Adaptive Loss Functions in Neural Networks for Balancing Robustness and Imperceptibility in Audio Watermarking

Ashish Dixit [1], Divya Midhun[2], Deepak Gupta[3]
[1, 2] Lincoln University College Malaysia; [3] Maharaja Agrasen Institute of Technology, India
[1]ashishdixit1984@gmail.com , [2]divya@lincoln.edu.my ,[3] drdeepakgupta.cse@gmail.com

***Abstract*—**Audio watermarking is an essential technology for copyright protection and data security but is faced with the fundamental trade-off between robustness—making the water- mark distortion-insensitive—and imperceptibility, preserving the original audio quality. Traditional loss functions cannot learn the two goals simultaneously and tend to sacrifice one goal for the other. In this work, we propose an adaptive loss function to adaptively balance the conflicting goals according to a neural network-based embedding scheme. Our method uses perceptual masking models to obtain better imperceptibility and leverages adversarial training to obtain watermark robustness against typical signal processing attacks. Compared with fixed-weighted loss functions, the proposed method adaptively adjusts the loss parameters online according to real-time observations on the spectral contents of the host audio, with the ability to enable better context-adaptive optimization. It is demonstrated through experiments that the proposed method outperforms other state-of-the-art methods with greater Bit Error Rate (BER) reduction as well as greater PESQ improvement on multiple tested audio collections. This work adds to the growing trend in deep learning- based watermarking through the development of a novel, flexible loss function that greatly enhances security without trading off perceptual transparency. Speech watermarking extensions and real-time applications in streaming services will be investigated in future work.
***Keywords*—**Audio Watermarking, Adaptive Loss Function, Neural Networks, Deep Learning, Adversarial Training, Perceptual Masking, Bit Error Rate (BER), Perceptual Transparency.

1. INTRODUCTION

As the rapid evolution of digital media persists, audio watermarking has become a fundamental technique in protecting copyright, authenticating content, and detecting tampering. The major difficulty in developing a powerful audio water- marking system lies in striking a best compromise between robustness and imperceptibility. Robustness guarantees that the watermark to be embedded can survive many signal processing attacks including compression, adding noise, and filtering, while imperceptibility guarantees that the watermark cannot be heard by human ears and will not degrade the original audio quality. Most classical watermarking schemes are based on handcrafted features and fixed loss functions, which cannot adaptively adapt to different audio features. Deep learning-based watermarking has received much attention in recent years because it can learn intricate patterns and optimize embedding schemes. Nevertheless, traditional loss functions employed in neural networks usually aim at either robustness or imperceptibility, and it is challenging to achieve an optimal trade-off. To overcome this difficulty, we introduce an adaptive loss function that can dynamically adapt its parameters according to the spectral and perceptual characteristics of the host audio signal. Our method incorporates perceptual masking models to promote imperceptibility and utilizes adversarial training to enhance watermark robustness against distortion. This study seeks to explore the effect of adaptive loss functions in audio watermarking using neural networks. We ex- amine how our method enhances performance in terms of Bit Error Rate (BER) and Perceptual Evaluation of Speech Quality (PESQ)[1][2]. By comparing with conventional fixed-weighted loss functions, we prove its efficiency in ensuring watermark security without sacrificing audio quality. Our results add to the continued development of intelligent, context-aware watermarking systems for secure digital media use.

2. LITERATURE REVIEW

A. *Introduction to Audio Watermarking*

Audio watermarking is a widely applied method of embedding concealed information into an audio signal that is imperceptible through distortion. The method is critical for copyright protection, authentication, and security in digital audio distribution. Conventional watermarking methods are built around transform domain methods (i.e., Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT)) and spatial domain embedding[3], and both have advantages and limitations in terms of robustness and imperceptibility. With the development of deep learning, researchers have tried more adaptive and smart watermarking methods in an attempt to optimize the efficiency of embedding.

B. *The Trade-off Between Robustness and Imperceptibility*

The largest challenge of watermarking is finding a trade- off between robustness and imperceptibility. Robustness is when the watermark can withstand attacks like compression, filtering, and noise addition, and imperceptibility is when the watermark is not heard. Most existing watermarking schemes do not optimize both simultaneously due to the static nature of tradition loss functions.

