing and proposed method. The attained MSE for the SLOA optimization for signal 5 is 0.039, and the attained variations for the MSE is 46.77% when compared to the DCNN with the HSO optimization method.

Fig. 9(b) represents the BER for both the signals as well as the corresponding existing and proposed method. The attained BER for the SLOA optimization for signal 5 is 0.089, and the attained variations for the MSE is 1.45% when compared to the DCNN with the HSO optimization method.

Fig. 9(c) represents the SNR for both the signals as well as the corresponding existing and proposed method. The attained SNR for the SLOA optimization for signal 5 is 48.861 dB, and the attained variations for the SNR are 2.05% when compared to the DCNN with the HSO optimization method.

4.4.2. Comparative analysis for image 2 based on signals

The SLOA optimization performance for five different signals, when compared to the other existing methods in terms of the MSE, BER, and SNR, is revealed in Fig. 10. Fig. 10(a) represents the MSE for both the signals as well as the corresponding existing and proposed method. The attained MSE for the SLOA optimization for signal 5 is 0.014, and the attained variations for the MSE is 62.88% when compared to the DCNN with the HSO optimization method.

Fig. 10(b) represents the BER for both the signals as well as the corresponding existing and proposed method. The attained BER for the SLOA optimization for signal 5 is 0.086, and the attained

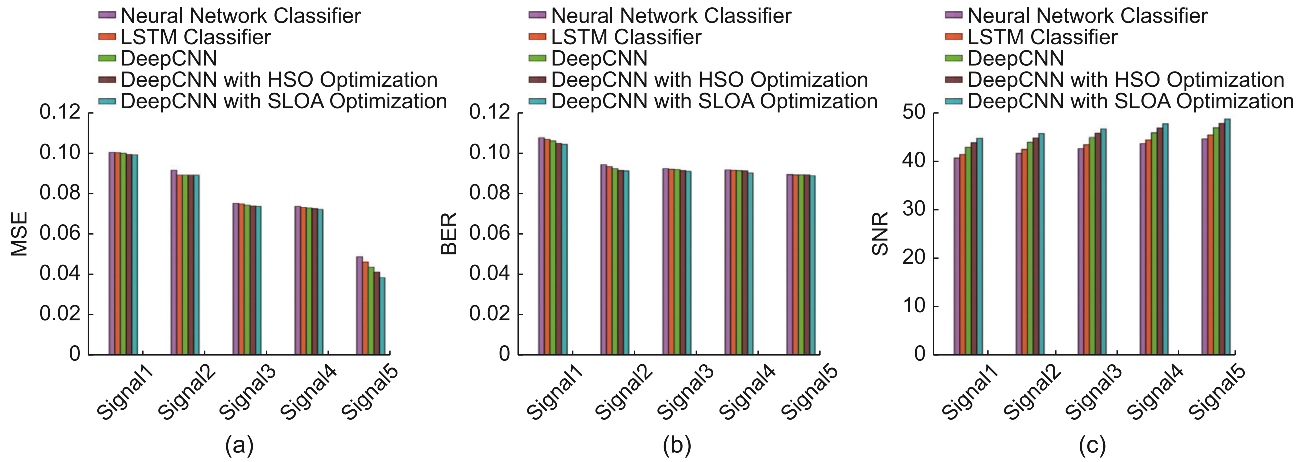


Fig. 9. Comparative analysis for input image 1. (a) MSE. (b) BER. (c) SNR.

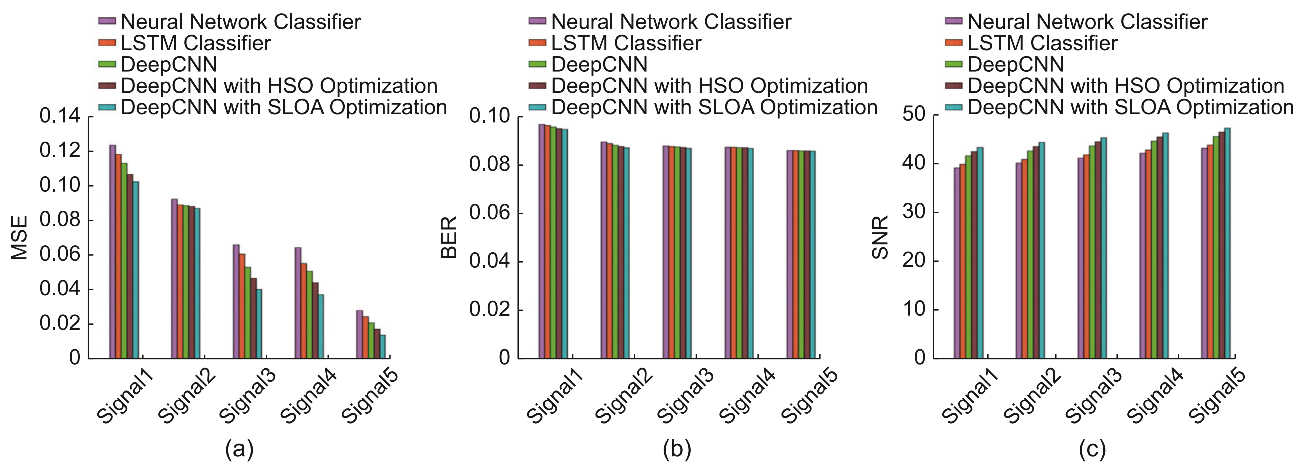


Fig. 10. Comparative analysis for image 2. (a) MSE. (b) BER. (c) SNR.

variations for the MSE is 1.34% when compared to the DCNN with the HSO optimization method.

Fig. 10(c) represents the SNR for both the signals as well as the corresponding existing and proposed method. The attained SNR for the SLOA optimization for signal 5 is 47.292 dB, and the attained variations for the SNR is 2.11% when compared to the DCNN with the HSO optimization method.

4.5. Comparative discussion

In this section, the MSE, BER, and the SNR of the various existing, as well as the proposed method, are described in Tables 4 and 5 for the input image 1 and image 2 depending on five different types of signals. The attained MSE for the five different signals in the input image 1 is 0.049, 0.046, 0.044, 0.041, and 0.039. The attained BER for the five different signals in input image 1 is 0.089, 0.089, 0.089, 0.089., and 0.089 The attained SNR for the five different signals in input image 1 is 44.753 dB, 45.540 dB, 47.050 dB, 47.961 dB, and 48.861 dB respectively.

The attained MSE for the five different signals in the input image 2 is 0.028, 0.025, 0.021, 0.017, and 0.014. The attained BER for the five different signals in input image 2 is 0.086, 0.086, 0.086, 0.086, and 0.086. The attained SNR for the five different signals in input image 2 is 43.130 dB, 43.810 dB, 45.575 dB, 46.441 dB, and 47.292 dB respectively.

