

## Article

# Weibull–Power Cauchy Modeling for Robust Transform-Domain Image Watermarking

Siyu Yang <sup>1,\*</sup> , Yufu Gao <sup>2</sup>  and Huiwen Zheng <sup>1</sup>

<sup>1</sup> School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China; 2405010305@xy.dlpu.edu.cn

<sup>2</sup> College of Marine Engineering, Dalian Maritime University, Dalian 116026, China; yufugao1998@163.com

\* Correspondence: yangsy\_1995@163.com

## Abstract

In a digital watermarking technology system, robustness, imperceptibility and payload capacity are the three core performance indicators that restrict one another. How to achieve the optimal balance between the three is still the key scientific challenge to be solved in this field. This paper proposes a digital watermarking algorithm based on the magnitude coefficient of Non-Subsampled Shearlet Transform Fast and Accurate Polar Harmonic Fourier Moments (NSST-FAPHFMs) and the Weibull–Power Cauchy (W-PC) statistical model. The algorithm consists of two stages: watermark embedding and detection. In the embedding phase, the original image is first decomposed by NSST multi-scale decomposition, and the high-frequency subbands are divided into non-overlapping blocks and partitioned. High-energy coefficient blocks are extracted to obtain NSST-FAPHFMs magnitude coefficient features, which serve as robust carriers for watermark embedding. In the detection phase, the W-PC distribution is used to accurately statistically model the above magnitude coefficients to characterize their heavy-tailed characteristics and strong correlation structure. Maximum likelihood estimation (MLE) is employed to estimate the model parameters, and a blind watermark detection mechanism is further constructed by integrating the W-PC model with the Local Optimal Detector (LOD) under the Neyman–Pearson (N-P) criterion. Experimental results show that the proposed algorithm has good imperceptibility, and the area under the receiver operating characteristic curve (AUROC) can reach 0.9991 without attack. The algorithm maintains strong robustness against various attacks and can effectively realize the joint optimization of the three core performance indicators of the watermarking system.

**Keywords:** NSST-FAPHFMs magnitude; Weibull–Power Cauchy; blind watermark detection mechanism; statistical model



Received: 23 April 2026

Revised: 8 June 2026

Accepted: 9 July 2026

Published: 16 July 2026

**Copyright:** © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

Digital watermarking has evolved into a compelling technical solution for addressing multimedia copyright protection and content authentication challenges, enabling the imperceptible integration of hidden information into host images without compromising their inherent visual integrity. From a functional perspective, digital watermarking technologies are mainly classified into two core types: detection-oriented and decoding-oriented. Watermark detection aims to verify the presence of a target watermark in a given image, whereas decoding focuses on the ability to accurately retrieve the concealed data content. For image copyright protection, mere confirmation of watermark existence suffices to establish legal ownership, making watermark detection the core focus of this study.

With the widespread dissemination of digital images, safeguarding copyright has become increasingly critical, yet balancing the three core indicators of watermarking systems—imperceptibility, robustness, and embedding capacity—remains a key challenge [1–3]. In order to solve this trade-off problem, recent related research has promoted the development of various transform-domain watermarking algorithms based on statistical modeling by mining the statistical characteristics of the image itself.

Spatial-domain watermarking embeds and extracts watermarks directly in the original pixel layer without complex transformation operations. It features low computational complexity, good real-time performance and easy implementation. However, it suffers from weak robustness and is vulnerable to compression, noise, filtering and other common distortions. Panigrahi et al. [4] proposed an image steganography technique based on the LSB (least significant bit), which combines symmetric encryption to hide text in grayscale and RGB images and is implemented with a simple GUI for encoding and decoding. When simultaneously implementing visible and invisible watermarks, invisible watermarks can generate QR codes for easy extraction. Li et al. [5] proposed a bipolar watermarking method, BiMarker. The generated text is divided into positive and negative poles, and the detection is enhanced by comparing the number of “green” marks in the two poles, without additional computing resources or prompt information, which effectively overcomes the problem of inaccurate estimation of non-watermark text distribution in traditional methods.

