Counterfeiting in Depth Synthesis based on Digital Watermarking

Yu Liang 1,*, Yadong Yu 1, Yina Wang 1, Dunjun Li 1, Zejiong Zhou 2, b, *

1 School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233030, China
2 School of Economics, Anhui University of Finance and Economics, Bengbu, 233030, China
1 1426972597@qq.com, b, * aczzj123456@163.com
* Corresponding author: Zejiong Zhou (Email: aczzj123456@163.com)

Abstract: The purpose of this paper is to discuss and apply digital watermarking technology to solve the forgery problem in depth synthesis. With the rapid development of deep synthesis technology and its application in various fields, it is particularly important to protect the authenticity and integrity of digital content. Based on the understanding of digital watermarking, this paper explores an experimental design, which uses watermarking embedding and extraction algorithms and forgery detection technology to solve the problem of deep forgery, protect the copyright, integrity and anti-copy of digital products. In order to improve the robustness and reliability of the watermark, a suitable watermark embedding and extraction algorithm is designed by analyzing the characteristics of deep synthesis forged media in the experimental process. Then select the data set containing the original digital media and the deep synthetic forged samples, extract the features of the two, and find out the features that distinguish the differences between the two. Finally, the forgery detection technology is used to evaluate the performance of digital watermarking technology in depth forgery detection. In this paper, digital watermarking technology is used to provide an effective solution to the problem of forgery in deep synthesis, which can be applied to protect intellectual property rights, prevent tampering and forgery, and protect the authenticity and integrity of digital media content.

Keywords: Deep Composition; Deep Forgery; Digital Watermarking; Authenticity; Integrity.

1. Introduction

Deep forgery [1] refers to the use of deep learning and artificial intelligence technology to generate realistic false content, especially the synthesis and tampering of human faces as the main object. The emergence of deep forgery technology is related to the development of deep learning and computer vision, the enhancement of big data and computing power, and the popularity of social media and digital content. Deep forgery technology brings potential application value in film and television special effects, artistic creation and other fields, but also brings a series of problems and challenges. Deep forgery technology may be used for malicious purposes, such as fraud, network fraud, false news dissemination, etc., which seriously affects social stability and personal privacy security.

Digital watermarking technology [2] is a technology that hides specific information in digital media, so that the information can be transmitted and stored safely, and can be identified and extracted when needed. Digital watermarking technology is widely used in copyright protection, content authentication, information hiding and so on. Its main application areas include: copyright protection, retrieval and traceability of audio, image, video, text and other digital content; digital information management and protection in digital libraries, e-commerce, network security and other fields; document identification, authentication and transaction audit in government, enterprises and other institutions.

Digital watermarking technology can be used to solve the problem of deep forgery. First, feature extraction is carried out on the original image, and the extracted feature information is embedded into the original image to form a digital watermark. Then, a deep learning model is used for training to classify the real face and the forged face, and the features that distinguish the two are extracted. When there is a suspected depth forgery image, the extracted features are compared with the watermark embedded in the original image for verification.

Figure 1. The process of digital watermarking technology to solve the problem of forgery

However, with the continuous development of technology, forgery algorithms are also improving, and there may be methods to combat watermarking. Therefore, it is necessary to study new algorithms, models and techniques to improve the performance and effect of watermark embedding and detection. For example, the technique based on blind watermarking [3] can detect the watermark without the original data, while the technique based on non-blind watermarking requires the original data to be encrypted before detection. In addition, there are some emerging areas of research, such as antagonistic watermarking, multimedia big data watermarking and so on.

In this paper, the watermark embedding and extraction algorithm is used to embed the watermark information into
the digital media efficiently and accurately, while ensuring the reliability and integrity of the watermark information when extracting. The research in this area needs to consider the robustness of the watermark, that is, the ability to resist attacks, and to minimize the impact on the quality of the original media. The watermark algorithm is designed to make the watermark information not easy to be detected or modified in the embedding process, thus increasing the difficulty of forgery. This paper designs the experimental process to solve the deep forgery, and provides some ideas for the research. It mainly uses forgery detection technology [4] to detect the digital media embedded with watermark to verify whether it contains effective watermark information, so as to solve the problem of whether the image is forged. The research in this area needs to consider the sensitivity and accuracy of the detection algorithm, as well as the adaptability to various attacks such as compression, filtering, and noise addition.