Compared to the existing classifiers and the optimization-based classifiers, the DCNN with SLOA optimization attains a higher SNR, low MSE, and BER. Thus, the more efficient output and enhanced signal quality are attained by the presence of enormous relevant information than the irrelevant noise signals. The time complexity of the proposed DCNN-SLOA model is compared with the conventional methods and tabulated in Table 6.

5. Conclusion

The audio signals are easily subjected to tampering and modifications, which may lead to various privacy and security issues. This research article implements and evaluates the audio watermarking model based on deep learning techniques. The DWT approach is utilized in this research to enhance the performance of the audio watermarking model. The importance of the selection of optimal locations for embedding the secret message is highlighted in the research, which is accomplished by deep learning techniques. The selection of the optimal embedding location is done by a DCNN classifier in which the hyperparameter is optimally tuned by the proposed search location optimization. The evaluation based on the performance metrics, such as BER, MSE, and SNR reveals that the optimization-based DCNN is found to be effective in the watermarking system. The robustness and performance improvement of the proposed model concerning MSE, BER, and SNR is obtained as 46.77%, 1.45%, and 2.05% while comparing

Table 4
Comparative analysis for image 1 based on signals.

Methods	Signal 1			Signal 2			Signal 3			Signal 4			Signal 5		
	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)
Neural Network	0.101	0.108	40.753	0.100	0.107	41.540	0.100	0.106	43.050	0.100	0.105	43.960	0.099	0.104	44.861
LSTM	0.092	0.094	41.753	0.089	0.093	42.540	0.089	0.092	44.050	0.089	0.091	44.960	0.089	0.091	45.861
DCNN	0.075	0.092	42.753	0.075	0.092	43.540	0.074	0.092	45.050	0.074	0.091	45.960	0.074	0.091	46.861
DCNN with HSO optimization	0.074	0.092	43.753	0.073	0.092	44.540	0.073	0.091	46.050	0.073	0.091	46.960	0.072	0.090	47.861
DCNN with SLOA optimization	**0.049**	**0.089**	**44.753**	**0.046**	**0.089**	**45.540**	**0.044**	**0.089**	**47.050**	**0.041**	**0.089**	**47.961**	**0.039**	**0.089**	**48.861**

Table 5
Comparative analysis for image 2 based on signals.

Methods	Signal 1			Signal 2			Signal 3			Signal 4			Signal 5		
	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)	MSE	BER	SNR (dB)
Neural Network	0.124	0.097	39.130	0.118	0.097	39.810	0.113	0.096	41.575	0.107	0.095	42.441	0.102	0.095	43.292
LSTM	0.092	0.090	40.130	0.089	0.089	40.810	0.089	0.088	42.575	0.088	0.088	43.441	0.087	0.087	44.292
DCNN	0.066	0.088	41.130	0.061	0.088	41.810	0.053	0.088	43.575	0.047	0.088	44.441	0.040	0.087	45.292
DCNN with HSO optimization	0.064	0.088	42.130	0.055	0.088	42.810	0.051	0.087	44.575	0.044	0.087	45.441	0.037	0.087	46.292
DCNN with SLOA optimization	0.028	0.086	43.130	0.025	0.086	43.810	0.021	0.086	45.575	0.017	0.086	46.441	0.014	0.086	47.292

Table 6
Time complexity analysis for DCNN-SLOA.

Methods	Computation time (ms)
Neural network	76
LSTM	82
DCNN	93
DCNN with HSO optimization	79
DCNN with SLOA optimization	68

the DCNN with the HSO model. However, there requires a performance increment in the proposed audio watermarking system in the presence of noise. Hence, in the future, the advanced filters will be utilized to remove the noise from the audio signals and the hybridization of the advanced classifiers will be utilized to mitigate the issues like vanishing gradients.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Nomenclature

S_{audio}	Cover audio signal
S_{tot}	Total signal
T_{tot}	Total time interval
b_{tot}	Total signal block
u	Dimensional exploring region
pth	Exploring period
$Y_j^p \in \mathbb{R}^u$	Present location1
θ_j^p	Follower's head angle
Y	Promoter
Y_x	Promoter zero degrees
Y_R	Promoter right side
Y_h	Promoter left side
R_1	Random numbers
$\varphi_{\max}$	Maximum pursuit angle
G	Global best location
C	Promoter scanning at zero degrees
α, γ	Constant parameters
C and D	Constant terms
$\beta_{\max} \in \mathbb{R}^1$	Maximum rotating angle
θ_j	Randomized head angle

References

[1] K. Galajit, J. Karnjana, M. Unoki, P. Aimmanee, Semi-fragile speech watermarking based on singular-spectrum analysis with CNN-based parameter estimation for tampering detection, APSIPA Trans. Signal Inf. Process. (8) (2019).

[2] P. Garg, R.R. Kishore, An efficient and secured blind image watermarking using ABC optimization in DWT and DCT domain, Multimedia Tools Appl. (2021) 1–18.

[3] H. Karajeh, T. Khatib, L. Rajab, M. Maqableh, A robust digital audio watermarking scheme based on DWT and Schur decomposition, Multimedia Tools Appl. 78 (13) (2019) 18395–18418.

[4] K. Galajit, J. Karnjana, P. Aimmanee, M. Unoki, Digital audio watermarking method based on singular spectrum analysis with automatic parameter estimation using a convolutional neural network, in: Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2018, pp. 63–73.

[5] D. Hu, D. Zhao, S. Zheng, A new robust approach for reversible database watermarking with distortion control, IEEE Trans. Knowl. Data Eng. 31 (6) (2018) 1024–1037.

[6] F. Deeba, S. Kun, F.A. Dharejo, H. Langah, H. Memon, Digital watermarking using deep neural network, Int. J. Mach. Learn. Comput. 10 (2) (2020) 277–282.

[7] A. Kaur, M.K. Dutta, J. Prinosil, General regression neural network based audio watermarking algorithm using torus automorphism, in: Proceedings of 41st International Conference on Telecommunications and Signal Processing (TSP), 2018, pp. 1–4.

[8] L. Sun, J. Xu, S. Liu, S. Zhang, Y. Li, C.A. Shen, A robust image watermarking scheme using Arnold transform and BP neural network, Neural Comput. Appl. 30 (8) (2018) 2425–2440.

[9] S.M. Pourhashemi, M. Mosleh, Y. Erfani, A novel audio watermarking scheme using ensemble-based watermark detector and DWT, Neural Comput. Appl. 33 (11) (2021) 6161–6181.

[10] Q. Su, B. Chen, Robust color image watermarking technique in the spatial domain, Soft Comput. 22 (2018) 91–106.

[11] C.I. Podilchuk, E.J. Delp, Digital watermarking: algorithms and applications, IEEE Signal Process. Mag. 18 (4) (2001) 33–46.

[12] R. Li, S. Xu, H. Yang, Spread spectrum audio watermarking based on perceptual characteristic aware extraction, IET Signal Process. 10 (3) (2016) 266–273.

