ISSN: 2710-0170

Research on Digital Watermarking Technology Based on Artificial Intelligence Generated Content Model

Yu Liang *, Chaoran Wu, Lin Zhang, Yadong Yu, Hao Hong

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, Anhui, 233030

*1426972597@qq.com

Abstract

With the rapid development of Generative Artificial Intelligence (AIGC) technology, the creation and dissemination of digital content have entered a new era. However, issues such as copyright protection, security, and traceability of these generated contents are gradually becoming prominent, and effective technological measures are still needed to address them. Digital watermarking technology, as an effective means to solve these problems, has a wide range of application prospects. This study proposes a new digital watermarking algorithm specifically targeting the characteristics of AIGC generated content, aiming to address the shortcomings of existing technologies in terms of concealment, robustness, and security. By comparing and analyzing existing technologies, improvement plans were proposed and optimized and experimentally verified. The experimental results show that the proposed algorithm can effectively improve the copyright protection capability of content while ensuring the concealment of watermarks, and has good application prospects.

Keywords

AI generated content; Digital watermark; Copyright protection; Concealment; robustness.

1. Introduction

The application of Generative Artificial Intelligence (AIGC) technology has received widespread attention and practice in recent years. The images, audio, videos, and other content generated by AI not only provide great convenience for creators, but also bring copyright protection and security issues. How to effectively trace and protect generated content, avoid copyright disputes and illegal tampering, has become a problem that needs to be solved.

Digital watermarking technology, as an effective means of copyright protection, can embed identifiable identification information in digital content to ensure the authenticity and copyright ownership of the content. However, existing digital watermarking technologies have not been optimized for the special requirements of generative artificial intelligence generated content, resulting in poor concealment, weak robustness, and poor security. To address these issues, this study proposes a digital watermarking algorithm based on AIGC generated content and optimizes it.

2. Research Background and Current Situation

2.1. Overview of Digital Watermarking Technology

Digital watermarking technology is a technique that identifies the source, copyright, and other information of digital content by embedding detectable and invisible information. According to the manifestation of watermark information, watermarks are divided into explicit watermarks

ISSN: 2710-0170

and implicit watermarks. Explicit watermarks typically affect the appearance of content, while implicit watermarks can embed information without significantly altering the content. Hidden watermarking has a wide range of applications in AIGC, especially in protecting image and video content.

2.2. Challenges of Artificial Intelligence Generated Content

AIGC's generated content presents new challenges in copyright protection due to its automation, complexity, and diversity. Firstly, the copyright ownership of the generated content is unclear, which can easily lead to illegal use and plagiarism. Secondly, the diversity and constantly changing nature of generated content make it difficult for traditional watermarking techniques to adapt, especially in the process of real-time content generation and updating, where the stability and robustness of traditional watermarking techniques are particularly inadequate.

2.3. Current Status and Problems of Digital Watermarking Technology

The existing digital watermarking technology has been applied to copyright protection to a certain extent, but there are still many problems in protecting AIGC generated content. For example, existing algorithms struggle to balance concealment and robustness, and cannot effectively address the loss of watermark information during the processing, modification, and conversion of generated content. In addition, security issues are often exposed, and watermark information may be maliciously attacked or removed.

3. Research Objectives and Methods

3.1. Research Objectives

3.1.1. Propose watermark technology suitable for generative artificial intelligence generated content

Currently, although some watermarking techniques have been applied to protect digital content, these technologies face unique challenges in the field of generative artificial intelligence (AIGC), especially in terms of the invisibility, robustness, and efficient extraction of watermarks during content generation. The content generated by generative artificial intelligence includes text, images, audio, videos, etc. Their generation methods are relatively complex, and watermarking technology must be able to ensure the stability and security of watermarks in different forms and environments. Therefore, this study aims to propose a new watermarking technique that can effectively embed and extract watermark information, addressing the applicability and traceability issues of existing technologies in generative content.

3.1.2. Calculation of Digital Watermark Algorithm for Generative Artificial Intelligence Generated Content

In order to meet the requirements of AIGC, we have designed a novel digital watermarking algorithm. The algorithm mainly has the following characteristics:

High concealment: The watermark information should be embedded in the content without significantly changing the appearance and quality of the content. This means that the algorithm needs to ensure the anonymity of the watermark to prevent users and audiences from detecting its existence.

High availability: Watermarks should be able to be effectively preserved and extracted after common signal processing operations such as compression, filtering, format conversion, etc., to ensure that the watermark can remain effective in practical applications.

ISSN: 2710-0170

High security: Watermark algorithms should have the ability to resist tampering, noise, and deletion to prevent external malicious attacks or operations from damaging the integrity of watermark information.

High robustness: Watermarks should remain stable in different generation environments and content formats, such as different formats and quality requirements for content types such as images, audio, and video.

3.1.3. Algorithm optimization and application verification

Based on the preliminary design of the watermark algorithm, algorithm optimization is carried out to improve its computational efficiency, reduce complexity, enhance robustness, and improve its applicability in different types of generated content. In the optimization process, the focus is on solving the problem of watermark embedding and extraction of generative content in different types (images, audio, video), while verifying the algorithm's application in real scenarios to ensure its stability and efficiency in complex environments.

