UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
RPW-EEG:An Unified Framework for Robust and Practical Watermark of EEG

Permalink
https://escholarship.org/uc/item/96h9c291

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors
Qin, Tianyang
Yi, Hangjie
Qian, Jingsheng
et al.

Publication Date
2025

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

RPW-EEG:An Unified Framework for Robust and Practical Watermark of EEG

Tianyang Qin (241050021@hdu.edu.cn) **Hangjie Yi (yihangjie@hdu.edu.cn)**
Jingsheng Qian (jingshengqian@hdu.edu.cn) **Xuanyu Jin (xy_jin@hdu.edu.cn)**
Honggang Liu (lhg@hdu.edu.cn) **Wanzeng Kong * (kongwanzeng@hdu.edu.cn)**
School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

Abstract

With the rapid growth of the metaverse and advancements in extended reality (XR) technologies, brain-computer interfaces (BCIs) are expanding their applications beyond the medical field into various consumer industries. However, this broader adoption has raised significant concerns about the privacy and security of EEG data. To address this challenge, we propose the RPW-EEG framework, which embeds copyright information as perturbations to enhance the security and traceability of EEG data, while ensuring the robustness and usability of the data. The framework adopts an encoder-decoder architecture for end-to-end training, incorporating a noise layer to enhance the stability and anti-attack capabilities of the watermark data. Additionally, to prevent the loss of task-related features in EEG data, we introduce a plug-and-play fine-tuning module that restores these features within the watermark-embedded signals. Experimental results demonstrate that RPW-EEG outperforms baseline models in terms of watermark quality, with watermark extraction accuracy consistently exceeding 61% under various attack scenarios. Moreover, the classification accuracy for task paradigms reaches 88.5%, indicating that RPW-EEG effectively balances data copyright protection with the preservation of EEG data usability and practicality in analysis.
Keywords: EEG; robust watermarking; brain-computer interfaces; privacy protection; copyright protection.

Introduction

Electroencephalogram (EEG) data is a highly sensitive biological signal(Buzsáki, Anastassiou, & Koch, 2023) that carries information closely related to an individual's mental state, cognitive function, and health conditions. Not only does it reflect a person's psychological activity, but it also reveals deeper insights into emotional tendencies, interests, health issues, and even personality traits(Huang et al., 2021), making it a repository of private information. However, with the ongoing advancements in technologies such as the metaverse and extended reality , the use of Brain-Computer Interfaces(Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002) has extended beyond the medical field(Pan, Cai, Huang, He, & Li, 2023) and is gradually infiltrating consumer industries, such as entertainment. This shift raises significant concerns regarding ownership and security related to personal privacy data, particularly during the use and sharing of EEG data(Yi, Qian, Ming, & Kong, 2024; Yi, Ming, Liu, & Kong, 2024; Zhang, Yi, & Kong, 2024). The critical issue of ensuring data ownership and security has thus become an urgent problem that needs addressing.

In this context, robust watermarking technology, as an important tool for tracking and verifying the source of data, has been widely applied(Xiao, Zhang, Hua, Xia, & Weng, 2024; Xiong, Han, Yang, & Shi, 2022; H. Wu, Liu, Yao, & Zhang, 2021; You, Wang, Zhu, & Kwong, 2022). These techniques embed hidden watermark signals into host data, making them nearly imperceptible to human observers while retaining the ability to trace the data's origin. However, traditional robust watermarking methods face the challenge of poor robustness when applied to EEG data. EEG data typically contains a large amount of noise and individual variability, and its high-frequency components and weak signals are prone to interference from external factors, leading to poor stability of the watermark in EEG data. Therefore, improving the robustness of watermarking in EEG data has become a key challenge in the field.

EEG signals contain a wealth of task-related features, such as motor imagery, ERP, and other time-frequency-space features, which are crucial for downstream tasks like classification(D. Wu, Jiang, & Peng, 2022; Huebner, Verhoeven, Mueller, Kindermans, & Tangermann, 2018; Shanechi, 2019). However, due to the low signal-to-noise ratio of EEG data, watermark embedding can result in the loss of task-related features, negatively impacting downstream applications. Specifically, in most task-related classification tasks, the presence of robust watermarks tends to decrease classification accuracy, which contradicts the goal of data protection. The reason behind this phenomenon is that previous studies often treat watermark embedding and the usability of host data as separate issues(Li, Yang, Wang, Zhang, & Wen, 2024; Zhu, Kaplan, Johnson, & Fei-Fei, 2018). To effectively address this challenge, watermark embedding techniques should minimize damage to the multi-dimensional features of EEG data while ensuring copyright protection, enabling both data usability and copyright preservation to coexist. Therefore, a critical challenge in designing EEG watermarking techniques is to strike a balance between watermark embedding and the preservation of task feature integrity.