The research objectives of this paper include: Embedding the watermark information into the image by using the watermark embedding and extraction algorithm, and comparing the original image with the watermark detection technology to determine whether the image is true or false. Improve the robustness and security of digital watermarking technology, making digital watermarking more difficult to be forged and tampered with. Provide efficient and accurate digital watermark embedding and extraction algorithms to reduce the impact of the embedding process on the quality of the original media.

2. Introduction of Digital Watermarking Technology

2.1. Principles and Classification of Digital Watermarking

Digital watermarking is a technique for embedding invisible identification information in digital media, which is used for copyright protection, content authentication and forgery detection. Its principle is based on the appropriate modification or embedding of information in the original media data, so that the modified media data can still maintain the invariance of audio-visual effects.

According to the embedding and extraction methods of digital watermarking, digital watermarking can be divided into the following categories: (1) Spatial domain watermarking [5]. Spatial domain watermarking refers to the embedding of information directly on the pixel values of the original media. It hides the watermark information by adjusting the brightness or color value of the pixel. The common techniques are least significant bit replacement method and spread spectrum technology. (2) Transform domain watermarking [6]. Transform domain watermarking refers to the transformation of the original media data into a mathematical transform domain, in which information is embedded. Common transforms include Fourier transform, wavelet transform and discrete cosine transform. Watermarking in transform domain has better robustness and imperceptibility, and the commonly used methods are frequency domain coding and parity check code. (3) Feature-based watermarking [7]. Feature-based watermarking refers to embedding and extracting watermark information by using some specific features in the original media. These features may be image textures, edges, local statistical features, etc. By embedding and extracting operations on these features, the watermark can be hidden and extracted. (4) Time-domain watermarking [8]. Temporal watermarking is a digital watermarking technology applied to temporal media such as video and audio. It embeds and extracts the watermark according to the time information of the media data. The common technologies are time sequence coding and watermarking algorithm based on human auditory model. (5) Multimedia synchronous watermarking [9]. Multimedia synchronization watermarking is a watermarking technique for maintaining synchronization in multimedia content. For example, watermark information is embedded in video and audio at the same time to ensure timing synchronization between the two.

2.2. Digital Watermarking Implementation Procedure

The implementation process of digital watermarking is as follows. (1) Select the type of watermark. Select the appropriate type of digital watermark according to the requirements and application scenarios. Common types of digital watermarking include visible watermarking, invisible watermarking, blind watermarking and so on. (2) Feature selection and extraction. Determine the watermark information to be embedded into the original image, such as text, image, binary code, etc. For invisible watermarking, we usually use image processing or feature extraction algorithms to extract the features of the original image. (3) Embedding watermark. The selected watermark information is embedded into the original image. There are many specific embedding methods, such as frequency domain embedding, spatial domain embedding, spread spectrum embedding and so on. The embedding process needs to pay attention to maintain the quality and visibility of the image, as well as the appropriate encryption protection of the watermark. (4) Design of detection algorithm. An appropriate detection algorithm is designed to extract and verify the existence of the watermark. The detection algorithm may be based on techniques such as alignment, feature extraction, machine learning, and the like. The design goal of the detection algorithm is to judge the existence of the watermark as accurately as possible, and to deal with attacks such as depth forgery robustly. And (5) extract and verifying that watermark. Watermark is extracted and verified from the watermarked image. In the extraction process, the watermark information is extracted by reverse processing of the image according to the embedding algorithm and the key. The verification process can be judged by comparing the extracted watermark information with the expected watermark information, or analyzed by feature extraction and other methods. (6) Optimization and adjustment. According to the actual application requirements and effects, the digital watermarking scheme is optimized and adjusted. This may involve the adjustment of parameters, the improvement of algorithms, the enhancement of encryption protection, etc.