Transform-domain watermarking first decomposes an image using classical frequency-domain transformations such as DCT, DWT, contourlet, and NSCT, followed by watermark embedding via coefficient modification. This category of methods exhibits strong robustness, good imperceptibility, and excellent anti-attack capability. It can be further combined with statistical models to boost detection accuracy, making it the mainstream solution for robust watermarking. Nevertheless, it suffers from relatively high computational and model complexity. Marzieh et al. [6] proposed a color image multiplicative watermark detection method based on a Hidden Markov Model (HMM). In the contourlet domain, the HMM is employed to simultaneously capture inter-channel dependencies and cross-scale correlations. A detector is then constructed using the likelihood ratio test. Wang et al. [7] proposed an image watermarking detector based on a Weibull mixtures–HMT model. In this method, the watermark is embedded in the NSST-FPCET magnitude domain in a multiplicative way, the embedded region is selected by Otsu Canny edge detection, and the model parameters are estimated by variance-reduced random expectation maximization. Iqbal [8] proposed a multi-channel color image watermarking detection framework based on HMMs. In the contourlet domain, the HMM is used to model the dependency and scale correlation between RGB channels, which overcomes the limitation of ignoring channel correlation in traditional methods. The watermark is embedded in a multiplicative way, and the statistical hypothesis test is used for blind detection. Djordje et al. [9] proposed an image watermarking detection method based on the Rayleigh distribution and the corresponding embedding scheme. The watermark is embedded in  $8 \times 8$  blocks in the DCT domain by multiplication, and the block selection is based on variance and anti-JPEG compression ability. In the detection phase, using the theory of Local Optimal Detector, the Rayleigh distribution detector suitable for the statistical characteristics of the selected intermediate frequency coefficient is derived. Zhang et al. [10] proposed a dual watermarking algorithm for remote sensing image copyright protection and integrity authentication, which combines discrete wavelet transform (DWT), singular value decomposition (SVD), and Chebyshev chaotic mapping construction. A semi-fragile weak watermark with quantized index modulation (QIM) is embedded in the DWT high-frequency domain for tamper detection; robust watermarking is designed based on information entropy and SVD, with strong resistance to attacks, and can meet the security authentication requirements of

remote sensing images. Yuan et al. [11] proposed a color image quantum watermarking algorithm by combining the quantum discrete wavelet transform with Sinusoidal Tent mapping and Baker mapping. This method determines the embedding channel through chaotic sequences, uses quantum Haar wavelet decomposition, and embeds the scrambled watermark into high-frequency subbands. While ensuring high concealment and strong robustness of the watermark, it effectively reduces computational complexity and provides an efficient and feasible quantum implementation scheme for color image copyright protection. Niu et al. [12] proposed an image watermarking method based on the non-subsampled contourlet transform (NSCT) and pseudo-Jacobi–Fourier moment (PJFM) magnitudes, combined with the Weibull–Burr integrated bivariate distribution (WBIBD). The magnitude of PJFMs is calculated in the high-frequency subband of the NSCT, and its edge distribution and dependence are simultaneously modeled by the WBIBD. The parameters are estimated by trimmed L-moments, and the detector is constructed by combining them with the locally most powerful (LMP) test.

Watermarking technology based on deep learning utilizes the feature learning capability of neural networks to adaptively realize watermark embedding and anti-interference optimization. It achieves an excellent balance between robustness and imperceptibility and has strong generalization ability, enabling it to effectively cope with complex attacks and deep forgery scenarios. However, it relies heavily on large-scale training data, accompanied by high computational overhead and resource consumption, as well as poor model interpretability. Wang et al. [13] proposed an active deep forgery detection framework combining a Quaternion Polar Harmonic Fourier Moment (QPHFM) watermark and Dual-Task Mutual Learning (DTML). The watermark was embedded in the magnitude of QPHFMs, and the knowledge distillation of watermark extraction and deep forgery detection was realized through DTML, which performed well on the benchmark dataset. Singh et al. [14] proposed a digital image watermarking algorithm based on a CNN and autoencoder, which integrates the carrier and watermark features into the embedded network to generate watermarked images. The extraction network introduces a denoising autoencoder to enhance the anti-attack ability, which can resist various common attacks, but the embedding capacity needs to be improved. Ahmed et al. [15] proposed a deep learning model copyright protection watermarking scheme, which uses a Gold sequence and chaotic sequence to generate dual watermarks, embeds model weights through RDWT-SVD, and enhances security by chaos scrambling the weights. This method does not modify the loss function and has almost no impact on the accuracy of the model, effectively resisting pruning and fine-tuning attacks. Mahapatra [16] proposed a dual-network watermarking method based on a convolutional autoencoder, which achieves embedding through CNN feature extraction and concatenation and achieves watermark extraction through transposed convolution. It has high robustness, high concealment, and low computational overhead.