To address these challenges, we propose the RPW-EEG framework, which utilizes an encoder-noise layer-decoder architecture for end-to-end training. The implementation of the framework is divided into two stages. In Encoder-Decoder Training stage, the encoder and decoder are jointly trained in an end-to-end manner, incorporating a noise layer to enhance the robustness of the watermark. The noise layer incorporates various transformations across the time, frequency, and spatial domains, commonly employed during the preprocessing and data augmentation stages of EEG signals, enabling the decoder to accurately extract copyright information from both normal and attacked watermarked data. Since the embedded watermark can be regarded

* Corresponding author.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

as a form of adversarial perturbation(Hou et al., 2023), benign adversarial fine-tuning can enable the model to retain task-relevant information within the watermark data. Therefore, during the adversarial task fine-tuning stage, the robust watermark is further refined into an adversarial watermark.

To evaluate the performance of RPW-EEG, we assessed it across three key aspects: watermark quality, data usability, and robustness. In terms of watermark quality, RPW-EEG outperformed baseline models in both noisy and noise-free environments, as measured by PSNR and SSIM metrics. Regarding data usability, RPW-EEG achieved a task prediction accuracy of 88.5%, surpassing the baseline models. In terms of robustness, RPW-EEG demonstrated over 61% watermark extraction accuracy under various attack methods (e.g., time-frequency domain, spatial domain), showcasing its superior resistance to attacks and confirming its reliability in practical applications. Our contributions are as follows:

- We propose a traceable EEG data protection scheme, RPW-EEG, which introduces a specialized noise layer and a plug-and-play fine-tuning module to generate traceable adversarial watermark data with both robustness and practicality.
- To the best of our knowledge, we are the first to achieve a balance between usability and robustness in EEG data watermark embedding.
- We evaluated our approach using two publicly available motion imagination datasets, and the experimental results demonstrate that RPW-EEG outperforms existing technologies in terms of EEG data quality, robustness, and usability.

Method

Architecture Overview

To ensure copyright and privacy protection of EEG data, we propose an end-to-end traceable EEG watermarking adversarial model, RPW-EEG. As illustrated in Figure 1, the implementation of RPW-EEG consists of two distinct stages. In the first stage, the encoder and decoder are jointly trained in an end-to-end manner, with a noise layer introduced during training to enhance the resilience of the EEG watermark against adversarial attacks. In the second stage, the encoder and decoder are fine-tuned to optimize the robust watermark into an adversarial watermark, aiming to preserve the integrity of the original task-specific features while enhancing the practical utility of the watermark data. During the inference phase, only the finalized, fine-tuned watermark encoder and decoder are utilized to achieve copyright traceability.

Encoder To address the complex spatial features of EEG data, the encoder structure has been specifically designed. The encoder's core is composed of transition layers, which integrate convolution, batch normalization, and ReLU activation, along with a watermark embedding layer that fuses multi-scale convolutional features. The transition layers employ a default 3×3 kernel with a stride of 1 and padding of 1, ensuring the capture of detailed EEG data features. The embedding layer utilizes multi-scale convolution, processing EEG data in parallel with convolutional kernels of varying sizes to capture frequency components across different scales. By introducing convolutional kernels of different sizes in the embedding layer, redundancy is enhanced, improving the robustness of the embedded information against attacks during transmission. The architecture of the encoder is illustrated in Figure 2.

During the encoding process, copyright information is seamlessly integrated into the feature representations at various stages. The incorporation of concatenation operations and residual connections ensures that features from preceding layers are fully utilized. This design not only mitigates the issue of gradient vanishing but also enhances the model's ability to learn latent patterns within EEG signals. Consequently, the encoder's architecture addresses the common gradient vanishing problem in deep networks while facilitating the efficient propagation of watermark features, which is particularly advantageous for embedding high-dimensional and complex EEG data. Additionally, the transition layers ensure that the watermark information is effectively embedded without significantly altering the original EEG data.