[13] H. Latifpour, M. Mosleh, M. Kheyrandish, An intelligent audio watermarking based on KNN learning algorithm, Int. J. Speech Technol. 18 (4) (2015) 697–706.

[14] M.K. Dutta, V.K. Pathak, P. Gupta, An adaptive robust watermarking algorithm for audio signals using SVD, in: Proceedings of Transactions on Computational Science Vol. X, 2010, pp. 131–153.

[15] S.J. Mousavirad, H. Ebrahimpour-Komleh, Human mental search: a new population-based metaheuristic optimization algorithm, Appl. Intell. 47 (3) (2017) 850–887.

[16] S. He, Q.H. Wu, J.R. Saunders, Group search optimizer: an optimization algorithm inspired by animal searching behavior, IEEE Trans. Evol. Comput. 13 (5) (2009) 973–990.

[17] BraTS dataset, https://www.med.upenn.edu/cbica/brats2020/data.html.

[18] E.Le. Merrer, P. Perez, G. Trédan, Adversarial frontier stitching for remote neural network watermarking, Neural Comput. Appl. 32 (13) (2020) 9233–9244.

[19] A. Ferdowsi, W. Saad, Deep learning-based dynamic watermarking for secure signal authentication in the Internet of Things, in: 2018 IEEE International Conference on Communications (ICC), IEEE, 2018, pp. 1–6.

[20] S. Ingaleshwar, N.V. Dharwadkar, Water chaotic fruit fly optimization-based deep convolutional neural network for image watermarking using wavelet transform, Multimedia Tools Appl. (2021) 1–25.

SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I7P111

Volume 11 Issue 7, 108-116, July 2024

Original Article

AudioStamp: A Deep Learning Based Watermarking Procedure for Copyright Protection of Digital Audio Files

Abhijit Patil[1], Ramesh Shelke[2], Dilendra Hiran[3]

[1,3]*Department of Computer Engineering, PAHER University, Rajasthan, India.*
[2]*Department of EXTC Engineering, University of Mumbai, Maharashtra, India.*

[1]*Corresponding Author : abhijietsir@gmail.com*

Received: 08 May 2024 Revised: 08 June 2024 Accepted: 07 July 2024 Published: 27 July 2024

Abstract - *The ubiquitous use of digital media across various platforms has heightened the risk of copyright infringement and unauthorized distribution. Digital content such as images, audio, and video can be easily subjected to copyright violations if it is not adequately secured and protected using effective technological measures. In this paper, we explore different methods employed for safeguarding the copyright of digital media and propose a novel approach for copyright protection of audio files through the integration of watermarking techniques and neural networks. The proposed work concentrates on digital audio files. Our methodology leverages watermarking to embed ownership information or identifiers into audio files, ensuring their traceability and authenticity. Furthermore, we utilize neural networks, specifically encoder-decoder architecture, to enhance the robustness and security of the audio watermarking system. The primary objective of this innovative approach is to ensure robust protection of digital media without degrading the audio quality or clarity of embedded images. Utilizing sophisticated signal processing techniques, including wavelet transforms and denoising algorithms, the system embeds and subsequently reconstructs watermarked images within audio files with high fidelity. The goal is to strike an optimal balance between security and usability, providing content creators with a reliable method to safeguard their intellectual property. We evaluate the proposed method's performance against critical parameters such as Maximum Correlation and Peak Signal-to-Noise Ratio (PSNR), among others. By training neural networks to embed watermarks imperceptibly and detect them accurately, we aim to provide a robust solution for copyright protection in the digital audio domain.*

Keywords - *Copyright protection, Audio watermarking, Encoder-decoder, Deep Learning, Imperceptibility, Robustness.*

1. Introduction

The advent of digital technologies and the internet has revolutionized the creation, dissemination, and consumption of audio content, enabling unparalleled access to a vast array of music, podcasts, audiobooks, and other forms of personalized data in audio format. However, this digital landscape also presents significant challenges in terms of copyright protection, such as unauthorized copying, distribution, and sharing of audio files. Addressing these challenges necessitates robust mechanisms to ensure the integrity, authenticity, as well as ownership of digital audio content. In the literature, various approaches have been proposed to tackle these issues.

One of the promising techniques for safeguarding audio files against copyright infringement is the use of watermarking, which involves embedding another signal or ownership information directly into the host audio signal, making it difficult for attackers to extract or modify. Watermarking provides a means of uniquely identifying and tracing digital data, thereby enabling content creators and rights holders to prove ownership and enforce copyright protection measures.

Intellectual property protection encompasses proof of ownership, access control, tracing illegal copies, and other aspects related to music/audio files. Traditional transform domain approaches for audio watermarking include Fourier Transform (FFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Singular Value Decomposition (SVD), and hybrid techniques [7, 9, 14, 18]. Machine learning [6, 27, 28] and Deep learning [3] approaches are currently being utilized to improve the robustness and imperceptibility of the watermarked signal.

Additionally, cryptographic algorithms are being employed to enhance the security of the watermark and watermarked data. The most significant application of audio watermarking is to protect intellectual property rights and prevent unauthorized access. Consequently, there is a pressing need for a robust system for intellectual property protection of audio/music files. The goals of robustness and imperceptibility can be achieved through the synergy of deep learning and watermarking techniques.

Traditional audio watermarking techniques heavily rely on expert knowledge and empirical rules, which pose challenges in implementation and tend to offer limited encoding capacity while being susceptible to different types of signal attacks. Current advancements in deep learning and neural networks have paved the way for enhancing the performance and security of watermarking techniques. Deep learning architectures, including Convolutional Neural Networks (CNNs) [9, 25] and Recurrent Neural Networks (RNNs), have exhibited remarkable capabilities in feature extraction, pattern recognition, and signal processing tasks.

The approaches used for digital signal watermarking can be divided into three classes: i) Traditional, ii) Machine learning based and iii) Deep learning based. Deep learning techniques have demonstrated promising capabilities in audio watermarking, offering higher encoding capacity and enhanced robustness against attacks. Deep Neural Networks (DNNs) can automatically learn and adapt to predefined attacks, significantly reducing the complexity involved in designing encoding strategies.

This research paper aims to give a comprehensive review of the current techniques used in copyright protection of audio files using watermarking and deep learning. We explored the potential benefits of DNN-based audio watermarking. Throughout this paper, we analyzed the existing literature, highlighting key contributions and innovations in the field. Additionally, we will address the limitations and challenges of the current techniques and provide recommendations for future research endeavors. By harnessing the power of neural networks, we can develop more robust and resilient watermarking schemes capable of handling common attacks on audio signals without disturbing the quality and integrity of the original audio signal.

This research aims to provide an innovative approach for embedding a copyright of the owner (Image) into the audio without degrading the quality of the audio. Hence, In this paper a novel approach is proposed which makes use of DWT and Encoder-Decoder Network. The methodology combines state-of-the-art watermarking techniques with advanced neural network architectures to achieve robust and secure embedding of ownership information into audio content.

