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.

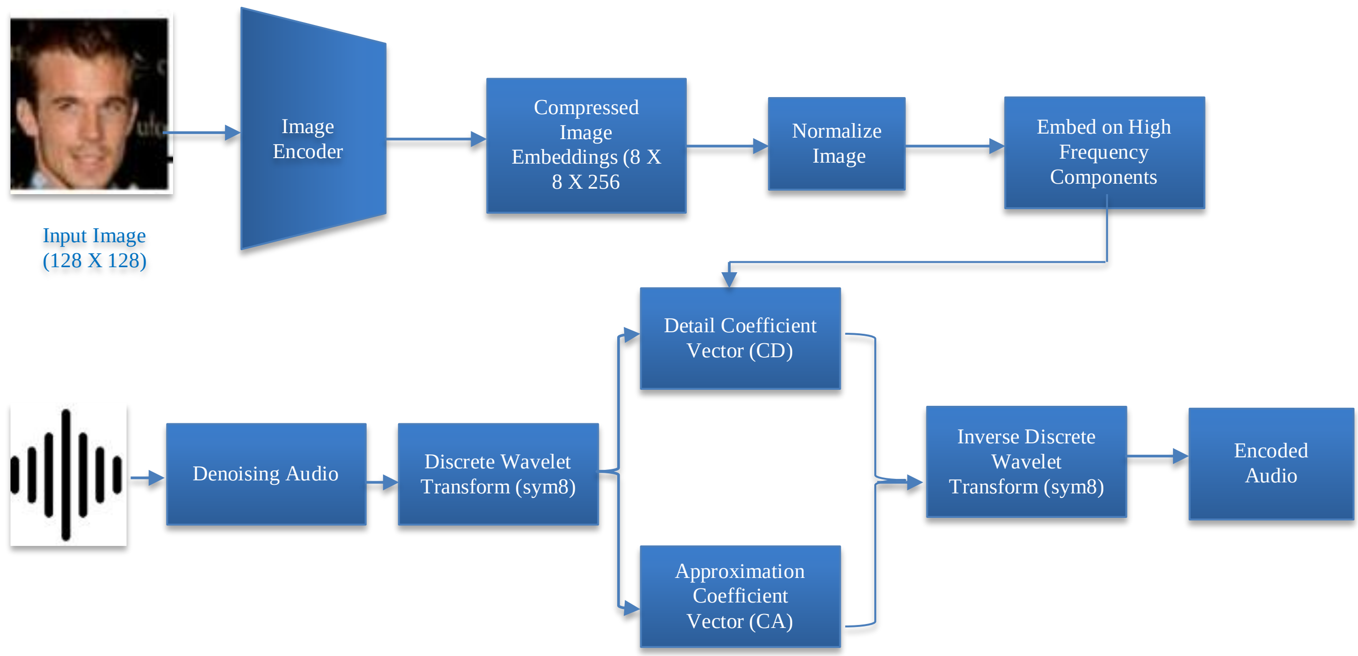

Fig. 1 Encoder architecture: Watermark embedding

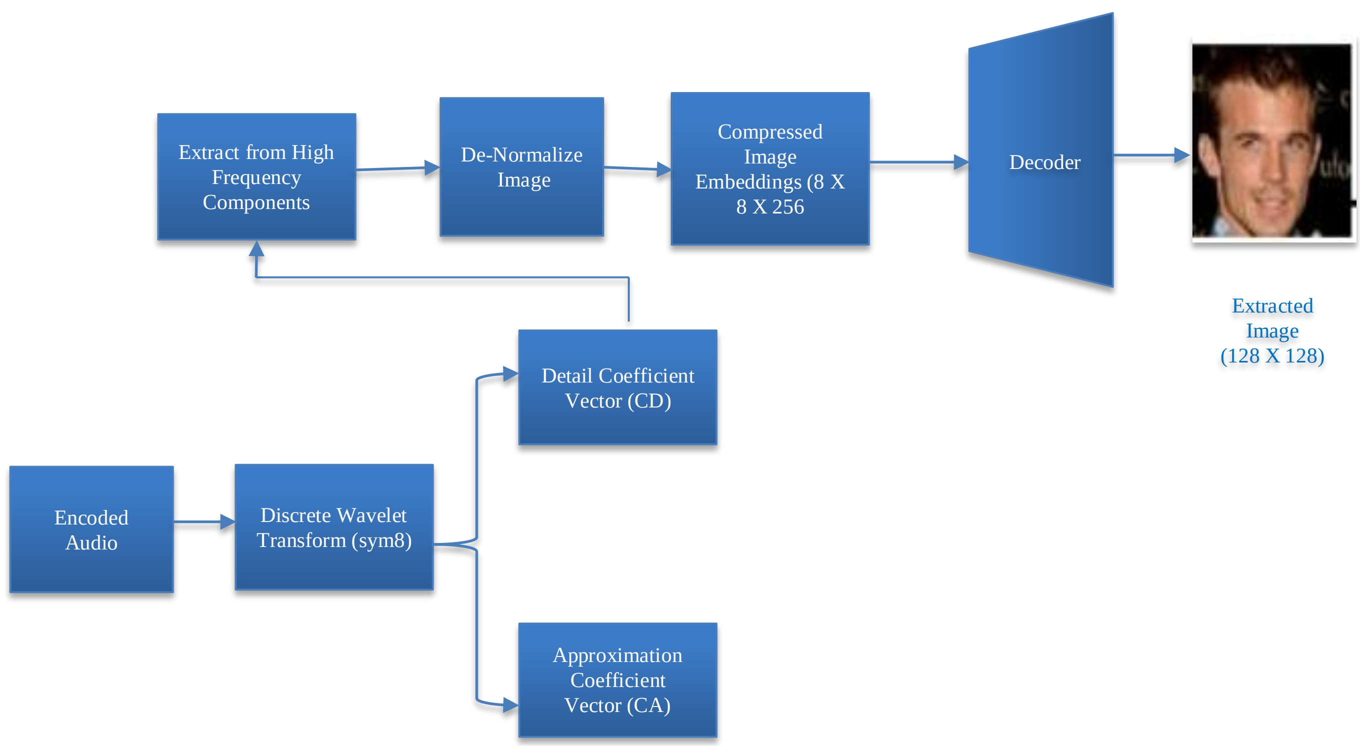

Fig. 2 Decoder architecture: Watermark extraction

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.

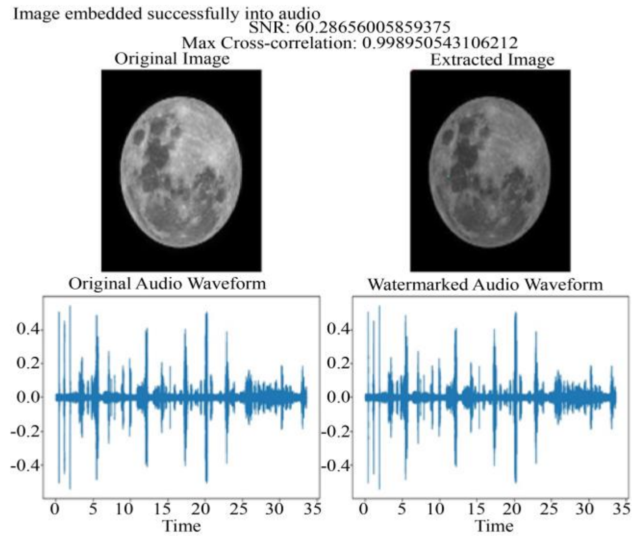

Fig. 3 Calculating SNR and cross correlation for pre-trained image

