applied sciences

MDPI

Article

High-Frequency Artifacts-Resistant Image Watermarking Applicable to Image Processing Models

Li Zhang, Xinpeng Zhang and Hanzhou Wu *

School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China; jmzhangli@shu.edu.cn (L.Z.); xzhang@shu.edu.cn (X.Z.)
* Correspondence: hanzhou@shu.edu.cn

Abstract: With the extensive adoption of generative models across various domains, the protection of copyright for these models has become increasingly vital. Some researchers suggest embedding watermarks in the images generated by these models as a means of preserving IP rights. In this paper, we find that existing generative model watermarking introduces high-frequency artifacts in the high-frequency spectrum of the marked images, thereby compromising the imperceptibility and security of the generative model watermarking system. Given this revelation, we propose an innovative image watermarking technology that takes into account frequency-domain imperceptibility. Our approach abandons the conventional convolutional neural network (CNN) structure typically used as the watermarking embedding network in popular watermarking techniques. This helps the image watermarking system to avoid the inherent high-frequency artifacts commonly produced by CNNs. In addition, we design a frequency perturbation generation network to generate low-frequency perturbations. These perturbations are subsequently added as watermarks to the low-frequency components of the carrier image, thus minimizing the impact of the watermark embedding process on the high-frequency properties of the image. The results show that our proposed watermarking framework can effectively embed low-frequency perturbation watermarks into images and effectively suppress high-frequency artifacts in images, thus significantly improving the frequency-domain imperceptibility and security of the image watermarking system. The introduced approach enhances the average invisibility performance in the frequency domain by up to 24.9% when contrasted with prior methods. Moreover, the method attains superior image quality (>50 dB) in the spatial domain, accompanied by a 100% success rate in watermark extraction in the absence of attacks. This underscores its capability to uphold the efficacy of the protected network and preserve the integrity of the watermarking process. It always maintains excellent imperceptibility and robustness. Thus, the framework shows great potential as a state-of-the-art solution for protecting intellectual property.

Keywords: model watermarking; security; high-frequency artifacts; imperceptibility; deep neural networks; image watermarking

check for updates

Citation: Zhang, L.; Zhang, X.; Wu, H. High-Frequency Artifacts-Resistant Image Watermarking Applicable to Image Processing Models. *Appl. Sci.* **2024**, *14*, 1494. https://doi.org/10.3390/app14041494

Academic Editor: Luis Javier Garcia Villalba

Received: 19 January 2024
Revised: 8 February 2024
Accepted: 9 February 2024
Published: 12 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, the field of generative modeling has experienced a substantial surge in development, resulting in the proliferation of numerous generative models. Artificial Intelligence-Generated Content (AIGC) [1–3] has garnered significant attention from both the scientific community and society at large. This trend underscores the substantial value of contemporary generative models. Consequently, the protection of intellectual property (IP) rights for these models has emerged as a prominent and pressing concern in the current landscape. To protect the IP of models, both academia and industry have developed a range of model watermarking techniques tailored for the protection of deep models. These models conventionally comprise three fundamental elements: the input, internal structure, and output. Many of these methods [4–10] mark the model by altering its internal parameters or structure, thus enabling IP protection of the model. On the other

hand, other methods [11–16] for verifying model ownership involve creating trigger sets and fine-tuning the model to produce aberrant outputs.

Recently, model watermarking techniques have been extended to protect generative models [17–20], which we denote as generative model watermarking. In this paper, we study generative model watermarking [17–20] for protecting an image generation network. Specifically, it involves embedding specific watermark information into the images generated by the network. In other words, a watermarked deep neural network is identifiable by the presence of watermarks in all the images it produces. Many of these generative model watermarking frameworks draw upon advanced techniques in image deep steganography [21,22]. In this paper, we will focus on the two most popular generative model watermarking techniques [17,19]. Wu et al. [17] introduced a framework that invisibly embeds watermarks into the output of the model by incorporating a watermark loss into the loss functions. The core of this watermarking framework includes a key-controlled watermark extraction network that is co-trained with the protected network. This watermark extraction network is able to extract the watermark from watermarked images. Additionally, any non-marked image or employing an incorrect key for copyright verification prompts the watermark extraction network to output noise, resulting in the failure of copyright authentication. Furthermore, the framework improves its resistance to common pre-processing attacks through adversarial training on manipulated samples. Note that this method utilizes the protected network for watermark embedding, which may potentially impact the quality of images generated by the network. Zhang et al. [18,19] proposed a novel watermarking framework designed for image processing networks, specifically aimed at countering surrogate model attacks. Such attacks involve collecting many inputs and outputs from the target model to guide the training of a surrogate model. The objective is to enable the surrogate model to achieve similar capabilities and performance levels as the target model. In this method, an imperceptible watermark is embedded into the model's output using a watermark embedding network. The watermarked output can then be extracted by the watermark extraction network to retrieve copyright information. In a surrogate model attack scenario, the attacker only has access to their own inputs and the outputs with the embedded watermark. Consequently, the output of their trained surrogate network also contains copyright information, preventing the surrogate model attack. However, it is important to note that this method may not be entirely robust against certain pre-processing attacks since the required consistency could be compromised.

The creation of a generative model of watermarking relies on the watermark embedding network to add a watermark to the output of the protected network for ownership verification. It is crucial to emphasize the watermark embedding network and the watermark addition process, which we will analyze in the following. It is evident that current generative model watermarking techniques [17–19] employ deep networks as the watermark embedding network. For instance, Wu et al. [17] utilize the host network as the watermark embedding network, while Zhang et al. [18,19] introduce an additional watermark embedding network. The watermark embedding network often involves numerous up-sampling and down-sampling operations, which may inadvertently introduce high-frequency artifacts into the output image [23]. Furthermore, the process of adding watermarks to an image is equivalent to introducing high-frequency perturbations, which directly leads to the emergence of high-frequency artifacts within the marked images [24]. Therefore, we find that all existing generative model watermarking methods encompass the watermark embedding network and watermark addition process as described earlier, inevitably introducing high-frequency artifacts into the images marked by generative model watermarking systems. These high-frequency artifacts undermine the ability of the model watermarking system to maintain invisibility in the frequency domain.

To ensure the effectiveness of watermarking systems, the embedded watermarks must be imperceptible to potential attackers [20,25,26], thus reducing the risk of a successful attack. Conversely, when a watermarking system lacks invisibility, it becomes vulnerable to direct exposure to attackers, making it susceptible to damage and invalidation. Our analysis

reveals the presence of high-frequency artifacts within the frequency domain of the images marked by generative model watermarking. Past generative model watermarking assert that their watermarks are invisibly embedded in the image's spatial domain, making it challenging for an attacker to detect the model watermarking in this spatial view. However, the existence of high-frequency artifacts in the frequency domain of the marked images renders the generative model watermarking conspicuous within the frequency domain, thus making it vulnerable to detection by potential attackers. The lack of invisibility within the frequency domain poses a significant challenge, as it could potentially enable attackers to identify and remove the embedded watermark using specific attacks. This vulnerability may lead to unauthorized watermark removal or fraudulent ownership claims. Hence, we advocate for the consideration of watermark invisibility in both spatial and frequency domains in the context of generative model watermarking. Additionally, we propose an image watermarking to address the existing issues of invisibility associated with generative model watermarks within the frequency domain.