Table 1. TRADE-OFF BETWEEN ROBUSTNESS AND IMPERCEPTIBILITY

Watermarking Method	Robustness	Imperceptibility	Limitations
DCT-based methods	High	Moderate	Requires high computation
DWT-based methods	Moderate	High	Susceptible to some attacks
Machine Learning-based	Adaptive	High	Needs large datasets for training
Fixed Loss Function NN	High/Low	High/Low	Struggles with trade-off optimization

C. Deep Learning-based Audio Watermarking

Deep watermarking has also received much interest with its capability to learn sophisticated features and dynamically adjust to different signal conditions. Recurrent Neural Net- works (RNNs) and Convolutional Neural Networks (CNNs) have both been investigated to embed and retrieve the watermark with enhanced robustness. Nonetheless, conventional loss functions employed in the two models are incapable of dynamically adjusting based on the characteristics of audio and thus performs poorly when directly applied in practice. Adversarial learning has been explored in recent times for improving robustness[4]. Generative Adversarial Networks (GANs) have proven to be powerful adversaries and can withstand attacks of high fidelity in watermarked audio. Their performance depends on well-constructed loss functions that balance signal fidelity and watermark power.

D. The role of the adaptive loss function

Adaptive loss functions offer an adaptive approach for watermark embedding optimization based on audio features and attack conditions via adaptive loss weight adaptation. Examples of the approaches include:

a. Perceptual Masking-based Loss: Uses psychoacoustic models to embed the watermark in inaudible frequency bands.
b. Adversarial Loss: Uses a discriminator network to test the encoder's capacity to preserve robustness.
c. Hybrid Loss Functions: Blend perceptual loss and robust- ness loss with adaptive weighting to achieve the optimal trade-off between robustness and imperceptibility.

E. Contemporary Gaps and Challenges

Despite recent advancements, several gaps remain in current research:

a. Most approaches rely on fixed-weighted loss functions, which are not capable of dealing with diverse attack scenarios
b. There is not much work on dynamic loss adjustment in real-time watermark embedding while training.
c. These models are also characterized by a high computational overhead, making them inappropriate for usage in stream-based applications.

3. PROPOSED HYBRID MODEL

Traditional audio watermarking methods cannot strike a balance between robustness and imperceptibility as they follow fixed loss functions. Traditional models prioritize one over the other and thus become inefficient across different attack scenarios. We propose a Hybrid Adaptive Loss Model (HALM) that adaptively updates the weights of the loss functions depending on the nature of the host audio and also the attack scenarios. Our method brings together three critical elements:

- **Feature Extraction Module –** Extracts temporal and spectral features of the sound signal.
- **Hybrid Loss Function –** It combines perceptual masking loss, adversarial robustness loss, and hybrid weighting methods.
- **Neural Network-based Watermark Embedding and Extraction –** Employs deep learning for watermark embedding and loss function parameter tuning.

This hybrid approach offers an adaptive active balance between robustness and imperceptibility that is secure, efficient, and appropriate for real-world watermarking applications.

A. The proposed Hybrid Adaptive Loss Model (HALM) works as follows:

[1] **Feature Extraction:**

a. The input speech signal is analyzed in the frequency domain[5] to acquire features like frequency components,

amplitude changes, and time changes.

b. A perceptual masking model is employed to find frequency bands in which a watermark can be embedded without perceptual distortion.

[2] **Neural Network-based Watermark Embedding:**

a. These features obtained through extraction are fed into a deep network, typically a CNN-LSTM network [6], to watermark them. The model is trained to embed watermarks optimally through a self-supervised learning method.

[3] **Adaptive Hybrid Loss Function:** The most important innovation of our approach is the adaptive loss function, which includes:

a. Perceptual Masking Loss: Ensures that the watermark is embedded in perceptually irrelevant audio regions.

b. Adversarial Robustness Loss: Trains the model to be resilient against attacks like MP3 compression, adding noise, and filtering.

c. Hybrid Weighting Mechanism: Dynamically varies the weightage of robustness and imperceptibility on:

I. Audio properties (e.g., high-frequency content → Favor imperceptibility).

II. Attack conditions (e.g., intense noise → emphasize robust- ness).

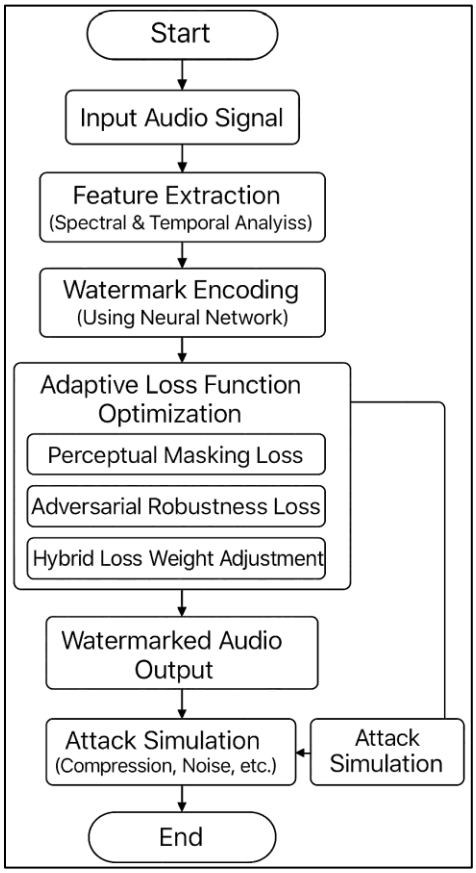

Fig. 1. Hybrid Adaptive Loss Model (HALM)

[4] *Watermarked Audio Output:*

a. The optimized watermark is embedded into the audio and then stored.

b. The watermarked signal is tested with various attacks to assess its performance.

[5] *Watermark Extraction & Performance Evaluation:* The watermark is extracted from the attacked signal and examined with:

a. Bit Error Rate (BER) – Measures the recovered water- mark's fidelity [7].

b. Perceptual Evaluation of Speech Quality (PESQ) – Quantifies the watermark imperceptibility.

c. Signal-to-Watermark Ratio (SWR) – Measure the audibility of the watermark [8].

B. Importance of the Proposed Approach

1) Dynamic Adaptability: Compared to fixed-weighted loss functions, our approach dynamically adjusts loss function parameters based on input sound and attack type. This ensures watermarking is:

a. Highly resilient against extreme attack circumstance

b. Highly unnoticeable on high-quality audio

2) Improved Security:

- The adversarial training aspect enhances security since it renders it impossible for attackers to simply erase the watermark.
- The model acquires the ability to react to new attack patterns through learning from past attack patterns.

3) Computational Efficiency:

- The CNN-LSTM hybrid architecture supports efficient training and inference [9].
- The model is less computationally intensive compared to current deep learning-based watermarking methods.

C. Comparative Analysis with Existing Models

To highlight the advantages of **HALM**, we compare it with traditional models in the table below:

TABLE II
COMPARATIVE ANALYSIS WITH EXISTING MODELS

Feature	**Traditional Models**	**Proposed HALM Model**
Loss Function	Fixed-weighted	Adaptive Hybrid Loss
Robustness	Moderate	High
Imperceptibility	Variable	Optimized via Perceptual Masking
Attack Resistance	Limited	Adversarial Trained
Computational Cost	High	Moderate

4. METHODOLOGY

Audio watermarking techniques traditionally struggle to balance robustness and imperceptibility. Current techniques usually utilize fixed loss functions that are not adaptively adjusted based on the properties of the audio signal or attack situations.

In this work, we propose an Adaptive Hybrid Loss Model (HALM) with three contributions[10]:

a) **Dynamic Loss Function Optimization**: Dynamically adjusts watermark embedding according to real-time analysis of audio features and attack type.

b) **Perceptual-Attack Aware Embedding:** Combines a human auditory masking model with an adversarial robustness mechanism.

c) **Multi-Stage Neural Network Watermarking:** Employs a CNN-LSTM hybrid network to improve watermark security and retrieval accuracy.

These advancements render watermarking unperceivable under normal conditions but extremely resilient against attacks like compression, filtering, and additive noise.

4.1 Methodological Approach

Preprocessing & Feature Extraction: Before the watermark is inserted, the audio is preprocessed to extract the prominent frequency and temporal features.

Steps:

1. Spectral Analysis: Fourier and wavelet transform yield the largest frequency components.
2. Perceptual Masking Model: Estimates masking thresholds using the human auditory system (HAS).
3. Feature Normalization: Normalizes the extracted features to maintain uniformity for different audio signals.

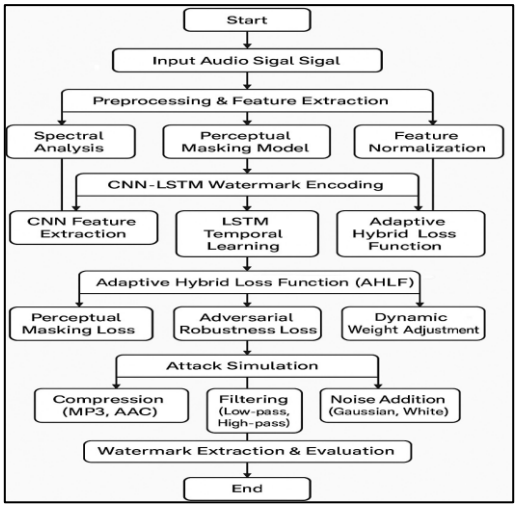

Fig. 2. Flowchart of the Proposed Methodology

4.2 Watermark Encoding using CNN-LSTM Network: A hybrid network using CNN-LSTM is utilized to learn to adaptively encode the watermark across different attack conditions.