2.3. Embedding and Extracting Algorithm of Digital Watermark

The specific implementation process of the digital watermarking embedding algorithm is as follows. (1) The original media data is transformed into a mathematical transform domain, such as Fourier transform, wavelet transform and so on. (2) select that watermark information, namely selecting a part from the target watermark information, such as a paragraph of text, a numb or a picture. (3) processing
the watermark information, namely encrypting the watermark information, scrambling the watermark information and the like to enhance the security of the digital watermark. (4) Embedding the encrypted watermark information into the media data: Embedding the encrypted watermark information into the media data by changing some specific positions of the media data, such as modifying the least significant bit in the pixel value. (5) Performing inverse transformation on the embedded media data, namely re-converting the embedded media data into original media data to complete the embedding process of the digital watermark.

![Figure 2. Digital watermark embedding process](image)

The specific implementation process of the digital watermarking extraction algorithm is as follows: (1) Transform the media data: convert the media data into the same mathematical transform domain as the embedding. (2) Positioning the position of the watermark information according to the rule used during embedding. (3) Extracting the watermark information, namely extracting the encrypted watermark information from a specific position of the media data. (4) Carrying out operations such as decryption and disorder on the extracted watermark information to obtain the original watermark information. (5) Judging whether the media data is tampered or forged according to the decrypted watermark information, and completing corresponding functions such as authentication, protection, detection and the like.

![Figure 3. Digital watermark extraction process](image)

### 2.4. Advantages and Disadvantages of Digital Watermarking

**Advantages of digital watermarking technology.** (1) Identity authentication and traceability. Digital watermarking technology can provide identity authentication and traceability for deeply synthesized digital media, so that the true source of the media can be confirmed, which is helpful to maintain intellectual property rights and copyright protection. (2) Concealment. The embedding of digital watermark can be carried out without affecting the perception of human eyes, maintaining the visual quality of the media, while providing a covert way to protect the media content. (3) Enhance the authentication ability. Digital watermarking technology can enhance the authentication ability of the media to a certain extent, so that people can verify the authenticity of the media through specific technical means, which is helpful to prevent forgery and tampering.

**Disadvantages of digital watermarking technology.** (1) Capacity limitation. Digital watermarking technology has certain limitations on the capacity of embedded information, which may not be able to meet the needs of large-scale data, especially for high-resolution images and videos. (2) Robustness challenge. Some digital watermarking technologies may have poor robustness to compression, cropping and other operations, resulting in the vulnerability of embedded watermarks and reducing their reliability. (3) Privacy issues. In some scenarios, digital watermarking technology may involve the embedding and dissemination of personal privacy information, and special attention should be paid to privacy protection.

### 3. Analysis of Forgery in Deep Synthesis

#### 3.1. Types of Depth Synthesis Image Forgery

Depth synthesis image forgery refers to the use of computer graphics and artificial intelligence technology to produce false visual information by synthesizing, modifying or generating images. This kind of forgery can use a variety of technical means, and the common types of deep synthesis image forgery include the following. (1) Face-changing technology. Using deep learning technology, one person's facial features are synthesized into another person's image to create a camouflage effect. This technique is widely used in videos and pictures, and is also used to create fake videos and pictures. (2) Human pose and motion synthesis. Using computer graphics and deep learning technology, one person's pose and motion can be synthesized into another scene, thus creating a false picture that looks real. (3) Scene synthesis and background replacement. Different scenes, environments or backgrounds can be synthesized into the original image through the synthesis technology, so that the image presents an illusion inconsistent with the actual situation. (4) Forgery of words and slogans. Image processing technology can be used to forge words, slogans and other information in pictures, including adding, deleting or modifying the text information in pictures to create misleading picture content. (5) Image restoration and tampering. Image processing technology is used to restore and tamper with images, including repairing missing parts, deleting or adding specific objects, etc., in order to deceive the audience. (6) Generation of false scenes. Computer graphics technology is used to generate seemingly real false scenes, including false landscapes, buildings, cities, etc. This technology is widely used in the field of film and television special effects and virtual reality.