In order to enhance the imperceptibility of generative model watermarking, we introduce a novel image watermarking framework designed to mitigate the high-frequency artifacts. Our approach bypasses traditional watermark embedding networks typically used for image marking, thus effectively addressing the issue of high-frequency artifacts associated with deep networks. We propose a frequency perturbation generation networks responsible for generating frequency perturbations, which can flexibly fine-tune the perturbation range adjustments. These perturbations are subsequently integrated into the frequency domain of the original image to create the marked image. By limiting the size of the watermark perturbation and adding it to the low frequencies of the carrier image (the central region of the frequency domain), we successfully prevent it from affecting the high-frequency regions of the image. Subsequently, we train the watermark extraction network to accurately recover the watermark from the marked image, ensuring high-quality retrieval. Our results conclusively demonstrate the effectiveness of our proposed watermarking framework in embedding the low-frequency perturbation watermark into the image while effectively suppressing high-frequency artifacts. This significantly enhances the frequency-domain imperceptibility and security of the watermarking system. Experimental findings affirm that our method does not compromise the performance of the protected network. It consistently delivers exceptional imperceptibility and robustness. Consequently, our framework exhibits substantial potential as an advanced solution for safeguarding IP rights.

The contributions of this paper are summarized as follows:

- Our investigation has revealed the presence of high-frequency artifacts in marked images, emphasizing the limitations of existing watermarking techniques in achieving concealment in the frequency domain. This novel discovery underscores the imperative for further research and refinements in the frequency domain to enhance the concealment capabilities of watermarking.
- We propose an image watermarking system with the aim of improving watermark imperceptibility in the frequency domain. Unlike existing methods that use a standard watermark embedding network for marking, our approach involves adding adjustable frequency perturbations to the low-frequency components of the cover image for marking. This strategy minimizes the impact on high-frequency elements of the image, thereby reducing high-frequency artifacts and enhancing the effectiveness of watermark concealment.
- A series of experiments validate the efficacy of our proposed method. Across various tasks, such as paint transfer and de-raining, our approach exhibits a notable absence of artifacts, demonstrating its strong concealment characteristics in both the spatial and frequency domains of the image.

The rest of this paper is organized as follows. We introduce preliminaries and the proposed method in Section 2, followed by the experimental results and analysis in Section 3. The discussion is provided in Section 4. Finally, we conclude this work in Section 5.

2. Preliminaries and Proposed Method

2.1. Detection of High-Frequency Artifacts

High-frequency artifacts manifest as unnatural and inconsistent textures, such as box-like patterns, in the frequency domain. In this section, we show the high-frequency artifacts present in the images that marked by the generative model watermarking, thus revealing the flaws in their covertness. We employed a frequency domain detection method [23] to detect high-frequency artifacts. Specifically, we used the Type II Discrete Cosine Transform (2D-DCT) to generate a DCT spectrum, which was then represented as a heatmap. Within this heatmap, the size of each data point indicates the coefficient of the respective spatial frequency, with larger values leading to brighter points. Frequency levels gradually increase from left to right and from top to bottom on the heatmap. The upper left section of the heatmap corresponds to lower frequencies, while the lower right section represents higher frequencies. In our experiments, we used two datasets: the Danbooru2019 dataset [27] and the De-raining dataset [28], to evaluate the previous generative model watermarking methods [17,19].

For the spatial domain, the evaluation results for these datasets are depicted in Figure 1. In Figure 1a, randomly selected ground-truth images from the Danbooru2019 dataset are showcased, while in Figure 1b, carrier images generated by the generative network [29] are presented. Figure 1c,d display images marked by [19] and [17], respectively. It is evident that the images in Figure 1a,b exhibit high similarity, indicating that the generative network [29] effectively accomplishes the image task. Simultaneously, the images in Figure 1c,d also closely resemble those in Figure 1a,b, suggesting that the previous generative model watermarking method successfully achieved the watermarking task without compromising the quality of the spatial domain in the original image. Figure 1e–h demonstrates similar results.

More importantly, our investigation delves into the characteristics and variations of the test images in the frequency domain. Typically, in natural images, low-frequency components depict smoothly varying image intensities and carry the majority of the image information [30,31]. Conversely, higher-frequency components approximate image edges, representing abrupt pixel transitions. The DCT heatmap for the various methods is presented at the bottom of Figure 1. At the bottom of Figure 1a, the DCT heatmap for the natural image are depicted. In the bottom of Figure 1b, the DCT heatmap for the generated image are shown, presenting box-like high-frequency artifacts. In the bottom of Figure 1c,d, the DCT heatmap for the marked images displays even more pronounced high-frequency artifacts. Similar results are observed at the bottom of Figure 1e–h. Therefore, through the DCT frequency domain analysis, we identify severe high-frequency artifacts in the marked images generated by existing generative model watermarking methods, highlighting a deficiency in frequency domain imperceptibility.

In summary, prior generative model watermarking methods have focused on the invisible embedding of watermarks in the spatial domain of images, making it difficult for potential attackers to discern the model watermarking within this spatial context. However, the identification of high-frequency artifacts in the frequency domain of the marked images renders generative model watermarking conspicuous within this frequency domain, thereby compromising the overall imperceptibility and security of the generative model watermarking system. Consequently, generative model watermarking becomes susceptible to detection by potential attackers. This situation underscores the imperative need for enhancing frequency-based imperceptibility in generative model watermarking.



Figure 1. Visual examples of paint transfer and de-raining. The left column (**a**–**d**) depicts paint transfer, while the right column (**e**–**h**) illustrates de-raining. (**a**) Ground-truth images from dataset [27], (**b**) carrier images generated by generative network [29], (**c**) marked images generated by method [19], (**d**) marked images generated by method [17], (**e**) ground-truth images from dataset [28], (**f**) carrier images generated by generative network [28], (**g**) marked images generated by method [19], (**h**) marked images generated by method [17]. The bottom part of the images presents the corresponding DCT heatmaps (the heatmap is the average of the corresponding test dataset).

2.2. Origins of High-Frequency Artifacts

In this section, we investigate the origins of high-frequency artifacts in the current generative model watermarking [17,19], building upon insights from previous research. We identify two primary sources of these artifacts: the watermark embedding network and the watermark embedding process. Odena et al. [32] attributed grid-like anomalies in spatial images to the up-sampling process. Similarly, Frank et al. [23] explicitly highlighted that incorporating up-sampling within the generative network introduces high-frequency artifacts in the frequency domain. These findings collectively suggest that the use of CNNs for image generation inherently results in structural high-frequency artifacts. Thus, we contend that the prevalent CNN-based watermark embedding networks in generative model watermarking inevitably introduce high-frequency artifacts when generating marked images. Moreover, Zeng et al. [24] observed that employing image patching triggers in backdoor attacks also leads to the emergence of high-frequency artifacts. Further insights from Zhang et al. [21] explained that embedding information using deep neural networks essentially adds coded high-frequency information to the carrier image. We posit that the watermark embedding process in generative model watermarking shares similarities with backdoor attacks or DNN-based information embedding. This process requires introducing perturbations to the marked image, leaving a distinguishable trace in the frequency domain due to time–frequency coherence. Without appropriate constraints, watermarking perturbations tend to exhibit high-frequency characteristics, inevitably introducing high-frequency artifacts into the marked image.