To ensure that the embedded watermark does not significantly alter the characteristics of the original EEG signals, the loss function L_{MSE} is employed to evaluate the point-wise difference between the EEG data with embedded watermark and the original EEG data. This loss function facilitates the assessment of the encoder's precision in preserving the features of the original EEG data. The L_{MSE} loss is calculated as follows:

$$L_{\mathrm{MSE}}(E_{\mathrm{or}}, E_{\mathrm{we}}) = \frac{\|E_{\mathrm{or}} - E_{\mathrm{we}}\|^2}{N_{\mathrm{ch}} \times N_{\mathrm{t}}} \tag{1}$$

where E_{or} represents the original EEG data, E_{we} represents the watermarked EEG data, and N_{ch} and N_{t} denote the number of EEG channels and sampling points, respectively.

To further enhance the realism and completeness of reconstructed EEG data, We utilize the adversarial optimization process between the Encoder and discriminator. The adversarial optimization process leverages both cross-entropy loss and adversarial loss to facilitate dynamic adversarial optimization between the encoder and the classifier. The cross-entropy loss L_{GAN} is defined as follows:

$$L_{\mathrm{GAN}} = \log(1 - A(E_{\mathrm{or}})) - \log(A(E_{\mathrm{we}})) \tag{2}$$

where A represents the classifier. This loss function aims to confuse the classifier, making it difficult to distinguish between the original EEG data and the protected EEG data.

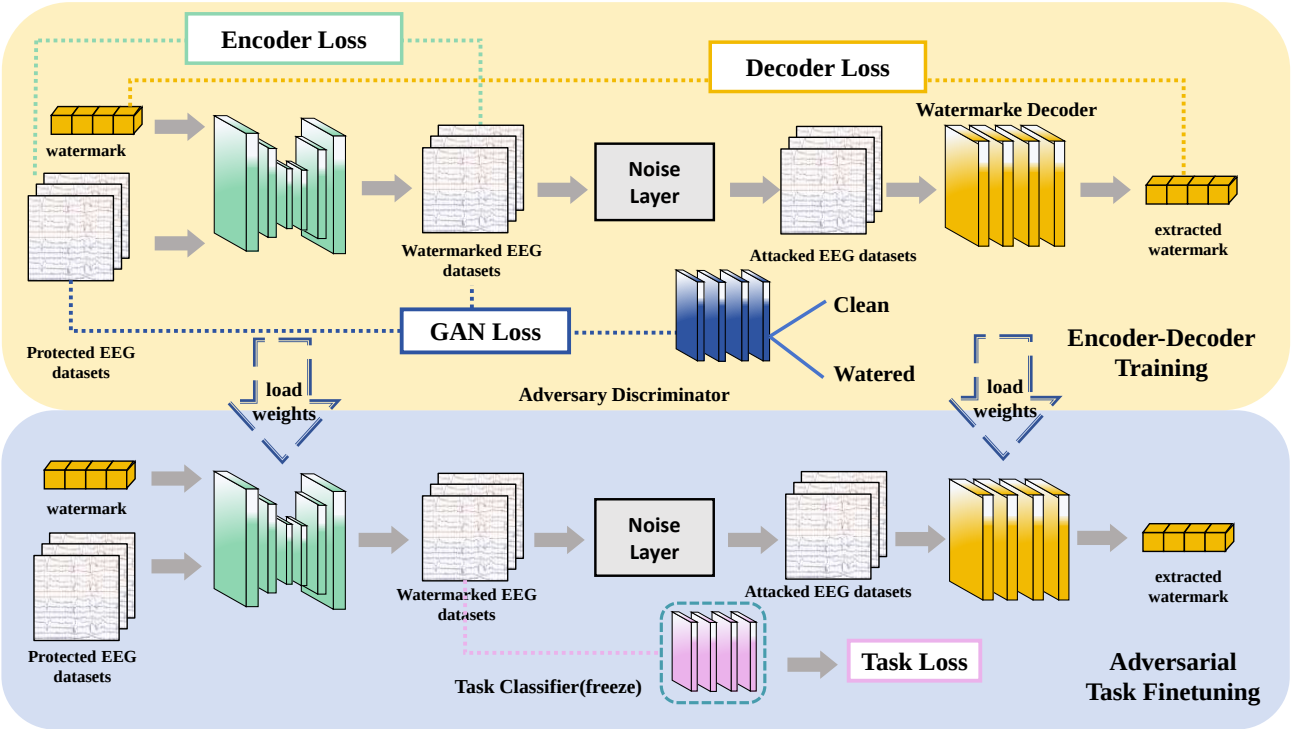

Figure 1: The framework diagram of RPW-EEG.