The existing digital watermark embedding methods generally perform embedding and statistical modeling based on transform-domain coefficients. However, the watermark detection performance is highly dependent on the accuracy of the statistical model. The statistical models constructed by most current watermarking algorithms have limited representation ability and cannot effectively capture the complex dependencies between transform coefficients, which restricts further improvement in watermark detection performance. In order to effectively improve the statistical representation and detection performance of the watermarking algorithm, this paper constructs an image watermarking scheme based on a new statistical model. This method integrates the magnitude coefficient characteristics of NSST-FAPHFMs and the W-PC model to realize watermark embedding and detection.

The main contributions are as follows:

1. We propose a robust invariant NSST-FAPHFM magnitude coefficient feature for watermark embedding and detection.
2. We introduce the W-PC distribution to statistically model the magnitude coefficients of NSST-FAPHFM, which can accurately describe the edge distribution characteristics and strong correlation structure of the coefficients and effectively improve the modeling accuracy and detection performance.
3. We design a blind watermarking detector based on the W-PC model and LOD with the Neyman–Pearson decision rule. The experimental results fully verify the effectiveness and superiority of the proposed detector.

## 2. Statistical Modeling of Local NSST-FAPHFM Magnitude

Frequency-domain transforms have the characteristics of multi-scale and multi-directional image decomposition, which makes the frequency-domain coefficients more robust than spatial pixels. However, in the face of attack, the anti-interference ability of frequency-domain coefficients is obviously insufficient. Orthogonal moments have good robustness and numerical stability and have been widely recognized and applied in the field of image processing. Based on this, this paper proposes combining Fast and Accurate Polar Harmonic Fourier Moments (FAPHFM) and Non-Subsampled Shearlet Transform (NSST) coefficients to construct the magnitude coefficient characteristics of NSST-FAPHFM with excellent robustness.

As a classical non-subsampled multi-scale geometric analysis method, the NSST exhibits excellent properties, including multi-directional decomposition, translation invariance, and alias-free decomposition. It can accurately capture the geometric structures of images and provides superior sparse representation capability. Therefore, it has become a core analysis tool in image denoising, fusion, segmentation, and other related fields [17]. As a new type of continuous orthogonal moment, FAPHFM effectively improve the image reconstruction accuracy of high-order moments and significantly reduce the computational time complexity. They demonstrate better comprehensive performance in various image processing tasks based on moment analysis [18]. In this paper, the NSST is employed to perform multi-scale and multi-directional decomposition on the image, and the high-energy subbands are selected to compute FAPHFM, which helps obtain robust and stable magnitude features.

The construction process is divided into three steps: first, the original image is decomposed using the two-level NSST (with parameter bits [2, 2]), yielding one low-frequency subband, four directional high-frequency subbands at the first scale, and four directional high-frequency subbands at the second scale. Then, the high-frequency subband with the highest energy is divided into non-overlapping  $8 \times 8$  subblocks. Finally, third-order FAPHFM are computed on each frequency-domain block to generate the local magnitude coefficients of NSST-FAPHFM.

To accurately capture edge distributions and strong dependencies, we employ the Weibull–Power Cauchy (W-PC) distribution [19] to model the local NSST-FAPHFM magnitude coefficients. Let  $g(x)$  denote the probability density function (PDF) of the W-PC distribution.

$$\begin{aligned}
 g(x) = & 2c\pi^{-1} \left(\frac{\alpha}{\sigma}\right) \left(\frac{x}{\sigma}\right)^{\alpha-1} \left[1 - 2\pi^{-1} \tan^{-1} \left(\frac{x}{\sigma}\right)^{\alpha}\right]^{-1} \\
 & \times \left\{ -\log \left[1 - 2\pi^{-1} \tan^{-1} \left(\frac{x}{\sigma}\right)^{\alpha}\right] \right\}^{c-1} \left[1 + \left(\frac{x}{\sigma}\right)^{2\alpha}\right]^{-1} \\
 & \times \exp \left\{ - \left[ -\log \left(1 - 2\pi^{-1} \tan^{-1} \left(\frac{x}{\sigma}\right)^{\alpha}\right) \right]^c \right\}, \\
 & x > 0, c, \alpha, \sigma > 0
 \end{aligned} \tag{1}$$

The random variable  $x$  obeys the W-PC distribution and is recorded as  $X \sim \text{W-PC}(c, \alpha, \sigma)$ . The maximum likelihood estimation of the three parameters in this distribution has a time complexity of  $O(N)$ , where  $n$  is the number of coefficients, and the iterative optimization converges rapidly.