These types of depth synthesis image forgery are becoming more and more common in today's society, which may lead to misleading information, privacy disclosure and social instability. Therefore, it is necessary to develop and improve digital watermarking, image recognition and other technical means to cope with the challenges of depth synthesis image forgery.

#### 3.2. Deep Forgery Detection Methods and Limitations

The detection methods of depth synthesis image forgery mainly include the detection method based on traditional image processing technology, the detection method based on physiological signal characteristics, the detection method based on GAN image characteristics and the detection method based on deep learning. (1) Detection methods based on traditional image processing techniques [10]. These methods mainly use some statistical features or geometric features of images, such as illumination inconsistency, noise,...
discontinuity, etc., to detect forged images. However, due to the development of deep learning models, traditional methods are facing great challenges in the detection of high-quality forged images. (2) Detection method based on physiological signal characteristics. refers to the technical method that uses various signals generated by human physiological activities, such as electrocardiogram, electroencephalogram, skin electrical activity, etc., to classify, identify, predict or monitor by extracting the characteristics of these signals. These methods have important applications in the fields of medicine, biometrics, health monitoring and disease diagnosis. (3) Detection method based on GAN image features. refers to a technology that uses the image features generated by the GAN model for image processing, analysis and recognition. GAN is an adversarial training framework consisting of a generator and a discriminator. The generator is responsible for generating fake image samples, while the discriminator is responsible for distinguishing the real image from the fake image generated by the generator. It has a wide range of applications in image generation, image editing, image recognition and other fields. (4) Deep learning-based detection method [11]. This method uses deep learning models to learn and capture features in forged images. For example, models such as a convolutional neural network (CNN) or a generative adversarial network (GAN) can be used for image forgery detection. These models can discriminate by learning the feature differences between a large number of real and fake images. However, there are also some limitations of such methods.

Although the deep learning model can effectively detect some common forged images to some extent, there are still some limitations. (1) Anti-sample attack [12]. For deep learning-based detection methods, the attacker can modify the input image so that it is misclassified as the real image. This kind of attack is called anti-sample attack, which can bypass the detection model and make the forged image not be detected correctly. (2) Dataset limitation. Training deep learning-based detection models usually requires a large number of real and fake image datasets. However, it is not easy to obtain an accurate and comprehensive set of forged images, which limits the training and generalization ability of the detection model. (3) Mixed forged images. In reality, forged images are often the mixed results of a variety of technologies, including artificial synthesis, image processing and so on. For such complex hybrid forged images, a single detection method may not be able to accurately identify them.

4. Experiments and Results

4.1. Introduction to Experimental Design and Dataset

4.1.1. Experimental Procedure

The experimental design steps based on this study are as follows. (1) Determine the purpose of the experiment. In the research of digital watermarking technology to solve the problem of deep forgery, the purpose of the experiment may be to verify the effectiveness of digital watermarking technology in the detection of deep forgery. (2) Design method. A digital watermark embedding and extraction method based on deep learning can be used, and a deep forgery detection model based on convolutional neural network can be used for detection. (3) Determine the experimental settings and parameters. It is necessary to determine the experimental settings such as the deep learning framework, the type and number of graphics cards, the data set division, and the experimental parameters such as the parameters of digital watermark embedding and extraction, the structure and hyperparameters of the deep forgery detection model. (4) Experimental operation and data collection. It is necessary to embed and extract digital watermark on the training set, train and verify the deep forgery detection model, and then test and evaluate it on the test set. (5) Analyze and summarize the experimental results. Accuracy, recall and F1-score can be used to evaluate the performance of digital watermarking technology in depth forgery detection, and the experimental results are analyzed and summarized.

4.1.2. Data Set

In combination with the above experiments, the data sets related to image processing used are shown in the following table.