Noise Layer To enhance the robustness and attack resistance of the protected EEG data, a multifunctional noise network is employed, incorporating EEG data augmentation techniques to simulate various potential attack scenarios. This approach is designed to prevent attackers from rendering the watermark ineffective by enhancing the data without altering its intrinsic characteristics, thereby significantly improving the stability and attack resistance of traceable adversarial examples. Within the noise network, attacks on the EEG data are simulated by introducing noise in the temporal, spatial, and frequency domains.

$$E_{\text{no}} = N_{\text{no}}(E_{\text{we}}) \tag{3}$$

Here, N_{no} represents the noise addition process. The network consists of three types of noise layers, each contributing a specific type of noise during the training cycle. These include Gaussian noise in the time domain, time-axis and symbol inversion; displacement, transformation, and filtering in the frequency domain; and symmetry and dropout processing across channels in the spatial domain.

Decoder The core design of the decoder lies in accurately extracting and recovering the watermark copyright information from the watermarked EEG data. The decoder is composed of a series of convolutional layers, batch normalization, and ReLU activation transition layers. It is responsible for progressively extracting the copyright information from the watermarked EEG data, followed by the application of a global average pooling layer that combines the spatial dimensions of the feature maps into a single value. Finally, a fully connected layer converts these compressed features into a vector of the copyright length, representing the predicted copyright information. This process involves mapping from a high-dimensional feature space to a low-dimensional binary information space, ensuring that the output matches the format of the original secret information. Furthermore, the L2 norm loss between the original secret information and the decoded information is utilized to enhance the accuracy of information recovery.

$$L_{\text{water}}(W_{\text{in}}, W_{\text{out}}) = \frac{\|M_{\text{in}} - M_{\text{out}}\|_2^2}{L} \tag{4}$$

Adversarial Task Finetuning Enhancing task-relevant features helps preserve the functional integrity of EEG signals. By focusing on neural activity patterns closely associated with specific tasks, we ensure that even after watermark embedding, the signal still accurately represents the brain's state during task performance. This not only aids in maintaining the interpretability of the signal, thereby enhancing its practical utility and imperceptibility, but also ensures that the watermarking process does not compromise the signal's functionality. In light of this, we propose a plug-and-play adversarial fine-tuning process, which adjusts the robust watermark into an adversarial watermark with the aim of restoring the task-specific features of the EEG data through the classifier. To preserve the integrity of these features, we freeze the task classifier T and update the encoder to force the recovery of task-related features for each sample. The corresponding loss function is defined as follows:

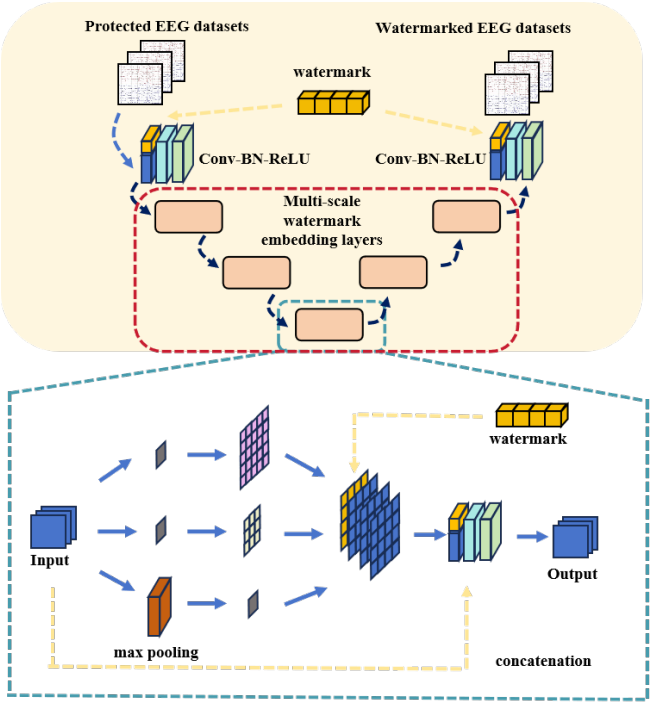

Figure 2: The structure diagram of the encoder.