3.2. Research Methods

This study proposes an improved digital watermarking algorithm based on deep learning [3] and generative adversarial networks (GANs) [4], aimed at applying it to watermark embedding and extraction of AI generated content (AIGC). The core objective of the research is to design a method that can effectively embed watermark information while ensuring the quality of generated content, and to ensure the concealment, robustness, and security of the watermark. Therefore, research methods can be summarized from the following aspects.

3.2.1. Design a new watermark embedding and extraction strategy

Traditional digital watermarking methods typically embed watermarks directly into images or other content, but for AIGC generated content, watermark embedding must be more intelligent and adaptive. The study adopted a multi-level embedding strategy by embedding watermark information at different levels of content, such as frequency domain or depth features, to effectively store it in different data levels. During the embedding process, adaptive adjustments were also made to dynamically adjust the strength of watermark embedding based on the type and quality of generated content, ensuring that the watermark can be effectively stored and extracted even under compression or low quality conditions. At the same time, the study also adopted an information fusion strategy, combining the features of the generated content (such as generation time or network parameters) with watermark information, so that the watermark is not only a copyright identifier, but also serves as metadata for the generated content, enhancing the depth and diversity of the watermark.

3.2.2. Use Generative Adversarial Networks (GANs) to optimize the embedding and extraction process of watermarks

GAN consists of a generator and a discriminator. The generator is responsible for generating content and embedding watermarks, while the discriminator is responsible for determining whether the content contains watermarks. In this study, the generator of GAN not only generates high-quality content, but also needs to embed watermark information to ensure that the watermark does not compromise the quality of the content during the generation process. The task of the discriminator is to evaluate the watermark in the generated content and ensure that the extraction of the watermark is not compromised. Through the mutual confrontation and optimization of the generator and discriminator, the algorithm can achieve the concealment, robustness, and security of the watermark while preserving the authenticity of the generated content.

ISSN: 2710-0170

3.2.3. Experimental analysis of the robustness, security, and concealment of watermarks

Robustness testing verifies whether watermark information can be effectively preserved by performing common operations such as compression, cropping, filtering, and rotation on different types of content such as images, audio, and video. Security testing targets malicious tampering attacks (such as cutting, splicing, etc.) to evaluate the watermark's ability to resist tampering. The concealment test verifies the degree of concealment of the watermark in the content through visual, auditory, and video quality evaluation indicators, ensuring that it does not significantly affect the quality of the generated content.

3.2.4. Conduct algorithm optimization and real-time performance analysis

By utilizing parallel computing, model compression, and other techniques, the computational efficiency of watermark embedding and extraction can be improved. In order to ensure the real-time performance of the watermark algorithm in practical applications, the research combines the edge computing scheme to ensure that the watermark processing can be completed with low delay when generating content, which meets the efficient requirements in practical application scenarios.

4. Research Content

The core of this project is to design and implement a digital watermarking technology for generative artificial intelligence content, and apply it to content types such as images, audio, and video. During the research process, the focus is on the following aspects:

4.1. Importance and application scenarios of digital watermarking technology in the field of generative artificial intelligence

With the development of generative artificial intelligence technology, a wide variety of content types are generated, including images, audio, video, text, etc. These contents are not only products of creation, but may also involve security risks such as copyright issues, data leaks, and malicious tampering. Digital watermarking technology can effectively embed information into generated content, ensuring copyright ownership and preventing unauthorized use.

In the field of AIGC, digital watermarking technology is mainly divided into two categories:

Generate content watermark [5]: Embed a watermark used to identify the generated content to indicate its origin from artificial intelligence.

Sample Protection Watermark [6]: Embed watermarks during the training process to protect data samples, prompts, and other information during the training process, in order to prevent unauthorized use and leakage of data.

Verify watermark: Provide a watermark that can verify whether the generated content complies with copyright and usage regulations, ensuring the legality of the generated content.

In this process, the robustness, algorithm efficiency, and complexity of digital watermarking technology are the main challenges faced. In addition, watermarking needs to balance protection with data quality and user experience. Therefore, research on digital watermarking technology should be closely integrated with practical application scenarios to provide effective copyright protection and regulatory support for future AIGC technologies.

4.2. Application of Digital Watermark Identification for Content Generated by Artificial Intelligence

One of the core applications of digital watermarking technology is content identification and copyright protection. Digital watermarking can not only clearly identify the source of generated content, but also help track data leakers, protect intellectual property, and more. In the generative artificial intelligence industry, digital watermarking technology is widely used,

ISSN: 2710-0170

especially explicit watermarking, which is used to directly identify the artificial intelligence source of the generated content and is suitable for scenarios where users need to be clearly informed of the generated content; There is also an invisible watermark used to covertly embed copyright information, generate model information, etc., to ensure that the copyright information of the generated content does not interfere with the user experience.

The application process of watermark identification includes the following steps:

a. Content generation: Generative artificial intelligence uses algorithms to generate digital content.

b. Watermark embedding: Relevant information during the generation process (such as the generator, generation model, generation time, etc.) is embedded as watermark information into the content.

c. Watermark extraction and verification: Extracting watermark information through specific algorithms to ensure that the copyright and creator information of the generated content can be effectively verified.