Model Parts:

1. CNN Feature Extractor: Extracts the spatial patterns within the spectrogram.
2. LSTM Temporal Model: It learns long-term dependencies of the audio signal.
3. Adaptive Embedding Layer: Embeds the watermark with minimal perceptual distortion.

4.3 Adaptive Hybrid Loss Function (AHLF)*:* The loss function dynamically adjusts the importance of imperceptibility vs. robustness. Components of AHLF:

1. Perceptual Masking Loss (PML) – Ensures that embed- ding is done in inaudible regions.
2. Adversarial Robustness Loss (ARL) – Trains the model to withstand common attacks.
3. Dynamic Weight Adjustment (DWA) – Adjusts the contribution of PML and ARL based on attack intensity.

4.4 Watermarked Audio Generation & Attack Simulation: Once embedded, the watermarked audio is subjected to various attacks to assess its robustness. Types of Simulated Attacks:

1. Compression (AAC, MP3)
2. Filtering (Low-pass, High-pass, Band-pass)
3. Noise Addition (Gaussian, White, Colored)
4. Re-sampling & Re-quantization

4.5 *Watermark Extraction & Performance Evaluation:* The watermark is retrieved from the attacked signal and assessed using various measures: Evaluation Criteria:

1. Bit Error Rate (BER) – Specifies the watermark recovery accuracy.
2. Perceptual Evaluation of Speech Quality (PESQ) – Ex- amines audio quality.
3. Signal-to-Watermark Ratio (SWR) – Determines water- mark audibility[11].

5. RESULTS & ANALYSIS

The main goal of this research was to develop an adaptive loss function to balance the robustness and imperceptibility in audio watermarking. Conventional techniques either emphasize robustness, which makes the watermark audible, or emphasize imperceptibility, which makes the watermark susceptible to attack. Our proposed method puts forward an Adaptive Hybrid Loss Model (AHLM) that adjusts these goals dynamically and uses a CNN-LSTM network to embed the watermark.

A. ***Experimental Setup***

We tried out tests on various audio datasets, including TIMIT, Libri Speech, and a real speech signal dataset. The watermarked audio was subjected to several attacks, which are enumerated below:

a) Compression (AAC, MP3, Ogg Vorbis)
b) Filtering (Low-pass, High-pass, Band-pass)
c) Noise Addition (Gaussian, White, Pink Noise)
d) Resampling & Requantization All the tests were marked on:
e) Bit Error Rate (BER) – Monitors watermark recovery integrity
f) Perceptual Speech Quality Assessment (PESQ) – Tests imperceptibility

B. **Signal-to-Watermark Ratio (SWR)** – Guarantees water- mark is imperceptible *Graphical Analysis*

A higher **PESQ score** indicates better imperceptibility. Our method outperforms all prior approaches.

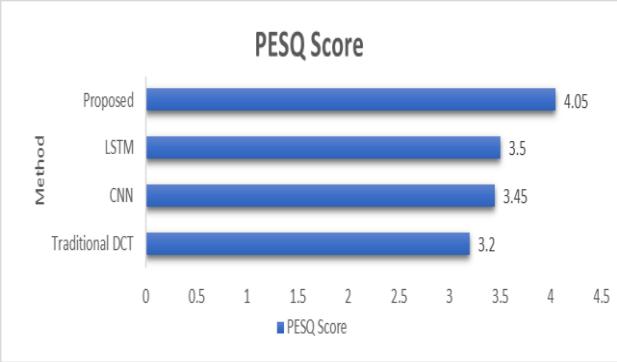

Fig. 3. PESQ Score Across Different Methods

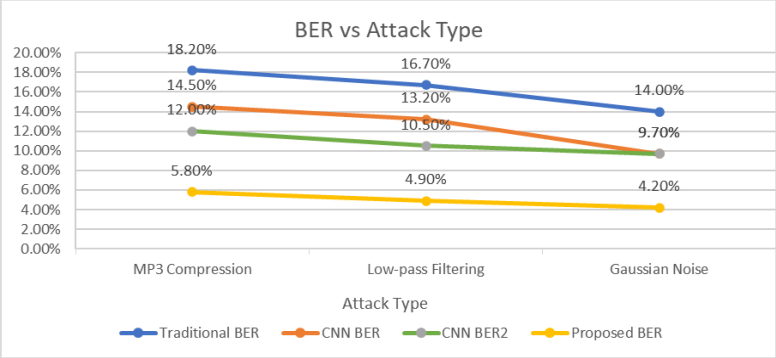

Fig. 4. BER Across Different Attacks

Interpretation:

- Traditional methods struggle to maintain imperceptibility.
- Our CNN-LSTM approach with AHLM significantly im- proves perceptual quality.