$$L_{\mathcal{T}} = \mathcal{F}(\mathcal{T}(x_w), y) = \mathcal{F}(\mathcal{T}(En(E_{\text{or}})), y) \tag{5}$$

After fine-tuning, we obtain the final encoder and decoder, where the embedded watermark is not only a robust watermark but also an adversarial watermark. To achieve the invisibility and robustness of the protected data, the objective of the model training is to minimize a loss function with various weighted terms. In summary, the loss functions in the encoding-decoding stage L_1 and the fine-tuning stage L_2 are expressed as follows:

$$L_1 = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{water}} + \lambda_3 L_{\text{GAN}} \tag{6}$$

$$L_2 = \lambda_4 L_{\text{MSE}} + \lambda_5 L_{\text{water}} + \lambda_6 L_{\text{T}} \tag{7}$$

Experiment

Dataset

This study utilizes the following two publicly available EEG datasets:

1) Dataset I(Tangermann et al., 2012): This MI dataset is from the BCI Competition IV, dataset 2a. It includes 22-channel EEG data from 9 subjects, each performing four different MI tasks (left hand, right hand, both feet, and tongue). Each task consists of 144 trials across two sessions.

2) Dataset II(Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004): This MI dataset comes from 109 subjects. Each subject performs four MI tasks while recording 64-channel EEG data. In this study, task 2 (imagining the opening and closing of the left or right fist) is used, with 45 EEG trials across three sessions for each subject.

Experiment Settings

To ensure a fair comparison, all experiments were conducted on an NVIDIA GeForce RTX 3090 GPU using the PyTorch framework implemented in Python. The binary code length in the model was set to $L = 30$ bits. During the initial stage of training the pretrained model, the loss function weights were set as $\lambda_1 = 0.7$, $\lambda_2 = 1$, and $\lambda_3 = 0.003$. In the subsequent model fine-tuning stage, the weights were adjusted to $\lambda_4 = 0.7$, $\lambda_5 = 1$, and $\lambda_6 = 0.7$. The learning rate in all stages was 10^{-3}, and the batch size was 64. To recover task-specific features during the fine-tuning stage, we employed three pretrained task classifiers (EEGNet, DeepConvNet, and ShallowConvNet) to enhance the detectability of task feature preservation and the adversarial transferability of the model(Schirrmeister et al., 2017; Lawhern et al., 2018).

Evaluation Metrics

To evaluate the quality of protected EEG data, this study employs two metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM)(Horé & Ziou, 2010). PSNR quantifies the difference between the original EEG data and the protected EEG data, while SSIM measures the structural similarity between the two, offering a comprehensive perspective on data fidelity.

For the accuracy of copyright extraction, we introduce the Bit Accuracy (BA) metric. BA is defined as the ratio of correctly identified watermark bits to the total number of watermark bits, providing an accurate assessment of the success rate in recovering copyright information from the protected EEG data.

Furthermore, to ensure that the generated EEG watermark data preserves the integrity of task-relevant features, this study employs Task Accuracy (TA) as a metric. Task Accuracy reflects the effectiveness of the EEG watermark data in performing the intended cognitive or physiological tasks, ensuring that even after the watermark embedding, the data retains its original functional utility.

Results

To validate the effectiveness of the proposed method, extensive experimental studies were conducted on two publicly available datasets. These experiments included watermark quality assessment, data usability evaluation, and robustness testing. Through this comprehensive series of rigorous tests, the study aims to systematically and thoroughly investigate the performance and characteristics of the proposed framework from multiple perspectives.