To illustrate the fitting performance between the W-PC distribution and the coefficients, the modeling results of NSST-FAPHFM magnitude coefficients using different distribution models are presented in Figure 1. The test images include Airplane, Boat, Mandrill, and Peppers. It can be observed from the figure that the fitting effect of the W-PC model is significantly better than those of other comparative distribution models. Therefore, it can be concluded that the W-PC model provides an excellent fitting performance for the NSST-FAPHFM magnitude coefficients.

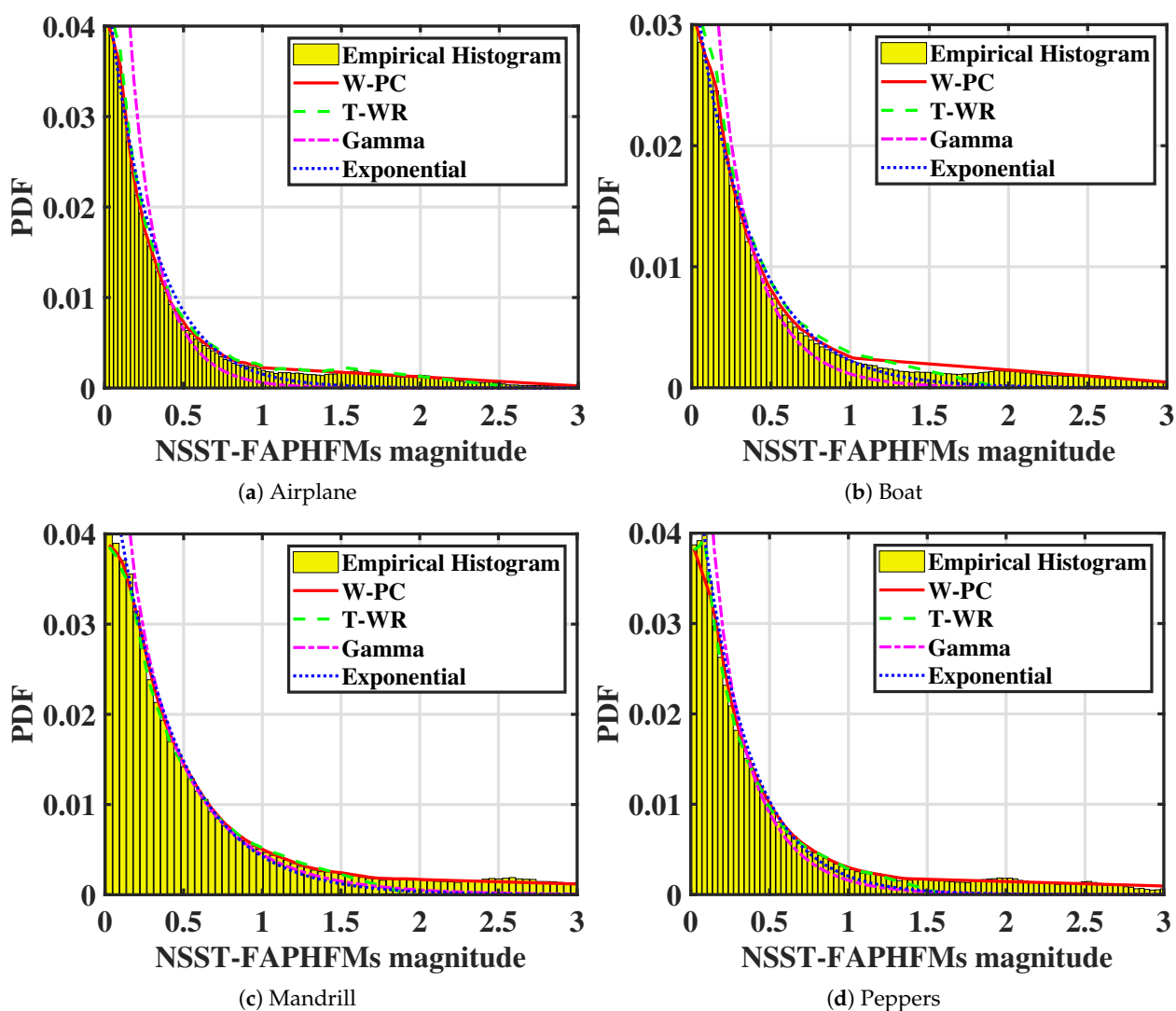


Figure 1. NSST-FAPHFM magnitude coefficient distribution histograms.

### 3. Watermark Embedding

In Section 2, we have obtained the NSST-FAPHFM magnitude coefficient matrix  $B_K$ . Next, the multiplicative embedding method modifies the NSST-FAPHFM magnitude coefficients, and watermark information  $w = \{w_k \in \{\pm 1\}, 1 \leq l \leq L\}$  is inserted into  $B_K$ .

$$y_i = \begin{cases} x_i \cdot (1 + \lambda) & \text{if } w_k = 1 \\ x_i \cdot (1 - \lambda) & \text{if } w_k = -1 \end{cases}, x_i \in B_k. \quad (2)$$

where  $x_i$  is the magnitude coefficient of the host NSST-FAPHFMs, and  $y_i$  is the magnitude coefficient of the watermark NSST-FAPHFMs.  $\lambda$  denotes the watermark embedding strength. In this paper,  $\lambda$  is associated with the watermarking-to-document ratio (WDR), which is defined as follows:

$$\lambda = \sqrt{10^{\frac{WDR}{10}} \times \sigma^2} \quad (3)$$

where  $\sigma^2$  is the variance of the magnitude coefficient matrix  $B_K$  of the selected NSST-FAPHFMs. Finally, the watermarked image is obtained by inverse FAPHFMs and the inverse NSST. The embedding process is shown in Algorithm 1's pseudocode. To verify the superiority of the NSST transformation used in this paper and the optimality of the hyperparameter selection scheme, relevant comparative verification experiments are included in the Appendix A.

---

#### Algorithm 1 Watermark embedding process

---