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Data type</th>
<th>Number of samples</th>
<th>Sample size</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCID [13]</td>
<td>Image</td>
<td>1338</td>
<td>384 x 256</td>
<td>Digital image processing</td>
</tr>
<tr>
<td>BOSSbase [14]</td>
<td>Image</td>
<td>10000+</td>
<td>Uncertain</td>
<td>Research on image steganography</td>
</tr>
<tr>
<td>CoMoFoD [15]</td>
<td>Image</td>
<td>13</td>
<td>Uncertain</td>
<td>Digital image processing</td>
</tr>
<tr>
<td>Lenna [17]</td>
<td>Image</td>
<td>1</td>
<td>512 x 512</td>
<td>Digital image processing</td>
</tr>
<tr>
<td>Stirmark [18]</td>
<td>Image</td>
<td>Uncertain</td>
<td>Uncertain</td>
<td>Digital copyright protection evaluation</td>
</tr>
</tbody>
</table>

The data sets in the above table are described as follows. (1) The UCID (University of California, Irvine Dataset) data set is a commonly used data set for digital watermarking research. The dataset contains 1338 uncompressed color images, including natural scenes, man-made images and computer-generated images. (2) BOSSbase (BOWS-2) dataset is a commonly used dataset for digital image forgery detection research. The dataset contains 10,000 unmodified grayscale images and their corresponding steganographic images, including 5,000 natural scene images and 5,000 computer-generated synthetic images. It is dedicated to the evaluation and comparison of digital image forgery detection algorithms. (3) The CoMoFoD (Color Model and Filtering Distortion Database) dataset is a commonly used dataset for evaluating image denoising algorithms. The dataset contains 500 color images divided into 250 reference images and 250 distorted images. The main purpose of this dataset is to study and evaluate the performance of image denoising algorithms in terms of color model and filtering distortion. (4) ImageNet data set is a widely used large-scale image data set. The dataset contains more than 14 million annotated images, covering more than 20,000 categories. This dataset is mainly used to train and evaluate deep learning models in computer vision tasks such as image classification, object detection and image segmentation. (5) Lenna dataset is a common test image dataset, which is often used in image processing, image
compression, digital signal processing and other fields. This dataset can be used to evaluate the performance of image processing algorithms in different scenarios. (6) The Stirmark dataset is a commercial dataset for evaluating digital media security technologies. The dataset contains a variety of digital media files, such as images, audio, video, and a variety of different types of digital watermarking and encryption techniques.

4.2. Digital Watermark Embedding and Extraction Experiment

The purpose of this experiment is to explore the effectiveness and stability of digital watermarking embedding and extraction technology, in order to deal with the problem of digital media tampering such as depth synthesis image forgery.

(1) Data preparation. First, the image data of the training set and the test set need to be prepared. These images should contain content that requires digital watermark embedding and extraction.

(2) Digital watermarking embedding. a. Select the digital watermarking algorithm: select the appropriate digital watermarking algorithm according to the research needs and experimental purposes. Common digital watermarking algorithms include Discrete Cosine Transform (DCT) [19], Discrete Wavelet Transform (DWT) [20], etc. B. Parameter setting: determine the parameters in the digital watermark embedding process, such as watermark strength, embedding rate, etc. These parameters affect the visibility and robustness of the watermark. And C. image processing, namely embedding the selected image by using a digital watermarking algorithm. Specific operations may include image transformation, frequency domain embedding or coding, and so on. The embedded image will contain digital watermark information.

(3) extract that digital watermark. a. realizing an extraction algorithm, namely realize a corresponding digital watermark extraction algorithm according to the selected digital watermark algorithm. The algorithm should be able to accurately extract the watermark information from the image embedded with digital watermark. And B. image processing, namely operating the image embedded with the digital watermark by using an extraction algorithm to extract digital watermark information. And C. comparing and verifying, namely comparing and verifying the extracted digital watermark with the original watermark, and evaluating the accuracy and the robustness of extraction.

(4) Evaluation and analysis. a. Performance evaluation: Some indexes are used to evaluate the performance of digital watermark embedding and extraction, such as embedding rate, extraction rate, distortion, anti-attack, etc. These indicators can measure the effect of digital watermarking algorithm. B. Result analysis: analyze and summarize the experimental results according to the results of the evaluation indicators. We can discuss the applicability, advantages and disadvantages of digital watermarking algorithm in different scenarios, as well as the direction of improvement.