Watermark Data Quality Testing

Through the visual presentation of the differences between the original EEG data and the watermark data, we further focus on the changes that occur in the original data during the watermark embedding process. This helps assess the impact of watermark embedding on the structure of the original data. Figure 3 clearly illustrates these differences, and compared

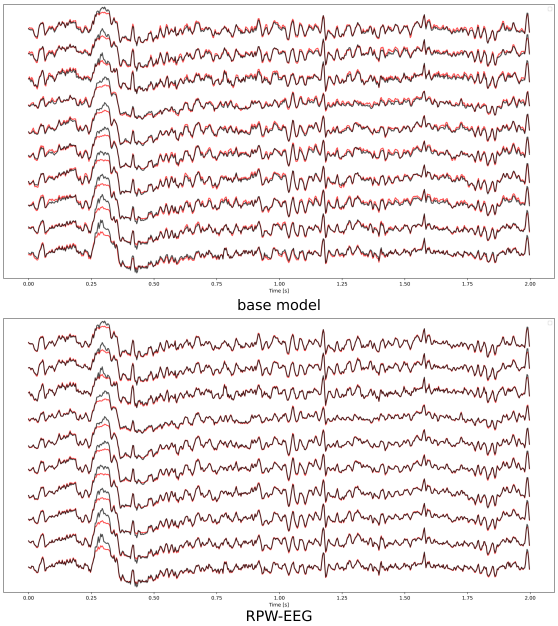

Figure 3: The differences between the original EEG (red) and the watermark EEG (black).

to the baseline models, the watermark data generated by our RPW-EEG shows significantly higher quality.

Table 1: Comparison of indicators under noise-free conditions.

Clean	Dataset I			Dataset II		
	PSNR	SSIM	BA	PSNR	SSIM	BA
Hidden	21.157	0.847	0.829	7.898	0.811	0.754
Dual	22.484	0.908	0.801	7.942	0.898	0.789
Ours	**25.256**	**0.918**	**0.896**	**11.864**	**0.913**	**0.803**

To comprehensively evaluate the performance of our proposed watermark model in terms of reconstructed data quality and copyright information recognition integrity, we conducted a detailed quantitative analysis. The experiment used objective quality assessment metrics, including PSNR and SSIM, to measure the quality of the reconstructed data. Additionally, bit accuracy (BA) was used to assess the accuracy of the extracted watermark copyright information. We define the watermark model without noise training as Ours and the model trained with noise layers is labeled as Noised. Our model was compared with two baseline models: the classic watermark model HiDDeN(Zhu et al., 2018) and the Dual model(Li et al., 2024), which employs a multi-layer residual convolution design.

Table 2: Comparison of indicators under noise conditions.

Noised	Dataset I			Dataset II		
	PSNR	SSIM	BA	PSNR	SSIM	BA
Hidden	17.732	0.758	0.711	7.511	0.851	0.717
Dual	18.631	**0.848**	0.741	7.686	0.867	0.709
Ours	**21.021**	0.827	**0.814**	**9.919**	**0.895**	**0.762**

The specific experimental results are presented in Table 1and Table 2, which compare the evaluation metrics of each model on Dataset I and Dataset II under noise-free (Clean) and noisy (Noised) conditions. Under noise-free training conditions, our model, along with the two baseline models, demonstrates outstanding performance in both PSNR and SSIM metrics. This can be attributed to the absence of noise constraints in the training environment, allowing the network to focus more on optimizing the decoding quality of EEG data without compromising watermark extraction accuracy. When noise constraints are introduced during training, our model significantly outperforms the two baseline models in PSNR and also achieves excellent performance in watermark extraction accuracy. This indicates that our model exhibits higher robustness and adaptability in balancing data quality and watermark extraction accuracy. It highlights its superior performance in mitigating the effects of noise interference. This enhanced balance provides a more reliable solution for protecting data integrity and copyright in practical applications.

Watermark Data Usability Testing

In the encoder-decoder training stage, the model focuses on optimizing the robustness of watermark embedding and extraction, which may lead to the loss of some task-relevant features in the original data. However, the introduction of the second-stage fine-tuning process effectively restores and strengthens these task-relevant features.

To evaluate the impact of adversarial training during fine-tuning, we conducted comparative experiments using fixed-task paradigm models, including EEGNet, DeepConvNet, and ShallowConvNet. The Adversarial Task Fine-tuning stage enables the model to relearn and adjust to the specific requirements of the task, ensuring that watermark protection does not compromise the effectiveness and usability of the original data for the target task.