**Input:**  $I$ : Original grayscale image;  $w$ : binary watermark bit.

**Output:**  $I_w$ : Watermarked image.

- 1: Perform two-level NSST decomposition on the original image  $I$ .
- 2: Select the subband with the largest energy (max-energy subband).
- 3: Divide the largest energy subband into  $8 \times 8$  non-overlapping blocks.
- 4: Select the high-entropy blocks, calculate the third-order FAPHFMs for each high-entropy block, and obtain the local NSST-FAPHFM magnitude coefficients.
- 5: Use Formula (2) to embed the watermark into the NSST-FAPHFM magnitude coefficient matrix  $B_K$ .
- 6: Perform inverse FAPHFM transformation and inverse NSST reconstruction.

**return**  $I_w$

---

## 4. Digital Watermark Detection

We developed a blind watermark detection algorithm based on the W-PC distribution and Local Optimal Detector (LOD) under the Neyman–Pearson criterion. This algorithm can detect the watermark without accessing the original host image or prior watermark information. Algorithm 2 presents the pseudocode of the proposed blind watermark detection method.

---

#### Algorithm 2 Watermark detection process

---

**Input:**  $I_w$ : Watermarked image.

**Output:**  $H_1$  or  $H_0$ , where  $H_1$  indicates that the watermark exists, and  $H_0$  indicates no watermark.

- 1: Perform two-level NSST decomposition on watermarked image  $I_w$ .
  - 2: Select the subband with the largest energy.
  - 3: Divide the largest energy subband into  $8 \times 8$  non-overlapping blocks.
  - 4: Select blocks used for watermark embedding; compute third-order FAPHFMs to obtain local NSST-FAPHFM magnitude coefficients.
  - 5: Select coefficient samples where the watermark is embedded.
  - 6: Estimate the W-PC model parameters via MLE using Equation (1).
  - 7: Use Formula (9) to determine whether the watermark exists.
  - 8: **if**  $l_{LO} > \tau$  **then**
  - 9:      $H_1$
  - 10: **else if**  $l_{LO} < \tau$  **then**
  - 11:      $H_0$
  - 12: **end if**
- return**  $H_1$  or  $H_0$
-