The main aim of the proposed method is to address the limitations of existing methods and provide a better solution for copyright enforcement in digital audio files. Despite different approaches being used by the researchers, some research gaps are identified in finding a novel solution for copyright protection. The main research gap is that no previous approach satisfies all of the required parameters at its highest level for achieving the security and robust, tamper proof watermarked audio data with respect to:

- Preserving the quality of the original audio is an important factor while doing all the required modifications to the original signal during the watermark embedding process.
- Making the watermarked signal more secure against attacks by applying the best suitable approach.
- Extraction of original watermark with the greatest accuracy. This can be tested using the Bit Error Rate.

Out of all the previously used approaches, no one had used and tested the autoencoder based approach for watermarking and extraction. This research work tested the autoencoders based procedure for audio watermarking to achieve the properties of Imperceptibility, Security and Robustness. The paper is organized as follows: Section 1 deals with the introduction of the topic, In Section 2, an overview of recent techniques being used and their applications is mentioned. It also discusses the role of neural networks in enhancing the security and robustness of watermarking schemes in copyright protection. Section 3 describes the proposed methodology in detail. Experimental results and performance metrics are discussed in Section 4. Finally, the paper is concluded in Section 5, highlighting the contributions and potential applications of our research. The future directions of this research are also highlighted.

2. Literature Review

While traditional watermarking methods have demonstrated effectiveness to some extent, they often lack robustness against common attacks on audio signal processing. In the literature, diverse approaches are employed for this task. Through the literature survey, we found various techniques being used and few of them are discussed here with their merits and demerits.

[1] Huang et al. proposed a “digital audio watermarking algorithm” to protect music multimedia works. The paper highlights the importance of copyright protection in multimedia and introduces a novel watermarking approach using sparse representation persistent-based techniques to embed imperceptible watermarks. They utilized Improved Singular Value Decomposition (iSVD) and Orthogonal Matching Pursuit (OMP) for implementation. Through experimental evaluation, the authors demonstrated the proposed algorithm’s effectiveness and robustness in preserving music content’s integrity and ownership. This research contributes to advancing copyright protection techniques in multimedia applications.

[2] Liu et al. introduced “DeAR, a deep-learning-based audio watermarking scheme” is resilient to re-recording attacks. The paper addresses protecting audio content against re-recording attacks by leveraging deep learning techniques.

Experimental validation demonstrated DeAR's effectiveness in embedding robust watermarks that withstand re-recording while maintaining perceptual quality. Their research contributes to developing resilient watermarking methods for copyright protection in audio applications. Their test results have shown SNR: 25.86 db and BER accuracy: 98.55%.

[3] Pavlović, Kovačević, and Đurović propose a novel speech watermarking technique based on Deep Neural Networks (DNNs). An encoder-decoder architecture is used for watermark embedding and extraction, achieving PSNR above 57dB and 100% transmission accuracy.

[4] Qu et al. propose AudioQR, a deep neural audio watermarking scheme using QR codes. They used an encoder-decoder framework for QR embedding and extraction. The proposed method is evaluated against Bit Error Rate (BER) and Signal-to-Noise Ratio (SNR). Their research contributes to developing QR-based audio watermarking techniques for multimedia applications. The test results have shown SNR: 31.84 and with data augmentation: 49.50 db BER Accuracy: 99.9%

[5] Singha and Ullah demonstrated an audio watermarking scheme with the decentralization of watermarks. The paper introduces a novel approach by distributing watermark data across multiple frequency bands using 16-level DWT along with the SVD technique for decentralized watermark embedding and extraction. They tested the system's effectiveness by applying "additive white Gaussian noise" to the watermarked signal. Their approach showed robustness against common attacks while maintaining watermark invisibility.

[6] Wang, Qi, and Niu proposed a new adaptive technique based on Support Vector Regression (SVR). The paper addresses the challenge of adaptability in watermarking schemes by leveraging SVR to dynamically adjust watermark embedding parameters based on the audio signal's characteristics.

Experimental validation demonstrated the approach's robustness in achieving imperceptible watermarking while maintaining high audio quality fidelity. The system is robust against common attacks. They obtained Normalized Cross-correlation (NC): 0.9858, Distortion Ratio (DR): 0.0081, Peak Signal-to-Noise Ratio (PSNR): 41.81.

[9] Galajit et al. proposed "A semi-fragile speech watermarking scheme based on Singular-Spectrum Analysis (SSA) with CNN-based parameter estimation for tampering detection." Experimental validation demonstrated the approach's effectiveness in detecting tampering and ensuring the integrity of speech signals.

[10] Kumsawat, Attakitmongcol, and Srikaew used a multiwavelet transform to achieve copyright protection in digital audio. They introduced a different approach to embedding watermarks into audio signals using multiwavelet transform, achieving improved robustness and imperceptibility. Experimental validation demonstrated the effectiveness of their approach in preserving audio content's integrity and ownership.

[14] Pourhashemi, Mosleh, and Erfani proposed an "ensemble-based watermark detector and discrete wavelet transform". The paper introduced an innovative approach to embedding watermarks into audio signals using DWT and employing ensemble-based detectors for watermark extraction. Their experimental demonstration showed the system's effectiveness in detecting watermarks and ensuring the authenticity of audio content.

[17] M. R. R. Ansori, Allwinnaldo, R. N. Alief et al. "HADES: A Hash-based Audio Copy Detection System" is implemented. It uses a novel approach with audio hash and integration of blockchain for a decentralized, transparent way of ownership protection. The proposed system is robust against many types of attacks.

2.1. Objectives of Research

Both CNNs and RNNs can be trained on large datasets of audio signals and watermarks, enabling them to learn robust and generalizable representations for watermarking. Additionally, these models can be combined with traditional signal processing techniques or other neural network architectures (e.g., autoencoders, GANs) to create hybrid or ensemble models for improved performance. Our research aims to test our custom autoencoders on audio files to address the challenges through several key objectives:

- Utilize autoencoder technology to embed imperceptible watermarks in audio streams without altering audio quality.
- Improve the reconstruction of images from watermarked audio to maintain original clarity and integrity.
- Achieve a practical equilibrium between imperceptibility security features and user experience without degrading the quality of the audio file.
- Rigorously test the system against common and sophisticated attacks to ensure robustness.

3. Materials and Methods

The proposed system introduces an innovative approach to multimedia data encoding and decoding along with the integration of image and audio signals for watermarking, employing meticulous preprocessing with the help of autoencoder and Discrete Wavelet Transform. Through encoding, image data is embedded into audio, while decoding efficiently retrieves and reconstructs the original image using a dedicated neural network, ensuring robust data integrity preservation across domains. Methodology: We perform a series of experiments to test the effectiveness of our approach in terms of watermark imperceptibility, resilience against various attacks, and the accuracy of watermark extraction. The whole process involves 7 steps, from pre-processing up to the final output. This research endeavor contributes to the advancement of sophisticated techniques aimed at safeguarding intellectual property rights in the digital audio domain.