In summary, existing model watermarking techniques employ watermark embedding networks that inherently introduce high-frequency artifacts. Additionally, the addition of watermarking perturbations naturally gives rise to high-frequency artifacts. In light of these findings, we propose an image watermarking technology designed to suppress high-frequency artifacts.

2.3. Method Overview

As mentioned previously, existing generative model watermarking techniques [17,19] suffer from noticeable high-frequency artifacts in the frequency domain, which compromise the overall concealment of the embedded watermark. In response to this challenge, we

introduce an innovative image watermarking framework. At its core, this framework revolves around bypassing the watermark embedding network to mitigate the inherent high-frequency artifacts caused by CNNs. We design a frequency perturbation generation network to create frequency perturbations, which are then incorporated into the carrier image to achieve watermarking. Finally, the watermark is extracted from the marked image using a watermark extraction network.

Our proposed framework comprises two primary components: a frequency perturbation generation network F, and a watermark extraction network E. As illustrated in Figure 2, the frequency generation network takes fixed watermarks as input and generates frequency perturbations. Given a $256 \times 256 \times 3$ carrier image (i.e., the image generated by the generating network to be protected), we target the central $128 \times 128 \times 3$ portion of its frequency domain as the low-frequency component. To insert a frequency perturbation into this low-frequency area, we create a $128 \times 128 \times 3$ perturbation and incorporate it into the carrier image's frequency domain to achieve the watermarking. Importantly, the frequency domain perturbation introduced only affects the $128 \times 128 \times 3$ low-frequency region at the center of the carrier image's frequency domain. Thus, our approach effectively embeds the frequency domain perturbation into the carrier image's low-frequency component. By controlling the embedding location of the frequency domain perturbation, we selectively modify the carrier image's low-frequency regions while leaving the high-frequency regions unaltered. This targeted strategy effectively mitigates the high-frequency artifacts commonly associated with the watermark embedding process. Moreover, by bypassing the watermark embedding network, we ensure that the carrier image remains free from the high-frequency artifacts introduced by CNNs. Finally, through joint training, the watermark extraction network is trained to accurately retrieve the watermark from marked images.

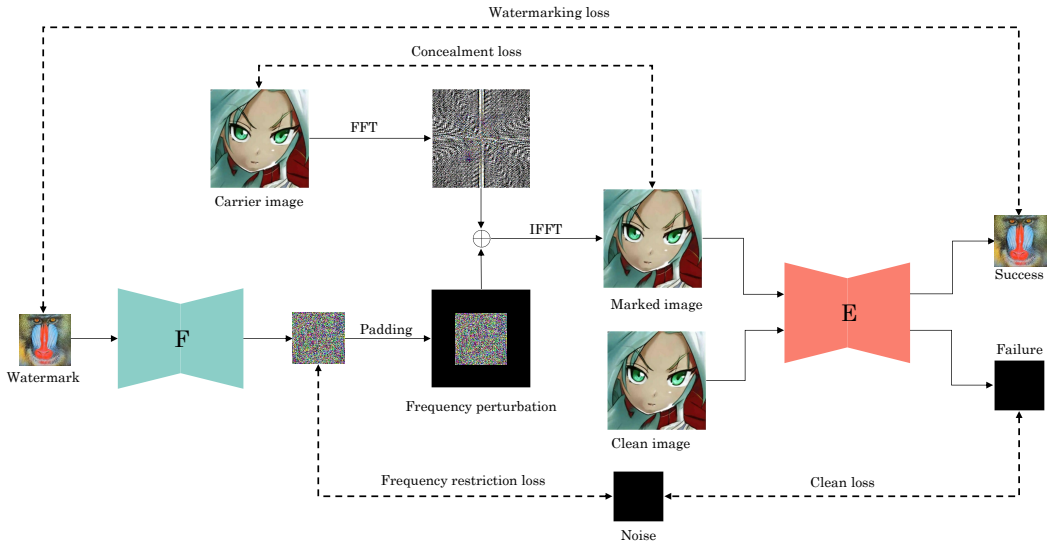


Figure 2. The diagram illustrates our proposed image watermarking framework, comprising a frequency generation network F and a watermark extraction network E. In our research, we use carrier images (i.e., the image generated by the generating network to be protected) with dimensions of $256 \times 256 \times 3$. The watermark is set to $128 \times 128 \times 3$, and the size of the frequency perturbation generated by network F is also $128 \times 128 \times 3$. Before adding this perturbation to the frequency domain of the carrier image, we must first pad it to match the dimensions of the carrier image, which are $256 \times 256 \times 3$. We applied the FFT operation with a shift (centering on low frequency) to transform the image into the frequency domain, and subsequently used IFFT to return the frequency domain representation to the spatial domain.

2.4. Framework

The proposed framework is shown in Figure 2. Carrier images are images generated by the protected model. A fixed watermark is input into the frequency perturbation generation

network F, and we expect F to generate the frequency perturbation f. As shown in Figure 2, we constrain the dimensions of f to $128 \times 128 \times 3$ and add it to the low-frequency domain of the carrier image. To match the frequency domain of the carrier image, we apply padding to the frequency perturbation (the filled area is set to 0). To obtain the frequency domain of the carrier image, we employ the Fast Fourier Transform (FFT). We add f to the frequency domain and subsequently perform an Inverse Fast Fourier Transform (IFFT) to convert it back to the spatial domain, resulting in the marked image. Similarly, our objective is to make the marked image as similar as possible to the carrier image, subject to the constraints of the loss function, while minimizing the impact of the frequency perturbation on the marked image. Lastly, we input the marked images into a watermark extraction network E, with the expectation that E learns how to extract watermarks from marked images. To prevent overfitting of the watermark extraction network E, we provide it with some negative samples from which we aim to extract noise. The above goals will be achieved through joint training of the watermark extraction network and the frequency perturbation generation network.

In our paper, we employ a Unet-like network structure [33] for the frequency perturbation generation network F, with the input and output dimensions set to $128 \times 128 \times 3$. For the watermark extraction network E, we utilize a ResNet-like architecture similar to the one used in paper [34], with input and output dimensions of $256 \times 256 \times 3$ and $128 \times 128 \times 3$. To transform signals from the spatial domain to the frequency domain and vice versa, we employ fundamental signal analysis techniques, namely the FFT for the former and the IFFT for the latter.