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.

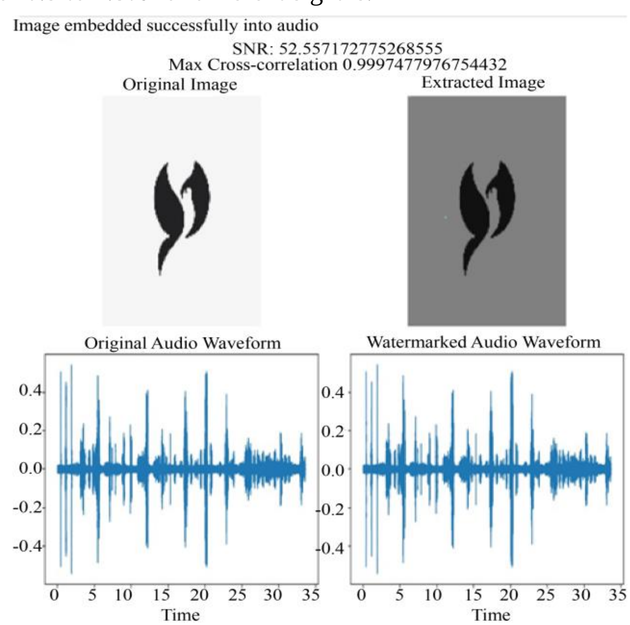

Fig. 4 SNR and Max correlation calculation for the test image

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

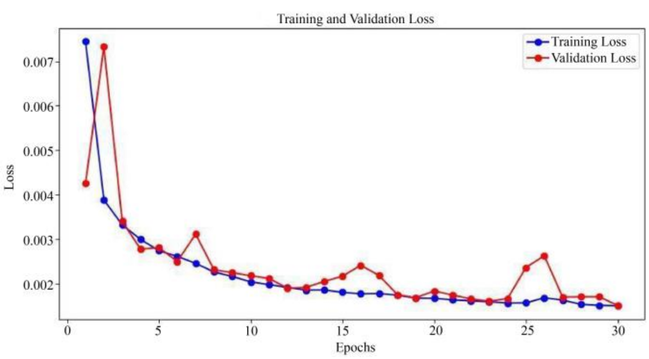

Fig. 5 Testing reconstruction loss using encoder-decoder architecture

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 $\geq$ 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]