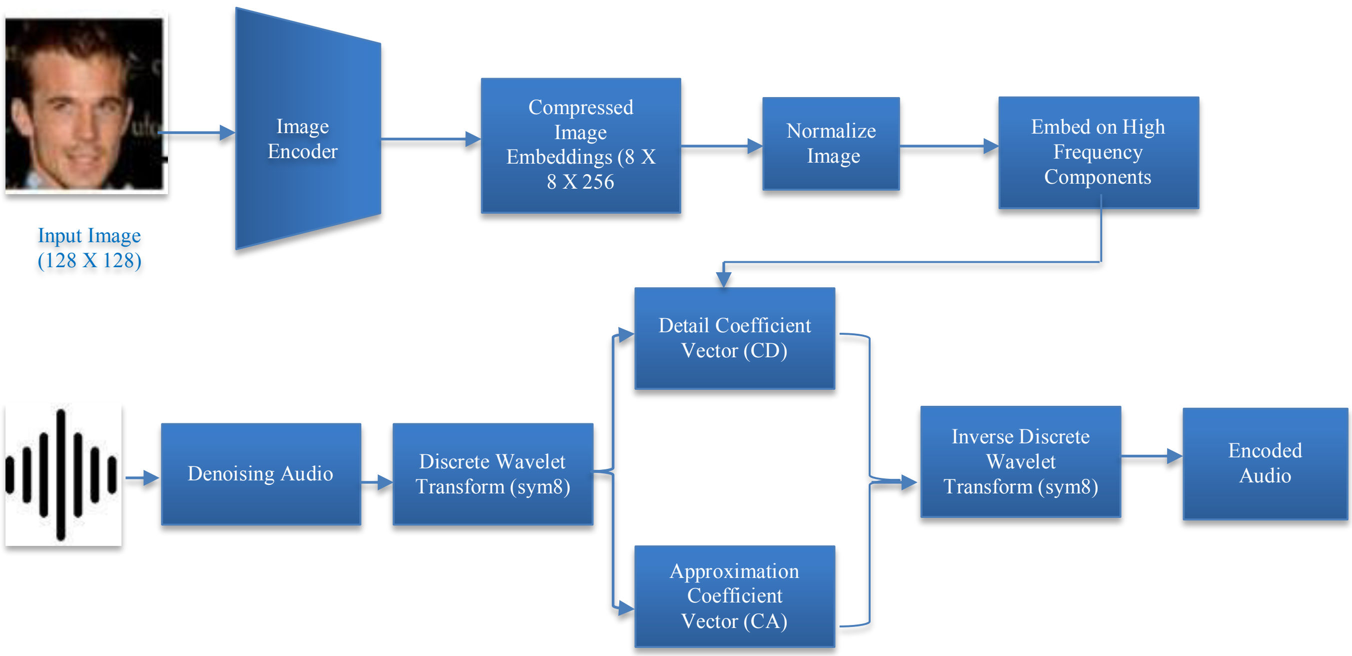


Fig. 1 Encoder architecture: Watermark embedding

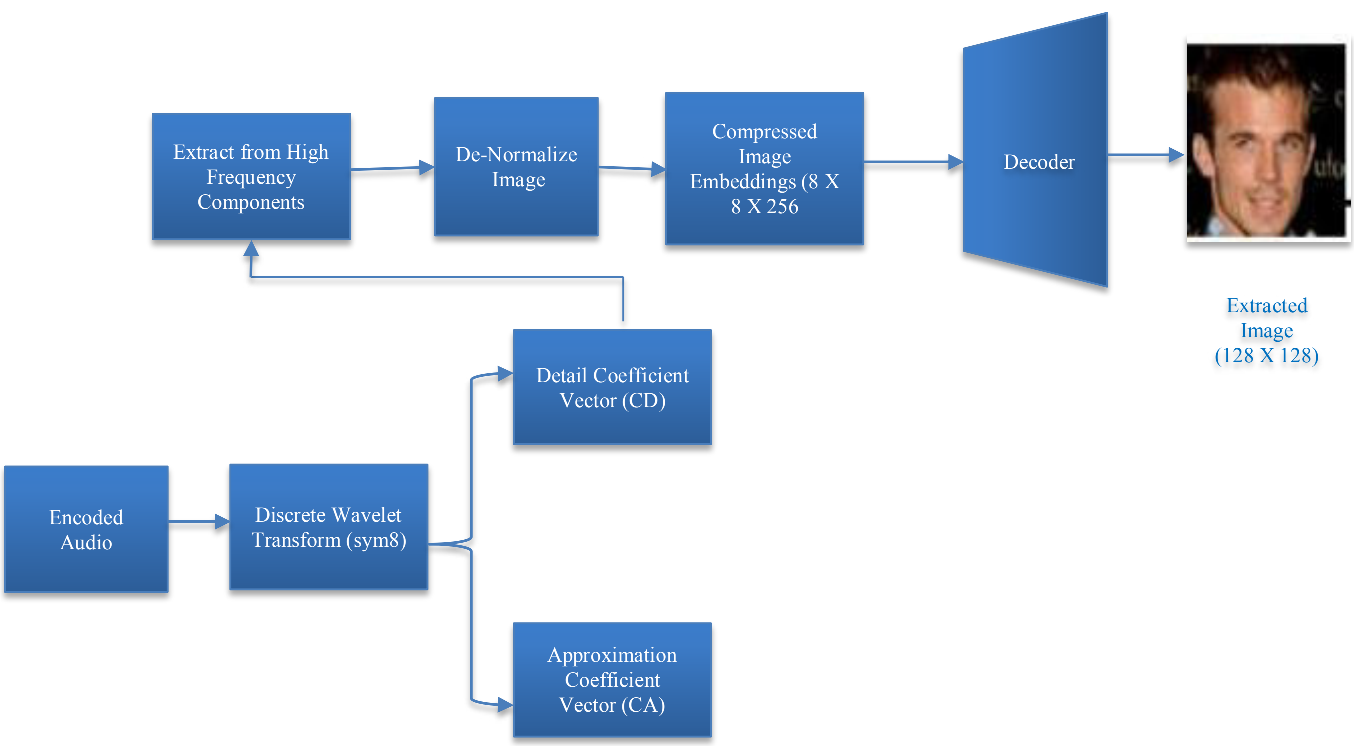


Fig. 2 Decoder architecture: Watermark extraction

3.1. Step 1: Image Encoding

3.1.1. Image Preprocessing

The initial step involves taking an input image with dimensions of 128 x 128 pixels. This image is preprocessed to ensure uniformity and quality before passing it to the image encoder neural network.

3.1.2. Image Embedding Using Neural Network

The preprocessed image is then fed into an image encoder neural network. This neural network is specifically designed to compress the image while retaining its essential features. The output from this network is a set of Image Embeddings that represent a compressed and feature-rich representation of the input image.