#### 4.1. Local Optimal Watermark Detector Based on W-PC

The detection process can be represented by a binary hypothesis test.  $H_1$  indicates the presence of a watermark.  $H_0$  indicates that there is no watermark. The binary hypothesis test for multiplicative embedding is defined as follows:

$$\begin{aligned} H_0 : y_i &= x_i \\ H_1 : y_i &= x_i(1 + \lambda w_i) \end{aligned} \quad (4)$$

where  $x_i$  and  $y_i$  are the NSST-FAPHFM magnitude coefficients of the original and watermarked images respectively, and  $w_i$  is a watermark bit. The watermark detector based on the Neyman–Pearson (N-P) criterion [20] is expressed as:

$$\Lambda(Y) = \frac{f_Y(\mathbf{y} | H_1)}{f_Y(\mathbf{y} | H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \quad (5)$$

where  $\Lambda(Y)$  is the likelihood ratio and  $\tau$  is the decision threshold, which is obtained by the N-P criterion. Assuming that the magnitude coefficient of the watermarked image obeys W-PC, the decision probability is

$$\begin{aligned} f_Y(\mathbf{y} | H_0) &= f_X(y_{1i}, y_{2i}) \\ f_Y(\mathbf{y} | H_1) &= g'(y_{2i})f_X(y_{1i}, g(y_{2i})) \end{aligned} \quad (6)$$

Here,  $\mathbf{y} = \{y_{ji}\}_{i=1}^N$ ,  $j = 1, 2$ , denotes the scale;  $y_{1i}$  and  $y_{2i}$  are respectively the set of local NSST-FAPHFM magnitude coefficients on each scale;  $g(y)$  is the inverse function of  $g(x) = x(1 + \lambda w)$ ; and  $g'(y)$  is the derivative of  $g(y)$ .

$$\begin{aligned} l_{LOD}(Y) &= \ln[\Lambda(Y)] = \sum_{i=1}^N \ln \frac{f_Y(\mathbf{y} | H_1)}{f_Y(\mathbf{y} | H_0)} \\ &= \sum_{i=1}^N \left[ \ln g'(y_{2i}) + \ln \frac{f_X(y_{1i}, g(y_{2i}))}{f_X(y_{1i}, y_{2i})} \right] \\ &= \sum_{i=1}^N l_{LOD}(y_{1i}, y_{2i}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \end{aligned} \quad (7)$$

Here,  $\tau = \ln(\eta)$ . We use the LMP test to expand Equation (7) into a Taylor series [20] at  $\lambda = 0$ .

$$\begin{aligned} l_{LOD}(y_{1i}, y_{2i})|_{\lambda} &= l_{LOD}(y_{1i}, y_{2i})|_{\lambda=0} + \left. \frac{\partial l(y_{1i}, y_{2i})}{\partial \lambda} \right|_{\lambda=0} \cdot \lambda + o(\lambda) \\ &\cong \left( \left. \frac{\partial g(y_{2i})}{\partial y_{2i} \partial \lambda} \right|_{\lambda=0} + h_{LO}(y_{1i}, y_{2i}) \right) \cdot \lambda \end{aligned} \quad (8)$$

where  $h_{LO}(y_{1i}, y_{2i})$  represents the local optimal nonlinearity. By introducing the PDF of W-PC into it, we can get the following equation:

$$h_{LO}(x) = -\alpha + \frac{2\alpha(x/\sigma)^{2\alpha}}{1 + (x/\sigma)^{2\alpha}} + \frac{2(x/\sigma)}{\pi[1 + (x/\sigma)^2]} \cdot \frac{1}{A(x)} \left[ \alpha - \frac{c-1}{L(x)} + c(L(x))^{c-1} \right] \underset{H_0}{\overset{H_1}{\gtrless}} \tau \quad (9)$$

Here,

$$u = \frac{x}{\sigma} \quad (10)$$

Here,  $x$  is the NSST-FAPHFM magnitude coefficient under a single scale (simplified method of writing the double scale  $y_{1i}$  and  $y_{2i}$ , which is the general form of the subsequent single-sample detection).