2.5. Loss Functions

As an image watermarking framework, the frequency perturbation generation network F is responsible for generating the frequency perturbation f. This perturbation is subsequently added into the frequency domain of the carrier image C, yielding the marked image C'. And C' is input into the watermark extraction network E, with the objective of extracting the watermark. It is crucial that, when E receives a clean image, it should generate a noise. To achieve these objectives, our design incorporates four key loss functions: the frequency restriction loss, watermarking loss, clean loss, and concealment loss. These losses play a pivotal role in optimizing the watermarking process and ensuring its effectiveness.

2.5.1. Frequency Restriction Loss

f represents the perturbation generated by the frequency perturbation network F:

$$f = F(wm), \tag{1}$$

where $F(\cdot)$ represents the output of the frequency perturbation network F and wm indicates the fixed watermark. To minimize the effect of adding the perturbation f to the carrier image, we aim to minimize the frequency perturbation f as much as possible. Therefore, we want to optimize:

$$\mathcal{L}_1 = \sum ||f||_1, \tag{2}$$

We will use ℓ_1 norm as the distance measure by default, unless otherwise specified.

2.5.2. Watermarking Loss

To watermark the carrier image C, it is necessary to add the frequency perturbation f into the frequency domain of C. Following this, we transform the frequency domain back to the spatial domain, yielding the marked image C'. The C' is

$$C' = IFFT(f \cdot \alpha + FFT(C)), \tag{3}$$

where α denotes the tunable hyperparameter. $FFT(\cdot)$ denotes the Fast Fourier Transform and $IFFT(\cdot)$ denotes the Inverse Fast Fourier Transform.

More importantly, To be able to extract a watermark, we use the watermarking loss

$$\mathcal{L}_2 = \frac{1}{|N|} \sum_{x \in C'} ||E(x) - wm||_1. \tag{4}$$

where x denotes the marked images that belong to C'. N denotes the pixel numbers. $E(x)$ represents the watermark recovered from the marked image by the watermark extraction network E.

2.5.3. Clean Loss

To prevent overfitting of the watermark extraction network, we must introduce a clean loss term:

$$\mathcal{L}_3 = \frac{1}{|N|} \sum_{x \in C} ||E(x) - noise||_1. \tag{5}$$

where x belongs to the clean carrier images C and we want the watermark extraction network to extract the noise from the clean image. In this paper, *noise* defaults to zero.

2.5.4. Concealment Loss

To embed the watermark in a spatially invisible way, we have to minimize the visual distance between the carrier image and the marked image, and the concealment loss can be expressed as

$$\mathcal{L}_4 = \frac{1}{|N|} \sum ||C' - C||_1. \tag{6}$$

Lastly, the frequency domain perturbation generation network and the watermark extraction network will be trained jointly with the total loss as shown below:

$$\mathcal{L}_{total} = \beta_1 \mathcal{L}_1 + \beta_2 \mathcal{L}_2 + \beta_3 \mathcal{L}_3 + \beta_4 \mathcal{L}_4, \tag{7}$$

3. Experimental Results and Analysis

3.1. Setup

Dataset. To validate the effectiveness of our approach, we conducted evaluations on two distinct tasks: paint transfer [29] and de-raining [28]. To protect paint transfer model, we employed the Danbooru2019 dataset [27] and adopted the training methodology [29]. Following the training of our paint transfer model, we randomly sampled 4000 images (i.e., carrier images) generated by the network, with 3000 allocated for image watermarking training, 500 for validation, and 500 for testing. To protect de-raining model, we employed the dataset [28] and followed their training procedure. Following the training of our de-raining model, we randomly selected 2000 images (i.e., carrier images) generated by the network, with 1800 were allocated for image watermarking training, 100 for validation, and 100 for testing.

Parameter Setting. All images were standardized to a size of $256 \times 256 \times 3$. We respectively utilized "Baboon" and "MDPI" as watermark images, with a fixed watermark size of $128 \times 128 \times 3$. Regarding adjustable parameters, we empirically set $\alpha = 5 \times 10^5$, $\beta_1 = \beta_3 = \beta_4 = 1$, $\beta_2 = 5$. We used the Adam optimizer for training with a learning rate of 2.0×10^{-4}. Our method was implemented on a single TITAN RTX GPU with CuDNN acceleration.

Evaluation Metrics. There are two commonly used metrics to assess the quality of marked images, Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). Bit Error Rate (BER) is a measure of the quality of binary watermark reconstruction. In addition, to evaluate the performance of watermark extraction, we define a new metric called success rate (SR). Watermark extraction is considered successful if the corresponding PSNR is greater than 25 dB or the corresponding BER is less than 1.0×10^{-2}. Finally, the DCT detection results are utilized to measure the frequency-domain invisibility of the image watermarking system. We define the frequency domain invisibility performance (FP), which

is expressed by calculating the Learned Perceptual Image Patch Similarity (LPIPS) [35] between the DCT detection result of the marked image and the DCT detection result of the ground-truth image, with the lower value of the metric being better.

3.2. Qualitative and Quantitative Results

The main objective of this paper is to mitigate the impact of high-frequency artifacts on generative model watermarking, focusing on enhancing imperceptibility and security. It is important to note that strong imperceptibility inherently enhances system security. To achieve this goal, we propose an innovative image watermarking framework that addresses imperceptibility in generative model watermarking within both the spatial and frequency domains. In this section, we will demonstrate the excellent imperceptibility of our proposed framework in both the spatial and frequency domains.

3.2.1. Spatial Invisibility

In this subsection, we assess the performance of our proposed framework by computing the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) between the carrier image and the marked image. Higher PSNR and SSIM values indicate a closer resemblance between the carrier and marked images, indicating that our framework preserves the quality of the carrier image during watermark embedding. Visual representations of the proposed framework are presented in Figure 3. In the spatial domain, the marked images are nearly indistinguishable from the carrier images, highlighting our framework's ability to proficiently embed watermarks in the frequency domain without compromising spatial image quality. Our approach achieves commendable visual results, ensuring excellent imperceptibility. Additionally, the extraction of high-quality watermarks from marked images is important for reliable ownership verification. The extracted watermarks from the marked images generated by our framework are also showcased in Figure 3, both exhibiting favorable visual results.



Figure 3. Visual examples of paint transfer and de-raining. The left column (**a**–**c**) depicts paint transfer, while the right column (**d**–**f**) illustrates de-raining. The top two rows display the results of embedding "Baboon," while the bottom two rows demonstrate the results of embedding "MDPI". (**a**) Represents the carrier images in paint transfer, (**b**) shows the marked images in paint transfer, (**c**) displays the extracted watermarks in paint transfer,(**d**) represents the carrier images in de-raining, (**e**) shows the marked images in de-raining, (**f**) displays the extracted watermarks in de-raining.

In further quantitative evaluation, as illustrated in Tables 1 and 2, the quality of both the marked images and the extracted watermarks is assessed. Note that, for the quality of the extracted watermark, we use PSNR and BER (Bit Error Rate) to evaluate it. These metrics demonstrate the high quality of both the marked images (>50 dB) and the extracted watermarks(>40 dB or ≤0.001). And the success rate (SR) of watermark extraction is 100%. Our image watermarking framework effectively ensures imperceptibility in the spatial domain, upholds carrier image quality, and maintains watermark extraction performance.