3.1.3. Normalization of Image Embeddings

The obtained image embeddings undergo a normalization process to standardize their values. This step ensures that the image data is uniformly scaled, facilitating consistent processing and integration with the audio signal.

3.1.4. Extraction of High-Frequency Components (HF)

From the normalized image embeddings, HF components are extracted. These components represent the fine details and textures of the image, which are crucial for preserving image quality during the embedding process.

3.1.5. Embedding onto High-Frequency (HF) Components

The extracted high-frequency components from the image are then embedded onto the HF components of the audio signal. This embedding ensures that the image information is seamlessly integrated into the audio without compromising the audio quality [22].

3.2. Step 2: Audio Preprocessing

3.2.1. Denoising of Input Audio

Simultaneously, the input audio signal undergoes denoising to eliminate any background noise or artifacts that may interfere with the watermarking process. This ensures a clean and clear audio signal for further processing.

3.2.2. Discrete Wavelet Transform (DWT)

Following denoising, the audio signal is subjected to a Discrete Wavelet Transform (DWT) using the 'sym8' symmetric biorthogonal wavelet. The DWT decomposes the audio signal into two primary components [22]:

3.2.3. Approximation Coefficients (cA)

These coefficients represent the low-frequency components of the audio signal, capturing the overall tone and base sounds.

3.2.4. Detail Coefficients (cD)

The detail coefficients contain the HF components of the audio signal, capturing the nuances and finer details of the sound.

3.3. Step 3: Integration of Image and Audio

3.3.1. Embedding Image onto Audio

The high-frequency components extracted from the image (Embedded on high-frequency components) are integrated with the detail coefficient vector (cD) obtained from the audio signal. This integration ensures that the image watermark is effectively embedded into the high-frequency components of the audio signal.

3.4. Step 4: Audio Reconstruction

3.4.1. Inverse Discrete Wavelet Transform (IDWT)

To reconstruct the audio signal with the embedded image information, the modified detail coefficient vector (cD) is combined with the approximation coefficient vector (cA) using the Inverse Discrete Wavelet Transform (IDWT) with the 'sym8' wavelet.

3.4.2. Watermarked Audio Output

The output from the IDWT is the final encoded audio, which is the original audio signal reconstructed with the embedded image watermark. This watermarked audio maintains the integrity and quality of the original audio while carrying the embedded image information.

3.5. Step 5: Audio Decoding

3.5.1. Discrete Wavelet Transform (DWT)

The initial step in the decoding process involves subjecting the Encoded Audio, which carries the watermarked image data, to a Discrete Wavelet Transform (DWT). This transformation utilizes the 'sym8' symmetric biorthogonal wavelet to decompose the watermarked audio signal into two primary components:

3.5.2. Approximation Coefficients (cA)

These coefficients represent the low-frequency components of the watermarked audio signal, capturing the fundamental tones and base sounds.

3.5.3. Detail Coefficients (cD)

The detail coefficients found within the cD vector encompass the HF components of the audio signal. Importantly, these detail coefficients contain the watermark information embedded during the encoding process.

3.6. Step 6: Watermark Extraction

3.6.1. Extraction of Detail Coefficient Vector (cD)

The Detail Coefficient vector (cD) is isolated from the DWT output, as it houses the watermark data that has been embedded into the HF components of the audio signal.

3.6.2. Transformation to 128x1 Watermark Array

From the extracted cD vector, a transformation is performed to derive a 128x1 watermark array. This array serves as a compressed representation of the embedded image information, maintaining the integrity of the watermark despite its compression.

3.7. Step 7: Image Decoding

3.7.1. Image Decoder Neural Network

The 128x1 watermark array containing the embedded image data is then inputted into an image decoder neural network. This neural network is specifically designed to decode and reconstruct the original 128x128 image from the compressed watermark data.

3.7.2. Neural Network Processing

The Image Decoder processes the 128x1 watermark array through its layers, utilizing learned weights and biases to reconstruct the image. The network leverages its architecture and training to map the compressed watermark data back to the original image features, allowing for accurate image reconstruction.

3.7.3. Reconstructed Image Output

Upon completion of the decoding process, the Image Decoder outputs the reconstructed 128x128 image. This image mirrors the original image that was initially embedded into the audio signal during the encoding phase, effectively recovering the embedded image data.

4. Results and Discussion

We designed the system in Python with the use of libraries such as sound files, librosa, pywt, numpy, PIL etc. The model is trained with a sample image dataset from Kaggle ("https://www.kaggle.com/datasets/farzadnekouei/50k-celebrity-faces-image-dataset").

We trained the model on these images for better embedding and reconstruction of watermark images with maximum features. From these trained images any image can be used as a watermark. We trained the model in Google Colab for 30 epochs and then tested it for random audio files and watermarks.

For evaluating the performance of an autoencoder-based approach in watermarking audio signals with image data, SNR, MC, and BER are tested. The encoding process utilized a neural network to embed image information into audio, while the decoding process employed a neural network to extract and reconstruct the embedded image from the watermarked audio signal. Below are the detailed results and analysis based on the experiments conducted.

4.1. Performance Metrics

4.1.1. Signal-to-Noise Ratio (SNR)

The average SNR obtained from the decoding process was found to be approximately 60.285 dB. This high SNR value indicates that the watermarking process maintains the quality and fidelity of the audio signal well, with minimal distortion introduced during the encoding and decoding processes.

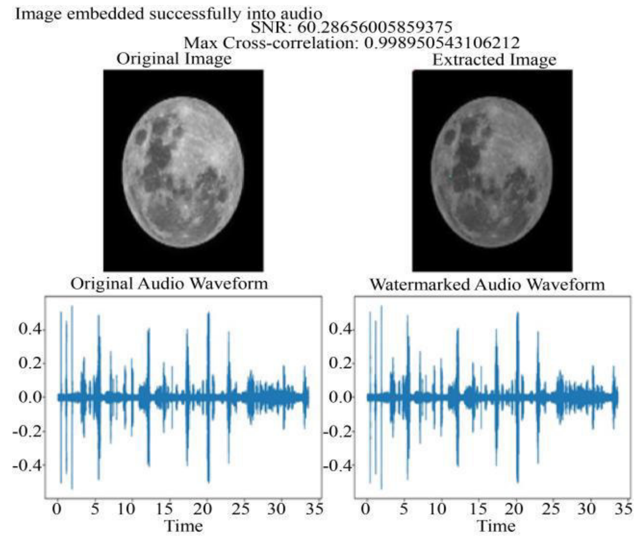


Fig. 3 Calculating SNR and cross correlation for pre-trained image

4.1.2. Maximum Correlation between Images

The average maximum correlation coefficient between the original and reconstructed images was calculated to be approximately 0.99895. This near-perfect correlation demonstrates that the autoencoder-based approach is highly effective in preserving image quality during the watermark extraction process.

4.1.3. Bit Error Rate

It is used to check the transmission accuracy and integrity of the signal. In this research, the embedded signal satisfies the imperceptibility at a high level. However, BER is compromised a little bit as compared to other methods used in the literature. The observed values of BER are in the range of 0.9 to 1.5% for different signals.