$$A(x) = 1 - \frac{2}{\pi} \tan^{-1}(x/\sigma) \quad (11)$$

Here,  $A(x)$  is the core intermediate term of the W-PC distribution, with a value range of  $(0, 1)$ .

$$L(x) = -\ln A(x) \quad (12)$$

By inserting (9) and (8) into Equation (7), we get the final decision formula for the LLOD (Log-Local Optimal Detector) watermark detector, as shown in Equation (13).

$$l_{LOD}(\mathbf{Y}) = \sum_{i=1}^N w_i \left[ -\alpha + \frac{2\alpha u_i^{2\alpha}}{1+u_i^{2\alpha}} - \frac{2u_i}{\pi(1+u_i^2)A(u_i)} \cdot \left( \alpha - \frac{c-1}{L(u_i)} + c(L(u_i))^{c-1} \right) \right] \quad (13)$$

Here,

$$u_i = \frac{x_i}{\sigma} \quad (14)$$

$$A(u_i) = 1 - \frac{2}{\pi} \arctan(u_i) \quad (15)$$

$$L(u_i) = -\ln A(u_i) \quad (16)$$

#### 4.2. Performance Analysis of Watermark Detector

The performance of the detector is analyzed by the probability of false alarm  $P_{fa}$ , the probability of missed detection  $P_m$  and the probability of detection  $P_{det}$ . According to the N-P criterion,  $p_{fa}$  is limited to a fixed threshold. If  $P_m$  decreases to the minimum,  $P_{det}(P_{det} = 1 - P_m)$  will reach the maximum. The calculation formula is as follows [21]:

$$\begin{aligned} P_{fa} &= P(l_{LOD}(Y) > \tau \mid H_0) \\ &= P\left(\frac{l_{LOD}(Y) - \mu_0}{\sigma_0} > \frac{\tau - \mu_0}{\sigma_0}\right) \\ &= Q\left(\frac{\tau - \mu_0}{\sigma_0}\right) \end{aligned} \quad (17)$$

where  $Q(x)$  is the complementary cumulative distribution function of the standard normal distribution.  $\mu_0$  is the mean value and  $\sigma_0$  is the standard deviation under the assumption of  $H_0$ .

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-t^2/2} dt \quad (18)$$

For a given  $P_{fa}$ , the decision threshold  $\tau$  can be obtained based on the N-P criterion:

$$\tau = \mu_0 + \sigma_0 Q^{-1}(P_{fa}) \quad (19)$$

If  $Q(x) = P_{fa}$ , then  $Q^{-1}(P_{fa}) = x$ ; the relationship between  $P_{fa}$  and  $P_{det}$  is

$$\begin{aligned} P_{det} &= Q\left(\frac{\tau - \mu_1}{\sigma_1}\right) \\ &= Q\left(\frac{\sigma_0}{\sigma_1} Q^{-1}(P_{fa}) - \frac{\mu_0 - \mu_1}{\sigma_1}\right) \end{aligned} \quad (20)$$

where  $\mu_1$  is the mean value under the assumption of  $H_1$ , and  $\sigma_1$  is the variance. We can calculate the mean and variance under two assumptions, namely,  $\mu_0$ ,  $\mu_1$  and  $\sigma_0^2$ ,  $\sigma_1^2$ , as follows:

Mean value  $\mu_0$  under the  $H_0$  assumption:

$$\mu_0 = E \left[ \sum_{i=1}^N w_i \cdot h_{LO}(x_i) | H_0 \right] = 0 \quad (21)$$

Mean value  $\mu_1$  under the  $H_1$  assumption:

$$\begin{aligned} \mu_1 &= E \left[ \sum_{i=1}^N w_i \cdot h_{LO}(x_i) | H_1 \right] \\ &= \sum_{i=1}^N E[w_i \cdot h_{LO}(x_i)] \\ &\approx \lambda \cdot N \cdot E \left[ h_{LO}(x) \cdot \frac{\partial}{\partial \lambda} g(x, \lambda) \right]_{\lambda=0} \\ &\approx \lambda \cdot N \cdot E[h_{LO}(x) \cdot x] \end{aligned} \quad (22)$$