As illustrated in Table 3, the fine-tuning process significantly improves the classification accuracy of the protected data across all three standard classification models. This underscores the importance of the second-stage adjustment in achieving a balance between robust watermarking and the preservation of task-relevant features.

This approach not only enhances the model's versatility and adaptability but also offers an effective solution to the potential conflict between data protection and task performance. Specifically, the pretraining stage enables the model to learn robust watermark embedding and extraction capabilities, while the fine-tuning stage ensures the recovery and enhancement of task-relevant features within the context of specific tasks. This dual-stage training strategy not only improves the overall performance of the model but also provides a feasible solution for data protection and task execution in complex application scenarios.

Table 3: Comparison of performance metrics for different methods.

	Dataset I			Dataset II		
	EEGNet	DeepConv	ShallowConv	EEGNet	DeepConv	ShallowConv
Raw Data	0.967	0.987	0.954	0.810	0.810	0.875
Hidden	0.655	0.633	0.609	0.715	0.711	0.761
Dual	0.623	0.664	0.608	0.725	0.768	0.731
Ours_no	0.607	0.679	0.735	0.618	0.748	0.663
Ours	**0.813**	**0.865**	**0.827**	**0.736**	**0.795**	**0.819**

Table 4: Robustness Experiment of EEG Watermark Data Against Attack Transformations.

Dataset I	**Time Domain**			**Frequency Domain**			**Spatial Domain**	
	Gaussian noise	Time reversal	Sign reversal	Frequency shift	Fourier transform	Band-pass filter	Channel symmetry	Channel dropout
DCT	0.44	**0.74**	0.18	0.48	0.51	0.24	0.51	0.42
SIRD	0.53	0.54	0.36	0.18	0.38	0.52	0.49	0.53
Hidden	0.47	0.53	0.55	0.57	0.58	0.55	0.62	0.53
Dual	0.48	0.52	0.48	0.49	0.45	0.51	0.58	0.50
Ours_no	0.53	0.55	0.59	0.53	0.51	0.56	0.56	0.55
Ours	**0.61**	0.65	**0.67**	**0.61**	**0.66**	**0.69**	**0.68**	**0.63**

Robustness testing

To evaluate the robustness of the proposed model in noisy environments, it is essential that copyright information can still be correctly extracted even when EEG data is subjected to various attacks. To this end, we designed a series of robustness testing experiments aimed at simulating real-world scenarios that could impact watermark recognition, thereby verifying the model's performance in practical applications. These experiments include not only simulations under noise-free conditions but also rigorous tests under noisy conditions, providing a comprehensive assessment of the model's anti-interference capabilities.

In this set of experiments, we focused on validating the robustness of watermark embedding in EEG data. Various attack methods, as shown in the tables, were employed for testing. These attacks cover different types in the time, frequency, and spatial domains, such as Gaussian noise addition, time reversal, symbol inversion, frequency shifting, Discrete Cosine Transform (DCT), channel swapping, and channel loss. The inclusion of these attacks ensures that the test results reflect the true performance of the model.

For a comprehensive and systematic comparison, we selected DCT and SIRD in the frequency domain as supplementary experimental methods and compared them with two classic baseline models (HiDDeN and Dual). All models were tested on EEG datasets with various types of attacks under the same conditions, ensuring the comparability and fairness of the results. During the evaluation, bit accuracy (BA) was used as the primary metric to measure watermark extraction accuracy.

As shown in Table 4, under different types of attacks, our model demonstrated significant robustness in copyright information extraction. In particular, when compared with common frequency-domain attacks (such as DCT and SIRD), our model's BA surpassed that of the baseline models. Despite the presence of complex attacks, our model effectively protects and extracts copyright information while maintaining data quality.

Conclusion

This paper introduces an innovative framework, RPW-EEG, designed to tackle the challenges of data usability and robustness in EEG data copyright traceability. In the first phase of training, the model focuses on generating high-quality watermark data and enhancing robustness. In the second phase, adversarial fine-tuning optimizes the watermark, transforming it into a robust adversarial watermark while preserving task-specific features, thereby improving the usability of the watermark-embedded data. Experimental results demonstrate that RPW-EEG outperforms existing models in key metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Bit Accuracy (BA), with exceptional performance under noisy conditions. Additionally, through adversarial fine-tuning, RPW-EEG effectively recovers task-related features, ensuring that the protected EEG data remains efficient for downstream tasks. Thus, RPW-EEG offers an effective solution for EEG data copyright protection and provides a novel technical approach to ensuring data security and functional preservation.