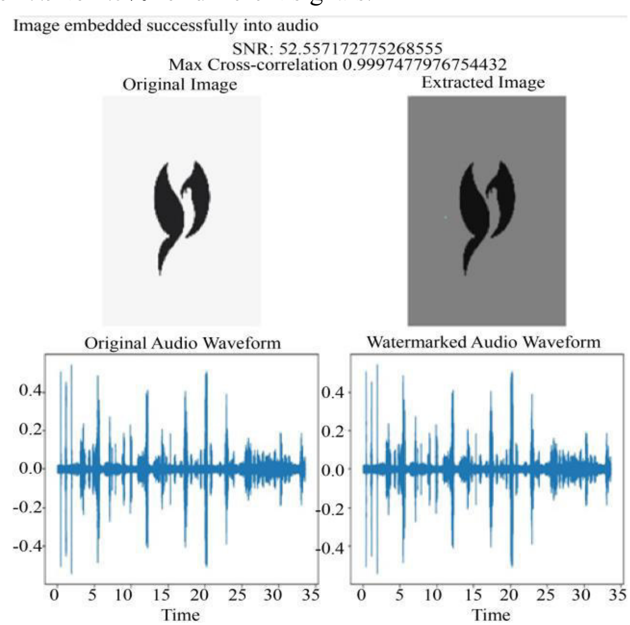


Fig. 4 SNR and Max correlation calculation for the test image

4.2. Loss Trend Analysis

As shown in Figure 5, the loss value shows a decreasing trend over the number of epochs, indicating that the autoencoder model is effectively learning to reconstruct the image from the watermark data embedded in the audio signal. The initial higher loss values gradually decrease, converging to smaller values as the model learns to optimize the reconstruction process.

$$\text{Reconstruction Loss} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{x_i} - \mathrm{x_i}^\wedge\right)^{2}$$

4.3. Training History and Loss Analysis

The training history of the autoencoder model provides valuable insights into the learning process and convergence of the neural network. The loss values, calculated during each epoch of training, are summarized in the table given below as follows:

Table 1: Training and validation loss analysis

Epoch	Training Loss	Validation Loss
1	0.0137	0.0043
2	0.0041	0.0073
3	0.0034	0.0034
4	0.0031	0.0028
5	0.0028	0.0028
10	0.0021	0.0022
15	0.0019	0.0021
20	0.0017	0.0018
25	0.0016	0.0017
30	0.0015	0.0015

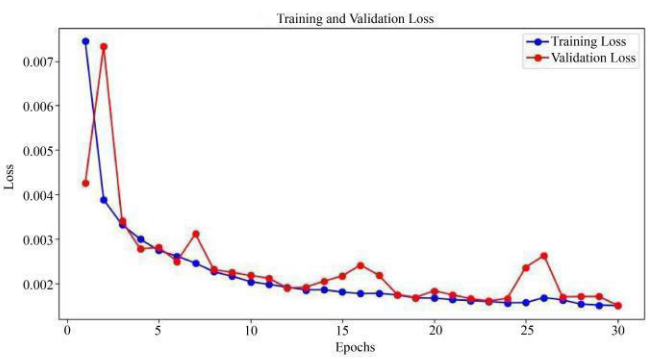


Fig. 5 Testing reconstruction loss using encoder-decoder architecture

4.4. Robustness Analysis

Because of the use of an encoder-decoder model of more than 5 layers, the watermark embedding and extraction process has become so complex that it is very difficult for the attacker to manipulate the signal or extract the watermark information. The embedded signal is robust against the attacks.

This deep learning-based audio watermarking using an encoder-decoder network presents a robust methodology for preserving audio quality while embedding copyrights into audio files. Through meticulous data pre-processing, model design, training, and evaluation, methodological rigor is ensured, laying a strong foundation for reliable watermarking processes. Leveraging tools like Librosa for audio processing and TensorFlow/Keras for deep learning, the proposed approach accurately reconstructs watermark images from processed audio signals, showcasing its effectiveness in content authentication and copyright protection. The study underscores promising applications in digital rights management, offering avenues for enhanced copyright protection and content authentication in various domains. We tested the approach on random samples of audio and watermark images and calculated SNR and Cross Correlation parameters. The observation shows that the approach is better than the previously used approaches in terms of imperceptibility and security of audio signal with a watermark. It is observed that BER is quite high; however, the quality of the audio is not degraded because of the use of bits from the high frequency components of the audio signal, which are inaudible to the human audio system. Also, the reconstruction of the watermark has shown good results. DWT-based signal decomposition with the help of the ‘sym8’ symmetric biorthogonal wavelet to decompose the watermarked audio signal into two primary components, cA and cD, made it easier to embed and extract the watermark. After calculating the parameters the average values obtained are SNR: 51.55, MC: 0.9989. The average BER is ≥ 1.2%. Though the approach gives a little higher BER as compared to the methods discussed in the literature it preserves the original quality of audio signal and imperceptibility. For testing the imperceptibility, 3 different samples of original and embedded audio with 15 different persons were tested. After careful listening to the samples, not a single user was able to identify any significant difference between the original and embedded signal.

5. Conclusion

A novel approach using autoencoders is presented in this paper. Audio watermarking using an encoder-decoder network presents a robust methodology for preserving audio quality while embedding copyrights into audio files. Through meticulous data pre-processing, model design, training, and evaluation, methodological rigor is ensured, laying a strong foundation for reliable watermarking processes. Leveraging tools like Librosa for audio processing and TensorFlow/ Keras for deep learning, the proposed approach accurately reconstructs watermark images from processed audio signals, showcasing its effectiveness in content authentication and copyright protection. The study underscores promising applications in digital rights management, offering avenues for enhanced copyright protection and content authentication in various domains. We tested the approach on random samples of audio and watermark images and calculated SNR and Cross Correlation parameters. The observation shows that the approach is better than the previously used approaches in terms of imperceptibility and security of audio signal with a watermark. It is observed that BER is quite high; however, the quality of the audio is not degraded because of the use of bits from the high frequency components of the audio signal, which are inaudible to the human audio system. The studies reviewed in this paper showcase the robustness of various watermarking techniques in safeguarding audio files against copyright infringement. Ranging from hash-based detection systems like DEAR [2], AudioQR [4], SVM Based [6], and HADES [17] to other deep learning-based approaches [21] and copyright-embedded watermarking schemes, each method presents unique advantages and challenges in achieving reliable copyright protection across decentralized music sharing platforms and other digital audio environments. Our approach demonstrates enhanced robustness against common signal processing attacks and has proven its effectiveness compared to previously employed methodologies. The experimental results indicate that the encoder-decoder-based method is more effective in protecting intellectual property rights in audio and music files. Overall, the findings presented in this paper contribute to the ongoing discourse on audio watermarking for copyright protection and provide valuable insights into the field of digital media security and intellectual property rights enforcement.