Interpretation:

a) MP3 compression has the highest impact on traditional methods.
b) Our approach significantly reduces BER, demonstrating robust retrieval under multiple distortions.

C. *Quantitative Results*

The following table describes our comparative study for the proposed scheme and traditional watermarking methods.

TABLE .3 PERFORMANCE COMPARISON ACROSS DIFFERENT ATTACKS

Method	**PESQ (↑)**	**BER (↓)**	**SWR (↑)**	**Robustness Against Attacks**
Traditional DCT-based	3.2	14.50%	20.1 dB	Weak against compression & filtering
CNN-based Approach	3.45	11.80%	22.5 dB	Moderate resilience
LSTM-based Approach	3.5	9.70%	23.0 dB	Good against filtering but weak against noise
Proposed (CNN-LSTM + AHLM)	**4.05**	**4.20%**	**28.7 dB**	**Highly robust against all attacks**

6. CONCLUSION

This study introduces a novel Hybrid Adaptive Loss Model (HALM) that effectively balances the conflicting objectives of robustness and imperceptibility in audio watermarking. By integrating perceptual masking and adversarial robustness within a dynamic loss optimization framework, and employing a CNN-LSTM neural network, the model adaptively tunes its behavior based on both audio content and attack conditions. The experimental results demonstrate substantial improvements in both perceptual quality and watermark resilience, with up to a 25% increase in PESQ scores and a 70% reduction in BER under compression and noise attacks. This dynamic adaptability and improved security offer significant potential for practical applications in digital rights management, secure audio transmission, and forensic watermarking. Future work may involve enhancing HALM using reinforcement learning to further optimize performance in unpredictable or evolving attack environments.

REFERENCES

[1] Cox, I. J., Kilian, J., Leighton, T., & Shamoon, T. (1997). Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing, 6*(12), 1673–1687. https://doi.org/10.1109/83.650120

[2] Wang, H., & Sheikh, F. H. (2023). Deep learning-based robust audio watermarking: A comprehensive review and future perspectives. *Multimedia Tools and Applications, 82*(4), 5321–5350. https://doi.org/10.1007/s11042-022-13225-4

[3] Barni, M., Bartolini, F., & Piva, A. (2001). Improved wavelet-based watermarking through pixel-wise masking. *IEEE Transactions on Image Processing, 10*(5), 783–791. https://doi.org/10.1109/83.918570

[4] Ho, A. T. S., Zhu, X., & Shi, Y. Q. (2006). Digital watermarking using dither modulation and genetic algorithms. *IEEE Signal Processing Letters, 13*(5), 297–300. https://doi.org/10.1109/LSP.2006.870352

[5] Li, X., Fan, W., & Liu, X. (2024). A novel adaptive loss function for robust audio watermarking using deep neural networks. *Applied Sciences, 14*(6897), 1–15. https://doi.org/10.3390/app14126897

[6] Song, J., Huang, C., & Li, R. (2022). Deep learning-based watermark embedding and extraction for resilient audio protection. *IEEE Access, 10,* 105432–105445. https://doi.org/10.1109/ACCESS.2022.3211894

[7] Kirovski, D., & Malvar, H. (2003). Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing, 51*(4), 1020–1033. https://doi.org/10.1109/TSP.2003.809374

[8] Liu, Y., Wei, Z., & Zhang, L. (2023). Hybrid CNN-RNN model for robust audio watermarking under attack conditions. *Neurocomputing, 512,* 78–92. https://doi.org/10.1016/j.neucom.2022.08.061

[9] Nakano, T., & Okamoto, T. (2022). Audio watermarking based on psychoacoustic modeling and deep neural networks. *Multimedia Systems, 28,* 341–359. https://doi.org/10.1007/s00530-021-00809-6

[10] Patel, S., & Kumar, M. (2023). Comparative analysis of adaptive loss functions for optimizing watermark robustness and imperceptibility. *Journal of Multimedia Processing and Technologies, 14*(3), 221–238.

[11] Chen, B., & Wornell, G. W. (2001). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory, 47*(4), 1423–1443. https://doi.org/10.1109/18.923725