Table 1. Quality assessment for the marked images and the extracted watermarks. All experimental results shown in this Table are mean values. "DE" means "de-raining" and "PT" means "Paint Transfer". PSNRw measures the quality of the extracted color watermarks.

Task	Watermark	Mean PSNR	Mean SSIM	PSNRw	SR
PT	*Baboon*	50.58 dB	0.998	42.88 dB	100%
DE	*Baboon*	53.34 dB	0.999	46.47 dB	100%

Table 2. Quality assessment for the marked images and the extracted watermarks. All experimental results shown in this Table are mean values. "DE" means "de-raining" and "PT" means "Paint Transfer". BER measures the quality of the extracted binary watermarks.

Task	Watermark	Mean PSNR	Mean SSIM	BER	SR
PT	*MDPI*	57.18 dB	0.999	0.0009	100%
DE	*MDPI*	54.59 dB	0.999	0.0010	100%

3.2.2. Frequence Invisibility

Our primary objective is to achieve imperceptibility of watermarking in the frequency domain by mitigating high-frequency artifacts. To illustrate the good imperceptibility of our novel image watermarking technique in the frequency domain, we examine the DCT heatmaps of the marked images generated by our framework, as displayed in Figure 4. It is evident that the DCT heatmap results of our framework closely mirror those of natural images at the bottom of Figure 1a,e, indicating that our proposed framework produces images that are exceptionally natural and devoid of high-frequency artifacts. In contrast, previous model watermarking methods, as depicted in at the bottom of Figure 1c,d,g,h, exhibit small square high-frequency artifacts, while our method conspicuously eliminates such artifacts.

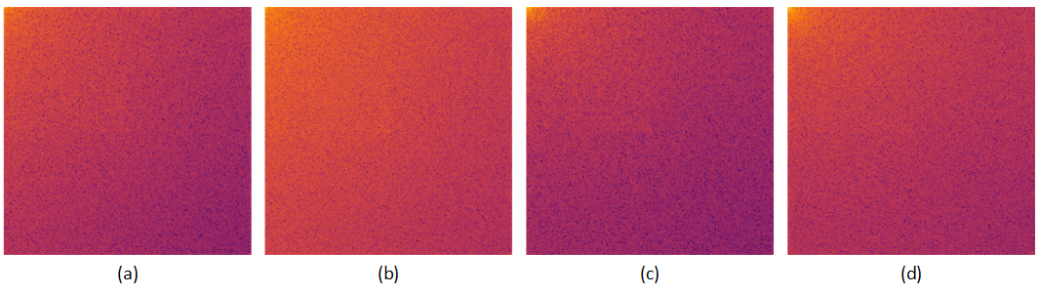


Figure 4. DCT heatmap of our method (average results): (**a**) the DCT heatmaps of embedding "Baboon" in paint transfer, (**b**) the DCT heatmaps of embedding "MDPI" in paint transfer, (**c**) the DCT heatmaps of embedding "Baboon" in de-raining, (**d**) the DCT heatmaps of embedding "MDPI" in de-raining.

In addition, we evaluate the frequency domain invisibility performance (FP).The results are shown in Figure 5. Where "Wu" denotes the method [17] and "Zhang" represents another method [19]. And "Original" represents the benchmark host network that performs specific image processing tasks such as paint transfer [29] and de-raining [28]. We observe

good performance across various tasks with our proposed method. For example, in the de-raining task with the "MDPI" watermark, our approach outperforms method [17] by 44%, showing a 24.9% improvement in average performance compared to method [17] and a 7.2% improvement over method [19]. These observations reveal that our method produces images with smooth surfaces, avoiding the introduction of high-frequency artifacts, a notable advantage.

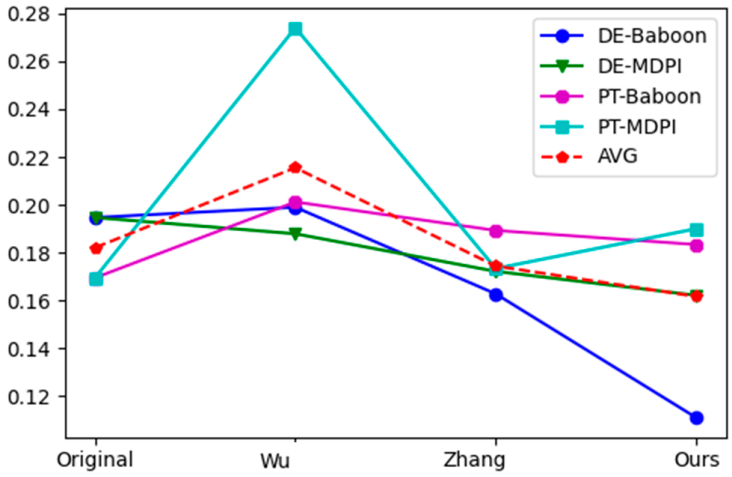


Figure 5. Frequency domain invisible performance (FP) of different methods on different tasks. Before—denotes the corresponding task, after—denotes the corresponding watermark. DE denotes "de-raining", PT denotes "paint-transfer". Watermarks include Baboon and IEEE [17,19].

This underscores the superior imperceptibility of our proposed framework in the frequency domain. In summary, based on the collective evidence from these experiments, we can assert that our method effectively reduces high-frequency artifacts, preserving the concealment of the embedded watermark in both the frequency and spatial domains, all while upholding the visual quality of the labeled image.

3.3. Robustness

A robust watermarking system must be able to counter attacks encountered in real-world scenarios. In this paper, our focus is on pre-processing attacks, which aim to manipulate the marked image to hinder watermark extraction. To enhance the robustness of our system against these attacks, a common strategy involves including pre-processed images in the training set. Due to limited computational resources, our study focuses on four common attacks: filtering, resizing, cropping, and flipping.

In the filtering attack, an adversary with prior knowledge of the frequency domain may attempt to eliminate the embedded watermark by filtering out high-frequency components from the marked image. To simulate this attack, we filter out high-frequency components from marked images. Specifically, for each channel of a marked image, we calculate the FFT coefficients. We preserve the central region, spanning size $[128^2, 256^2]$, while zeroing out the other coefficients. For the resizing attack, we randomly alter the input images to smaller or larger resolutions within the range $[128^2, 512^2]$ and then restore these manipulated images to their original size to extract the watermark. In the cropping attack, we retain only certain pixels within the range $[224^2, 256^2]$, setting the rest to zero to mimic cropping. In the flipping attack, the marked image undergoes a horizontal or vertical inversion. It is important to note that, while our paper tests a limited number of common attacks, there is potential for further exploration, given sufficient computational resources.

Figure 6 present examples of pre-processed images and the corresponding extracted watermarks. These examples illustrate that the embedded watermark can still be extracted with satisfactory quality despite various pre-processing operations. This underscores the effectiveness of adversarial training in enhancing robustness against pre-processing attacks.

Quantitative results in Tables 3–5 indicate that, while watermark quality may degrade as attack intensity increases, we generally maintain an acceptable level of quality. These results affirm that our proposed framework effectively withstands pre-processing attacks through the application of adversarial training.