Acknowledgements

This work was supported by National Natural Science Foundation of China (62471169), Key Research and Development Project of Zhejiang Province (2023C03026, 2021C03001, 2021C03003), and supported by Key

Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province (2020E10010).

The complete implementation code of this study has been open-sourced and is available on `https://github.com/qintianyang/RPW-EEG` to facilitate research reproducibility and method dissemination.

References

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2023, September). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes.

Horé, A., & Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *20th international conference on pattern recognition, icpr 2010, istanbul, turkey, 23-26 august 2010*.

Hou, Y., Guo, Q., Huang, Y., Xie, X., Ma, L., & Zhao, J. (2023). Evading deepfake detectors via adversarial statistical consistency. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 12271–12280).

Huang, K., Peng, D., Yao, Z., Xia, J., Zhang, B., Liu, H., … Huang, Y. (2021). Cathodic plasma driven self-assembly of heas dendrites by pure single fcc feconimncu nanoparticles as high efficient electrocatalysts for oer. *Chemical Engineering Journal*, *425*, 131533.

Huebner, D., Verhoeven, T., Mueller, K.-R., Kindermans, P.-J., & Tangermann, M. (2018). Unsupervised learning for brain-computer interfaces based on event-related potentials: Review and online comparison [research frontier]. , *13*(2).

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013.1-056013.17.

Li, M., Yang, Z., Wang, T., Zhang, Y., & Wen, W. (2024). Dual protection for image privacy and copyright via traceable adversarial examples. *IEEE Transactions on Circuits and Systems for Video Technology*, 1-1.

Pan, J., Cai, H., Huang, H., He, Y., & Li, Y. (2023). Multiple scale convolutional few-shot learning networks for online p300-based brain–computer interface and its application to patients with disorder of consciousness. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-16.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, *51*(6), 1034–1043.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., … Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, *38*(11), 5391–5420.

Shanechi, M. M. (2019). Brain–machine interfaces from motor to mood. *Nature Neuroscience*, *22*.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., … others (2012). Review of the bci competition iv. *Frontiers in neuroscience*, *6*, 55.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical neurophysiology*, *113*(6), 767–791.

Wu, D., Jiang, X., & Peng, R. (2022). Transfer learning for motor imagery based brain–computer interfaces: A tutorial. *Neural Networks*, *153*, 235-253.

Wu, H., Liu, G., Yao, Y., & Zhang, X. (2021). Watermarking neural networks with watermarked images. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(7), 2591-2601.

Xiao, X., Zhang, Y., Hua, Z., Xia, Z., & Weng, J. (2024). Client-side embedding of screen-shooting resilient image watermarking. *IEEE Transactions on Information Forensics and Security*, *19*, 5357-5372.

Xiong, L., Han, X., Yang, C.-N., & Shi, Y.-Q. (2022). Robust reversible watermarking in encrypted image with secure multi-party based on lightweight cryptography. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(1), 75-91.

Yi, H., Ming, Y., Liu, D., & Kong, W. (2024). Time-frequency jointed imperceptible adversarial attack to brainprint recognition with deep learning models. In *2024 ieee international conference on multimedia and expo (icme)* (pp. 1–6).

Yi, H., Qian, J., Ming, Y., & Kong, W. (2024). Independent components time-frequency purification with channel consensus against adversarial attack in ssvep-based bcis. *IEEE Signal Processing Letters*.

You, J., Wang, Y.-G., Zhu, G., & Kwong, S. (2022). Truncated robust natural watermarking with hungarian optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(2), 483-495.

Zhang, Y., Yi, H., & Kong, W. (2024). A privacy-preserving brainprint recognition system based on feature homomorphic encryption. In *2024 international joint conference on neural networks (ijcnn)* (pp. 1–10).

Zhu, J., Kaplan, R., Johnson, J., & Fei-Fei, L. (2018). Hidden: Hiding data with deep networks. In *Computer vision – eccv 2018: 15th european conference, munich, germany, september 8-14, 2018, proceedings, part xv* (p. 682–697).