This watermark identification not only helps with copyright protection, but also enhances users' awareness of the source of generated content and reduces confusion.

4.3. Research on Watermark Algorithm for Content Generated by Generative Artificial Intelligence

The content generated by generative artificial intelligence can be effectively identified and protected through digital watermarking technology. Watermarking technology has different application requirements and challenges in content types such as audio, video, and images. For example:

Image watermark: requires high concealment and robustness to avoid the watermark being removed by image processing operations.

Audio watermark: usually embedded in the time or frequency domain of audio to ensure that the watermark remains effective during audio compression, resampling, and other processes.

Video watermarking: Video watermarking needs to consider factors such as high compression ratio and temporal redundancy of the video, so watermarks can be embedded in multiple domains of the video (such as pixel domain, transform domain, time domain, etc.).

5. Experiment and Result Analysis

5.1. Experimental setup

In order to comprehensively test the performance of the proposed watermark algorithm, the experimental setup is as follows:

(1) Experimental dataset

AI generated images: High quality images generated using Generative Adversarial Networks (GANs), covering different types of content such as landscapes, people, animals, etc. These images simulate image generation tasks in real-world applications, with different image details and complexities.

AI generated audio: Several AI generated audio contents have been selected, including vocals, music, and natural sounds. These audio files are also generated through GAN models, representing the challenges of audio content generation.

AI generated video: AI generated video content has been selected, covering different scenes and motion modes. The video content is dynamic images with high complexity and diversity.

(2) Experimental scenario

ISSN: 2710-0170

Perform various transformation operations on generated content, including image cropping, rotation, compression, filtering, audio cropping, volume adjustment, filtering, video editing, cropping, frame rate changes, etc.

Each test scenario is compared with a control group using traditional watermarking algorithms to ensure the reliability of the experimental results.

(3) Evaluation indicators

Concealment: measured by evaluating the impact of watermarks on the quality of generated content. Common indicators include peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). A higher PSNR and SSIM value indicates that watermarking has a relatively small impact on content quality.

Robustness: Test whether the watermark can effectively recover content after various common image and video operations such as cropping, rotation, compression, etc. Evaluate robustness through the success rate of watermark extraction.

Security: Test against malicious attacks (such as tampering, deletion, compression, etc.) to evaluate whether the watermark can effectively resist attacks and maintain information integrity.

5.2. Result Analysis

The experimental results further validated the advantages of the proposed algorithm in AI generated content:

In terms of concealment, the algorithm successfully balances the relationship between watermark embedding and content quality through adaptive adjustment and multi-layer embedding strategy. In most experiments, watermarking hardly affects the perceived quality of content, which is crucial for copyright protection in practical applications.

In terms of robustness, the algorithm can still effectively extract watermarks even after encountering common transformations in images and videos, especially in situations of high compression and intensive editing. This indicates that the algorithm is capable of handling various operations in the real world and has good adaptability to generating content.

In terms of security, through various attack methods testing, watermarks can effectively prevent malicious tampering and compression attacks, demonstrating strong resistance to attacks. When facing different attacks, watermark information can maintain its integrity and effectively prevent content from being maliciously modified or plagiarized.

6. Conclusion and Prospect

The digital watermarking technology based on AIGC generated content proposed in this study solves the problems of watermark concealment, robustness, and security in generated content by improving existing algorithms. The experimental results show that the proposed watermarking algorithm can effectively protect the copyright of generated content, ensuring the authenticity and traceability of the content.

Future research can further optimize the real-time performance of algorithms and explore the application of watermarking technology in different fields, such as AI generated audio, virtual reality, etc. Meanwhile, with the continuous development of generation technology, digital watermarking technology needs to constantly adapt to new challenges, such as batch protection for large-scale generated content.

Acknowledgements

This research is funded by Anhui University of Finance and Economics College Student Innovation and Entrepreneurship Training Program Project (NO.: S202410378397).

ISSN: 2710-0170

References

[1] Guo Zhaojun, Li Meiling, Zhou Yangming, etc. Research progress on digital watermarking technology for content models generated by artificial intelligence [J]. Journal of Cyberspace Security Science, 2024,2 (01): 13-39.

[2] Liang Yan. Application of Digital Watermarking Technology in Digital Media Copyright Protection and Optimization [J]. Electronic Technology, 2024, 53 (05): 320-321.

[3] Wang Huibing. Research on Image Watermarking Algorithm Based on Deep Learning [D]. Jiangxi University of Science and Technology, 2024.

[4] Wei Ying. Research on Image Digital Watermarking Based on Generative Adversarial Networks [D]. Chongqing University of Technology, 2022.

[5] Liu Anan, Su Yuting, Wang Lanjun, etc. Progress in AIGC Visual Content Generation and Traceability Research [J]. Chinese Journal of Image and Graphics, 2024, 29 (06): 1535-1554.

[6] Wang Jie. Research on Copyright Algorithm Based on Adversarial Samples [D]. Dongguan University of Technology, 2022.