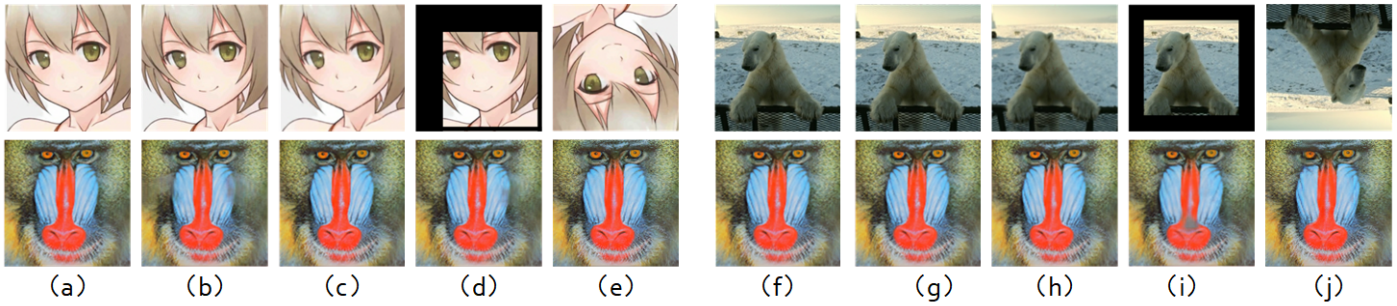


Figure 6. Examples of attacked images (top) and extracted watermarks (bottom) for paint transfer and de-raining. (**a**) the original marked image and the extracted watermark in paint transfer. (**b**) the filtered image (retains frequency domain details within the $128 \times 128 \times 3$ region) and the extracted watermark in paint transfer. (**c**) the resized image (resized to $148 \times 148 \times 3$) and the extracted watermark in paint transfer. (**d**) the cropped image ($cropsize = 64$) and the extracted watermark in paint transfer. (**e**) the rotated image and the extracted watermark in paint transfer. (**f**) The original marked image and the extracted watermark in de-raining. (**g**) The filtered image (retains frequency domain details within the $128 \times 128 \times 3$ region) and the extracted watermark in de-raining. (**h**) The resized image (resized to $148 \times 148 \times 3$) and the extracted watermark in de-raining. (**i**) the cropped image ($cropsize = 64$) and the extracted watermark in de-raining. (**j**) The rotated image and the extracted watermark in de-raining.

Table 3. Mean PSNRs (extracted color watermarks from the attacked images, dB). "PT" means "Paint Transfer" and "DE" means "de-raining".

Task	Watermark	Filtering			Resizing			Cropping			Flipping	
		$148^2 \times 3$	$160^2 \times 3$	$192^2 \times 3$	$128^2 \times 3$	$196^2 \times 3$	$512^2 \times 3$	Size = 16	Size = 32	Size = 64	Horizontal	Vertical
PT	*Baboon*	39.23 dB	40.88 dB	41.04 dB	32.98 dB	37.93 dB	40.16 dB	30.87 dB	33.64 dB	37.08 dB	28.85 dB	29.85 dB
DE	*Baboon*	44.85 dB	45.20 dB	45.52 dB	28.37 dB	44.30 dB	45.59 dB	25.55 dB	29.20 dB	32.92 dB	6.90 dB	36.12 dB

Table 4. Mean BERs (extracted binary watermarks from the attacked images). "PT" means "Paint Transfer" and "DE" means "de-raining".

Task	Watermark	Filtering			Resizing			Cropping			Flipping	
		$148^2 \times 3$	$160^2 \times 3$	$192^2 \times 3$	$128^2 \times 3$	$196^2 \times 3$	$512^2 \times 3$	Size = 16	Size = 32	Size = 64	Horizontal	Vertical
PT	*MDPI*	0.0025	0.0011	0.0010	0.0010	0.0011	0.0012	0.0010	0.0009	0.0009	0.0017	0.0021
DE	*MDPI*	0.0016	0.0014	0.0013	0.0028	0.0013	0.0015	0.0103	0.0116	0.0110	0.0026	0.0022

Table 5. SR SR against different preprocessing operations. "PT" means "paint transfer" and "DE" means "de-raining.

Task	Watermark	Filtering			Resizing			Cropping			Flipping	
		$148^2 \times 3$	$160^2 \times 3$	$192^2 \times 3$	$128^2 \times 3$	$196^2 \times 3$	$512^2 \times 3$	Size = 16	Size = 32	Size = 64	Horizontal	Vertical
PT	*Baboon*	94.2%	97.4%	97.6%	74%	94%	99.6%	95.6%	83.8%	75.4%	81.2%	83.4%
PT	*MDPI*	79.2%	87%	88.6%	69.2%	80.2%	84.8%	73.8%	73%	71.4%	88.8%	85%
DE	*Baboon*	94%	95.5%	97%	72%	92.5%	96%	80.5%	78.5%	64%	75.5%	77.5%
DE	*MDPI*	89.5%	90%	90.5%	80%	83%	89.5%	76.5%	72.5%	64.5%	88%	84.5%

3.4. Comparisons with Previous Model Watermarking Methods

In this section, we present a comparative analysis between our proposed method and prior approaches [17,19]. A comprehensive summary of the comparative results

is provided in Table 6, with a focus on resilience against pre-processing attacks and a watermark's imperceptibility in both spatial and frequency domains. Within Table 6, "Yes" denotes that the method exhibits robustness against the corresponding attack or achieves imperceptibility in the relevant domain, while "Partially" indicates that some processed images demonstrate good imperceptibility, while others may exhibit noticeable visual artifacts. Our analysis reveals that, in contrast to previous methods, which only partially satisfy these criteria, our proposed method successfully fulfills all the established objectives. This encompasses resistance against pre-processing attacks and ensuring that distortions caused by concealed watermarks remain imperceptible in both spatial and frequency domains.

Table 6. Comparison between different model watermarking methods in terms of robustness and imperceptibility.

Method	Robustness against Pre-processing Attacks	Imperceptibility	
		Spatial Domain	Frequency Domain
Ref. [17]	Yes	Partially	
Ref. [19]		Yes	
Proposed	Yes	Yes	Yes

3.5. Comparisons with Traditional Watermarking Methods

We have proposed an image watermarking technique that considers frequency domain imperceptibility, aimed at protecting the copyright of images generated by generative networks. In this section, we wish to highlight the distinctions between our proposed method and traditional image watermarking techniques, summarized as follows:

Novelty of the research problem: focusing on the high-frequency artifacts problem in generative model watermarking, an image watermarking algorithm that can reduce high-frequency artifacts and improve the effect of copyright protection is proposed.

Innovativeness of the solution: adopting an end-to-end deep learning approach, the watermark is directly embedded and extracted using neural networks to avoid high-frequency artifacts and improve concealment. In contrast, traditional methods require in-depth understanding of various mathematical techniques and careful design of embedding/extraction methods.