4.3. Forgery Detection Experiment and Result Analysis

4.3.1. Design of Experiment

The purpose of this experiment is to explore the accuracy and robustness of forgery detection technology, and evaluate its feasibility and effect in practical application through the analysis of the results. (1) Problem definition. specify the forgery detection problem to be solved, such as image tampering, audio forgery, or video obfuscation. (2) Data set preparation. select the data set containing the original sample and the forged sample. Ensure that the data set contains various types of fake samples with sufficient similarity to the original sample. (3) Feature extraction. feature are extracted from that original sample and the fake samples. The features may include textures of the image, color histograms, spectral features of the audio, and the like. (4) Classifier training. training a classifier model by using the extracted features and the labeled original/fake samples. The commonly used classifiers include support vector machine (SVM), random forest and so on.

4.3.2. Experimental Evaluation

Evaluation metrics are used to evaluate the performance of the classifier, such as precision, recall, F1 value, etc. (1) Accuracy rate. The accuracy rate is the proportion of the number of samples correctly predicted by the classifier to the total number of samples, that is, (true positive examples + true negative examples)/total number of samples. (2) Recall rate. Recall rate measures the prediction ability of the classifier for positive examples. The proportion of the number of samples predicted as positive samples to the number of all actual positive samples, and the calculation formula is the harmonic mean of the precision rate and the recall rate. It is the harmonic mean of the precision rate and the recall rate. The calculation formula is 2 * (precision rate * recall rate)/ (precision rate + recall rate).

4.3.3. Experimental Procedure

The specific process of the forgery detection experiment is as follows: (1) Data set division. The data set is divided into a training set and a test set. Cross-validation is usually used to ensure the reliability of the experimental results. (2) feature extraction. feature is extracted from that original sample and the fake samples. These features should be able to capture the difference between the original sample and the fake sample.

For example, the researchers found two graphs in the sample and extracted their characteristics as shown in the following figure.

Figure 4. Location information
5. Research Summary

5.1. Summary of the Study

Through research, researchers can use digital watermarking technology to solve the problem of depth forgery. Digital watermarking technology can embed the source data into digital media, so that the source data can be extracted in the subsequent detection. If the source data has been tampered with, it can be judged by comparing the extracted source data.

In this paper, the experimental process is designed to solve the deep forgery problem. According to the research requirements, select the appropriate digital watermarking algorithm. Based on the type of forgery studied, the data set containing the original sample and the forged sample was selected. Then feature extraction is carried out to distinguish the difference between the original sample and the forged sample, and the forged media is verified. Finally, a forgery detection experiment is carried out to evaluate the performance of digital watermarking technology in deep forgery detection.

In a word, the research based on digital watermarking technology to solve the problem of forgery has achieved some results. With the continuous development and application of technology, digital watermarking technology will play a greater role in anti-counterfeiting, copyright protection and so on. However, further research is still needed to deal with new forgery methods and improve the reversibility and anti-attack ability of digital watermarking technology.

5.2. Suggested Directions for Future Research

It is a research project with wide application prospect to solve the forgery problem based on digital watermarking technology. On the basis of the existing research results, we can also deepen and improve the watermarking algorithm, and think and expand the follow-up research direction.

For example, for different types of data, such as images, audio, video and so on, more efficient, secure and stable digital watermarking algorithms are studied. We can consider combining the research results of deep learning, cryptography and other fields to improve the robustness and invisibility of digital watermarking. In order to enhance the accuracy and reliability of forgery detection, multiple watermarking techniques can be studied to embed multiple types of watermark information into the data. By combining different types of watermarks, the threshold of forgery detection can be improved, and the detection results can be more accurate.

Through the continuous expansion of technology, digital watermarking technology can be combined with other fields such as cryptography, big data, Internet of Things and so on to open up new application scenarios. For example, supply chain management, medical data protection and other applications based on digital watermarking are studied.

In a word, it has important research value and application potential to solve the forgery problem based on digital watermarking technology. The follow-up research can be explored from multiple levels, aiming to lay a solid foundation for the development and application of digital watermarking technology in China.
Acknowledgments

This work is supported by Anhui University of Finance and Economics Provincial College Student Innovation and Entrepreneurship Training Program Project (S2022 1037 8124).

References