5.1. Future Work

Through this paper, we demonstrated the effectiveness of watermarking and deep learning with the traditional approaches for copyright protection of digital audio files. However, topics such as optimization-based watermarking, decentralized watermarking schemes, and the integration of blockchain technology for copyright enforcement present exciting opportunities for further exploration and development. Overall, this paper underscores the importance of continuous research and innovation in audio watermarking techniques to address the evolving challenges of copyright protection in digital audio content. By advancing the state of the art in watermarking technology, researchers and practitioners can contribute to the creation of more secure, robust, and efficient solutions for preserving the integrity and ownership of audio content in today's digital landscape. Looking ahead, future research directions may focus on optimizing the autoencoder architecture and extending real-time processing capabilities, further advancing the field of audio watermarking.

Funding Statement

All Authors and co-authors have seen and agree with the contents of the manuscript and we have NO affiliations with or involvement in any organization or entity with any financial interest. No funding is provided for the work by any organization. Hence, there is no financial interest to report.

References

[1] Wanxing Huang, "Copyright Protection of Music Multimedia Works Fused with Digital Audio Watermarking Algorithm," *International Journal of Grid and High Performance Computing*, vol. 15, no. 2, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Chang Liu et al., "DeAR: A Deep-Learning-Based Audio Re-Recording Resilient Watermarking," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, pp. 13201-13209, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Kosta Pavlovic, Slavko Kovacevic, and Igor Durovic, "Speech Watermarking Using Deep Neural Networks," *2020 28th Telecommunications Forum*, Belgrade, Serbia, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Xinghua Qu et al., "AudioQR: Deep Neural Audio Watermarks for QR Code," *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence AI for Good*, pp. 6192-6200, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Amita Singha, and Muhammad Ahsan Ullah, "Development of an Audio Watermarking with Decentralization of the Watermarks," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3055-3061, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xiangyang Wang, Wei Qi, and Panpan Niu, "A New Adaptive Digital Audio Watermarking Based on Support Vector Regression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2270-2277, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[7] Maha Charfeddine et al., "Audio Watermarking for Security and Non-Security Applications," *IEEE Access*, vol. 10, pp. 12654-12677, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Margarita N. Favorskaya, and Andrey I. Pakhirka, "Adaptive HVS Objectivity-Based Watermarking Scheme for Copyright Protection," *Procedia Computer Science*, vol. 192, pp. 1441-1450, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kasorn Galajit et al., "Semi-Fragile Speech Watermarking Based on Singular-Spectrum Analysis with CNN-Based Parameter Estimation for Tampering Detection," *APSIPA Transactions on Signal and Information Processing*, vol. 8, pp. 1-13, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Prayoth Kumsawat, Kitti Attakitmongcol, and Arthit Srikaew, "Digital Audio Watermarking for Copyright Protection Based on Multiwavelet Transform," *Intelligence and Security Informatics: European Conference*, Esbjerg, Denmark, pp. 155-164, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[11] Zulfiqar Ali, Muhammad Imran, and Mansour Alsulaiman, "An Automatic Digital Audio Authentication/Forensics System," *IEEE Access*, vol. 5, pp. 2994-3007, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[12] Chong Wang et al., "An Audio Watermarking Scheme with Neural Network," *Advances in Neural Networks - ISNN 2005: Second International Symposium on Neural Networks*, Chongqing, China, pp. 795-800, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[13] Farah Deeba et al., "Digital Watermarking Using Deep Neural Network," *International Journal of Machine Learning and Computing*, vol. 10, no. 2, pp. 277-282, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Seyed Mostafa Pourhashemi, Mohammad Mosleh, and Yousof Erfani, "A Novel Audio Watermarking Scheme Using Ensemble-Based Watermark Detector and Discrete Wavelet Transform," *Neural Computing and Applications*, vol. 33, no. 11, pp. 6161-6181, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Yassine Himeur, and Bachir Boudraa, "Secure and Robust Audio Watermarking System for Copyright Protection," *2012 24th International Conference on Microelectronics*, Algiers, Algeria, pp. 1-4, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[16] Zihan Wang et al., "Data Hiding with Deep Learning: A Survey Unifying Digital Watermarking and Steganography," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, pp. 2985-2999, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Muhammad Rasyid Redha Ansori et al., "HADES: Hash-Based Audio Copy Detection System for Copyright Protection in Decentralized Music Sharing," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 2845-2853, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Lukas Tegendal, "*Watermarking in Audio Using Deep Learning*," Master Thesis, Linköping University, pp. 1-50, 2019. [Google Scholar] [Publisher Link]

[19] Farida Aboelezz, "*Watermarking Audio Files with Copyrights*," Bachelor Thesis, the German University in Cairo, pp. 1-82, 2022. [Google Scholar]

[20] Preeti Garg, and R. Rama Kishore, "An Efficient and Secured Blind Image Watermarking Using ABC Optimization in DWT and DCT Domain," *Multimedia Tools and Applications*, vol. 81, pp. 36947-37964, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Saoussen Ben Jabra, and Mohamed Ben Farah, "Deep Learning-Based Watermarking Techniques Challenges: A Review of Current and Future Trends," *Circuits, Systems, Signal Process*, vol. 43, pp. 4339-4368, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Farnaz Arab, Salwani Mohd Daud, and Siti Zaiton Hashim, "Discrete Wavelet Transform Domain Techniques," *2013 International Conference on Informatics and Creative Multimedia*, Kuala Lumpur, Malaysia, pp. 340-345, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[23] Numpy.correlate, Numpy. [Online]. Available: https://numpy.org/doc/stable/reference/generated/numpy.correlate.html

[24] 50K Celebrity Faces Image Dataset, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/farzadnekouei/50k-celebrity-faces-image-dataset

[25] Nguyen Chi Sy, Ha Hoang Kha, and Nguyen Minh Hoang, "An Efficient Robust Blind Watermarking Method Based on Convolution Neural Networks in Wavelet Transform Domain," *International Journal Machine Learning Computing*, vol. 10, no. 5, pp. 675-684, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] N.V. Lalitha, Ch. Srinivasa Rao, and P.V.Y. Jaya Sree, "Localization of Copy-Move Forgery in Speech Signals through Watermarking Using DCT-QIM," *International Journal of Electronics and Telecommunications*, vol. 65, no. 3, pp. 527-532, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Arashdeep Kaur, Malay Kishore Dutta, and Jiri Prinosil, "General Regression Neural Network Based Audio Watermarking Algorithm Using Torus Automorphism," *2018 41st International Conference on Telecommunications and Signal Processing*, Athens, Greece, pp. 1-4, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[28] Hadi Latifpour, Mohammad Mosleh, and Mohammad Kheyrandish, "An Intelligent Audio Watermarking Based on KNN Learning Algorithm," *International Journal of Speech Technology*, vol. 18, pp. 697-706, 2015. [CrossRef] [Google Scholar] [Publisher Link]