Superior Performance: We have compared our method with traditional image watermarking techniques based on DWT-SVD-DCT [36]. Due to computational constraints, we only compare the case of embedding "MDPI" in the "de-raining" task. As shown in Table 7, traditional image watermarking techniques tend to be limited in watermark capacity due to the need for carefully designed embedding methods. In contrast, our deep learning-based image watermarking algorithm leverages the capabilities of neural networks to embed a large capacity of watermarks easily. Moreover, our proposed method outperforms in terms of spatial image quality, frequency domain invisibility, and watermark extraction quality.

Table 7. Comparative results of the proposed method with the method [36]. "DE" means "de-raining" and BER measures the quality of binary watermarks.

Method	Task	Watermark	Size	Mean PSNR	FP	BER
method [36]	DE	*MDPI*	128^2	54.59 dB	0.1622	0.0010
Ours	DE	*MDPI*	32^2	47.62 dB	0.1790	0.1291

3.6. Importance of Clean Loss

We have reason to believe that omitting the clean loss from the loss functions could lead to the problem of overfitting in the watermark extraction network, potentially allowing it to erroneously extract watermarks from any image. To empirically investigate this

hypothesis, we conducted an experiment where we trained two watermark extraction networks: "E", which incorporated the clean loss, and "$E1$", which excluded the clean loss. Subsequently, we input various clean images into $E1$ and observed that it mistakenly retrieved watermarks from these clean images, as illustrated in Figure 7. In contrast, E adeptly extracts watermark and noise from watermarked and clean images. Consequently, clean loss prevents the problem of overfitting in the watermark extraction network, making our copyright protection highly effective and practically relevant.

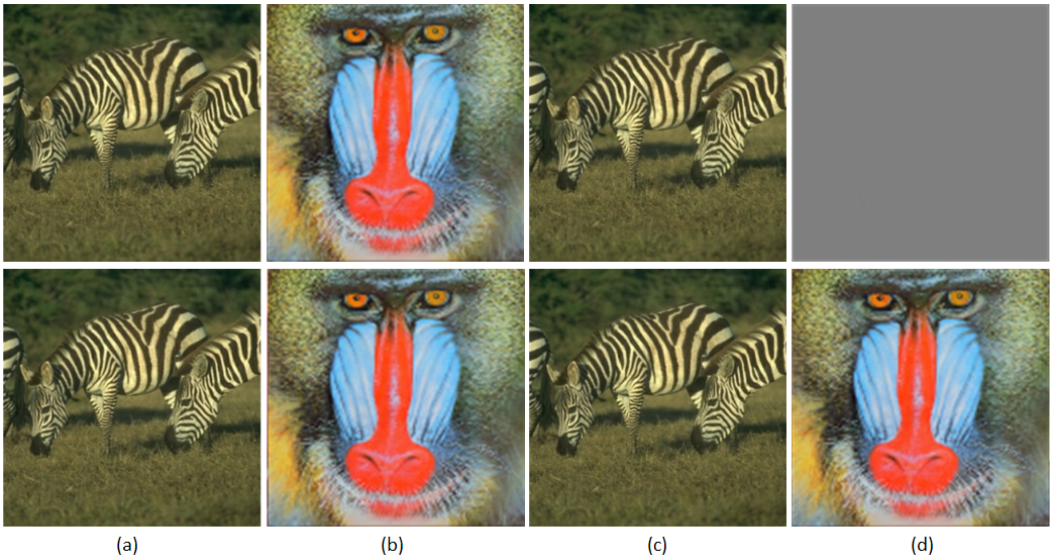


Figure 7. Experiment results of clean loss. The top row presents the results for "E", and the second row displays the results for "$E1$". (**a**) Marked image, (**b**) extracted results from the marked image, (**c**) clean image, (**d**) extracted results from the clean image.

4. Discussion

In the above paper, we have only considered four common pre6processing attacks given the limitation of arithmetic resources. It must be frankly admitted that our proposed framework does not guarantee absolute robustness in the face of a wide variety of real-world attacks. This is because we cannot predict and respond to all potential adversary attack strategies, and there are always attacks that may target weaknesses in our approach. For example, our framework shows some vulnerability against additive noise attacks due to the fact that the addition of noise can cause significant changes in the frequency domain, which in turn destroys the consistency upon which watermark extraction depends. In fact, existing methods also mostly show strong defense only in the face of specific attacks.

However, by reducing the artifacts in the watermarking system, we greatly enhance the covertness of the watermark and effectively hide the existence of the watermark, which not only enhances the imperceptibility of the watermark, but also reduces the possibility of arousing the suspicion of the attacker. In other words, by enhancing the covertness of watermarks, we significantly reduce the risk of watermarks being attacked, thereby enhancing the reliability of deep neural network (DNN) model-based IP protection. It is worth noting that our proposed method is not limited to protecting images generated only by generative models. Instead, it is an image watermarking technique that is applicable to protect any image. We expect that this exploration will inspire more innovative and advanced research in the future.

5. Conclusions

The remarkable success of DNN models underscores the increasing importance of safeguarding the IP of these advanced DNN models. To prevent potential infringements upon IP rights by DNN models, many researchers have employed model watermarking

techniques. Recent efforts in this domain have been dedicated to the protection of generative models, which apply watermarking to their outputs to preserve the integrity of the host network. In this study, we find that existing generative model watermarking methods yield conspicuous high-frequency artifacts within the images generated by the generative model. These artifacts significantly impair the imperceptibility and reliability of generative model watermarking systems. This discovery highlights the need for further research and advancements in the frequency domain to enhance the capabilities of watermark concealment. In response to this challenge, we introduce an image watermarking applicable to image processing models. Our approach diverges from the traditional reliance on CNNs as watermark embedding networks. Instead, we introduce a frequency perturbation generation network designed to create low-frequency perturbations. These perturbations are then incorporated as watermarks within the low-frequency components of the carrier image. This strategy minimizes the impact of the watermark embedding process on the high-frequency attributes of the image. As a result, our approach effectively suppresses the high-frequency artifacts in the marked image, significantly improving the imperceptibility of watermarking. The proposed method achieves good image quality (>50 dB) in the spatial domain, coupled with a 100% success rate in watermark extraction under attack-free conditions. Compared with the previous method, the average frequency domain invisibility performance (FP) can be improved by up to 24.9%. It enhances the safeguarding of the IP within generative models and elevates the effectiveness and robustness of watermark extraction through adversarial training. Our work aims to contribute to the advancement of research in watermarking and promote the adoption of more effective methods for IP protection.

Author Contributions: Conceptualization, L.Z. and X.Z.; methodology, L.Z.; software, L.Z.; validation, L.Z.; supervision, H.W.; project administration, H.W.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC) under Grants U22B2047.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The datasets can be found here: https://gwern.net/danbooru2021#danbooru2019 (accessed on 19 January 2024) and https://github.com/ZhangXinNan/RainDetectionAndRemoval (accessed on 19 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4217–4228. [CrossRef] [PubMed]
2. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: New York, NY, USA; 2022; Volume 35, pp. 27730–27744.
3. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 10674–10685. [CrossRef]
4. Uchida, Y.; Nagai, Y.; Sakazawa, S.; Satoh, S. Embedding Watermarks into Deep Neural Networks. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ICMR'17, New York, NY, USA, 6–9 June 2017; pp. 269–277. [CrossRef]
5. Darvish Rouhani, B.; Chen, H.; Koushanfar, F. DeepSigns: An End-to-End Watermarking Framework for Ownership Protection of Deep Neural Networks. In Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS'19, New York, NY, USA, 13–17 April 2019; pp. 485–497. [CrossRef]
6. Fan, L.; Ng, K.W.; Chan, C.S. Rethinking Deep Neural Network Ownership Verification: Embedding Passports to Defeat Ambiguity Attacks. In *Advances in Neural Information Processing Systems*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2019; Volume 32.

7. Li, Y.; Tondi, B.; Barni, M. Spread-Transform Dither Modulation Watermarking of Deep Neural Network. *J. Inf. Secur. Appl.* **2021**, *63*, 103004. [CrossRef]
8. Chen, H.; Rouhani, B.D.; Fu, C.; Zhao, J.; Koushanfar, F. DeepMarks: A Secure Fingerprinting Framework for Digital Rights Management of Deep Learning Models. In Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR'19, New York, NY, USA, 10–13 June 2019; pp. 105–113. [CrossRef]
9. Wang, T.; Kerschbaum, F. RIGA: Covert and Robust White-Box Watermarking of Deep Neural Networks. In Proceedings of the Web Conference 2021, WWW'21, New York, NY, USA, 19–23 April 2021; pp. 993–1004. [CrossRef]
10. Tartaglione, E.; Grangetto, M.; Cavagnino, D.; Botta, M. Delving in the loss landscape to embed robust watermarks into neural networks. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 1243–1250. [CrossRef]
11. Szyller, S.; Atli, B.G.; Marchal, S.; Asokan, N. DAWN: Dynamic Adversarial Watermarking of Neural Networks. In Proceedings of the 29th ACM International Conference on Multimedia, MM'21, New York, NY, USA, 20–24 October 2021; pp. 4417–4425. [CrossRef]
12. Adi, Y.; Baum, C.; Cisse, M.; Pinkas, B.; Keshet, J. Turning Your Weakness Into a Strength: Watermarking Deep Neural Networks by Backdooring. In Proceedings of the 27th USENIX Security Symposium (USENIX Security 18), Baltimore, MD, USA, 15–17 August 2018; pp. 1615–1631.
13. Chen, H.; Rouhani, B.D.; Koushanfar, F. BlackMarks: Blackbox Multibit Watermarking for Deep Neural Networks. *arXiv* **2019**, arXiv:cs.MM/1904.00344.
14. Guo, J.; Potkonjak, M. Evolutionary Trigger Set Generation for DNN Black-Box Watermarking. *arXiv* **2021**, arXiv:cs.LG/1906.04411.
15. Li, P.; Cheng, P.; Li, F.; Du, W.; Zhao, H.; Liu, G. PLMmark: A Secure and Robust Black-Box Watermarking Framework for Pre-trained Language Models. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 14991–14999. [CrossRef]
16. Hua, G.; Teoh, A.B.J. Deep fidelity in DNN watermarking: A study of backdoor watermarking for classification models. *Pattern Recognit.* **2023**, *144*, 109844. [CrossRef]
17. Wu, H.; Liu, G.; Yao, Y.; Zhang, X. Watermarking Neural Networks With Watermarked Images. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 2591–2601. [CrossRef]
18. Zhang, J.; Chen, D.; Liao, J.; Fang, H.; Zhang, W.; Zhou, W.; Cui, H.; Yu, N. Model Watermarking for Image Processing Networks. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12805–12812. [CrossRef]
19. Zhang, J.; Chen, D.; Liao, J.; Zhang, W.; Feng, H.; Hua, G.; Yu, N. Deep Model Intellectual Property Protection via Deep Watermarking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 4005–4020. [CrossRef] [PubMed]
20. Zhang, L.; Liu, Y.; Liu, S.; Yang, T.; Wang, Y.; Zhang, X.; Wu, H. Generative Model Watermarking Based on Human Visual System. In Proceedings of the Digital Multimedia Communications, Shanghai, China, 8–9 December 2022; Zhai, G., Zhou, J., Yang, H., Yang, X., An, P., Wang, J., Eds.; Springer: Singapore, 2023; pp. 136–149.
21. Zhang, C.; Benz, P.; Karjauv, A.; Sun, G.; Kweon, I.S. UDH: Universal Deep Hiding for Steganography, Watermarking, and Light Field Messaging. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: New York, NY, USA; 2020; Volume 33, pp. 10223–10234.
22. Baluja, S. Hiding Images in Plain Sight: Deep Steganography. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.
23. Frank, J.; Eisenhofer, T.; Schönherr, L.; Fischer, A.; Kolossa, D.; Holz, T. Leveraging Frequency Analysis for Deep Fake Image Recognition. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; Volume 119, pp. 3247–3258.
24. Zeng, Y.; Park, W.; Mao, Z.M.; Jia, R. Rethinking the Backdoor Attacks' Triggers: A Frequency Perspective. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 16453–16461. [CrossRef]
25. Guo, J.; Potkonjak, M. Watermarking Deep Neural Networks for Embedded Systems. In Proceedings of the 2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Diego, CA, USA, 5–8 November 2018; pp. 1–8. [CrossRef]
26. Li, H.; Wenger, E.; Shan, S.; Zhao, B.Y.; Zheng, H. Piracy Resistant Watermarks for Deep Neural Networks. *arXiv* **2020**, arXiv:cs.CR/1910.01226.
27. Danbooru2019: A large-Scale Crowdsourced and Tagged Anime Illustration Dataset. 2019. Available online: https://gwern.net/danbooru2021#danbooru2019 (accessed on 1 December 2022).
28. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep Joint Rain Detection and Removal from a Single Image. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1685–1694. [CrossRef]
29. Lata, K.; Dave, M.; Nishanth, K.N. Image-to-Image Translation Using Generative Adversarial Network. In Proceedings of the 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 12–14 June 2019; pp. 186–189. [CrossRef]
30. Burton, G.J; Moorhead, I.R. Color and spatial structure in natural scenes. *Appl. Opt.* **1987**, *26*, 157–170. [CrossRef] [PubMed]
31. Tolhurst, D.J; Tadmor, Y.T.C. Amplitude spectra of natural images. *Ophthalmic Physiol. Opt.* **1992**, *12*, 229–232. [CrossRef] [PubMed]

32. Odena, A.; Dumoulin, V.; Olah, C. Deconvolution and Checkerboard Artifacts. *Distill* **2016**, *1*, e3. [CrossRef]
33. Wu, H.; Li, C.; Liu, G.; Zhang, X. Hiding data hiding. *Pattern Recognit. Lett.* **2023**, *165*, 122–127. [CrossRef]
34. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251. [CrossRef]
35. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595. [CrossRef]
36. He, Y.; Hu, Y. A Proposed Digital Image Watermarking Based on DWT-DCT-SVD. In Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 25–27 May 2018; pp. 1214–1218. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.