Variance  $\sigma_0^2$  under the  $H_0$  assumption:

$$\begin{aligned} \sigma_0^2 &= Var \left[ \sum_{i=1}^N w_i \cdot h_{LO}(x_i) | H_0 \right] \\ &= \sum_{i=1}^N Var[w_i \cdot h_{LO}(x_i)] \\ &= \sum_{i=1}^N E[(h_{LO}(x_i))^2] \\ &= N \cdot E[(h_{LO}(x))^2] \end{aligned} \quad (23)$$

Variance  $\sigma_1^2$  under the  $H_1$  assumption:

$$\begin{aligned} \sigma_1^2 &= Var[T(Y) | H_1] \\ &\approx Var[T(Y) | H_0] \\ &= \sigma_0^2 = N \cdot E[(h_{LO}(x))^2] \end{aligned} \quad (24)$$

## 5. Experimental Results

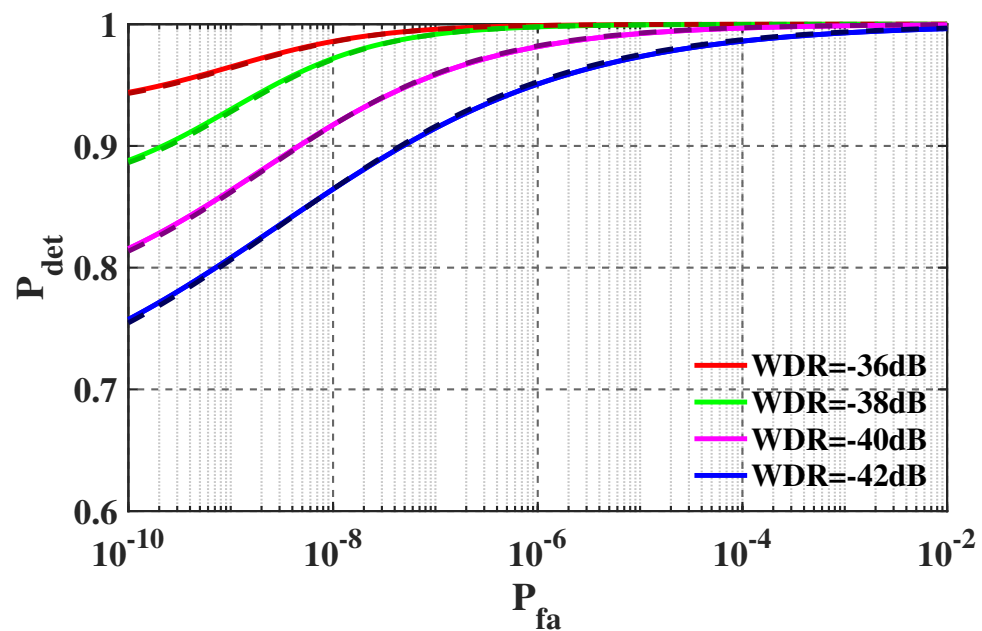
Through a series of experiments, the watermarking-to-document ratio (WDR), peak signal-to-noise ratio (PSNR), ROC curve, and strength of the watermark can be controlled by setting the WDR value. The ROC curve describes the relationship between  $P_{fa}$  and  $P_{det}$ . The watermark adopted in the experiments is a pseudorandom sequence composed of elements +1 and −1, which ensures the robustness of the watermarking system. The experiment was carried out using MATLAB R2022a. The gray test image is from a standard image database [22].

### 5.1. Performance Evaluation of Watermark Detector

#### 5.1.1. Accuracy Evaluation

In this section, the theoretical ROC curve is compared with the experimental ROC curve to validate the accuracy of the theoretical model. Specifically, 96 test images and 100 sets of watermarks were adopted for the experiments;  $P_{fa}$  was defined within the range of  $[10^{-10}, 10^{-2}]$ , and WDR varied from −38 dB to −30 dB. As shown in Figure 2, the theoretical ROC curve exhibits a high degree of consistency with its experimental counterpart, which confirms the accuracy of the theoretical model. This further demonstrates

that the theoretical ROC curve can serve as a reliable tool for evaluating the performance of watermarking systems.



**Figure 2.** The experimental (solid line) and theoretical (dashed line) ROC curves of different WDR cases.

#### 5.1.2. Imperceptibility Analysis

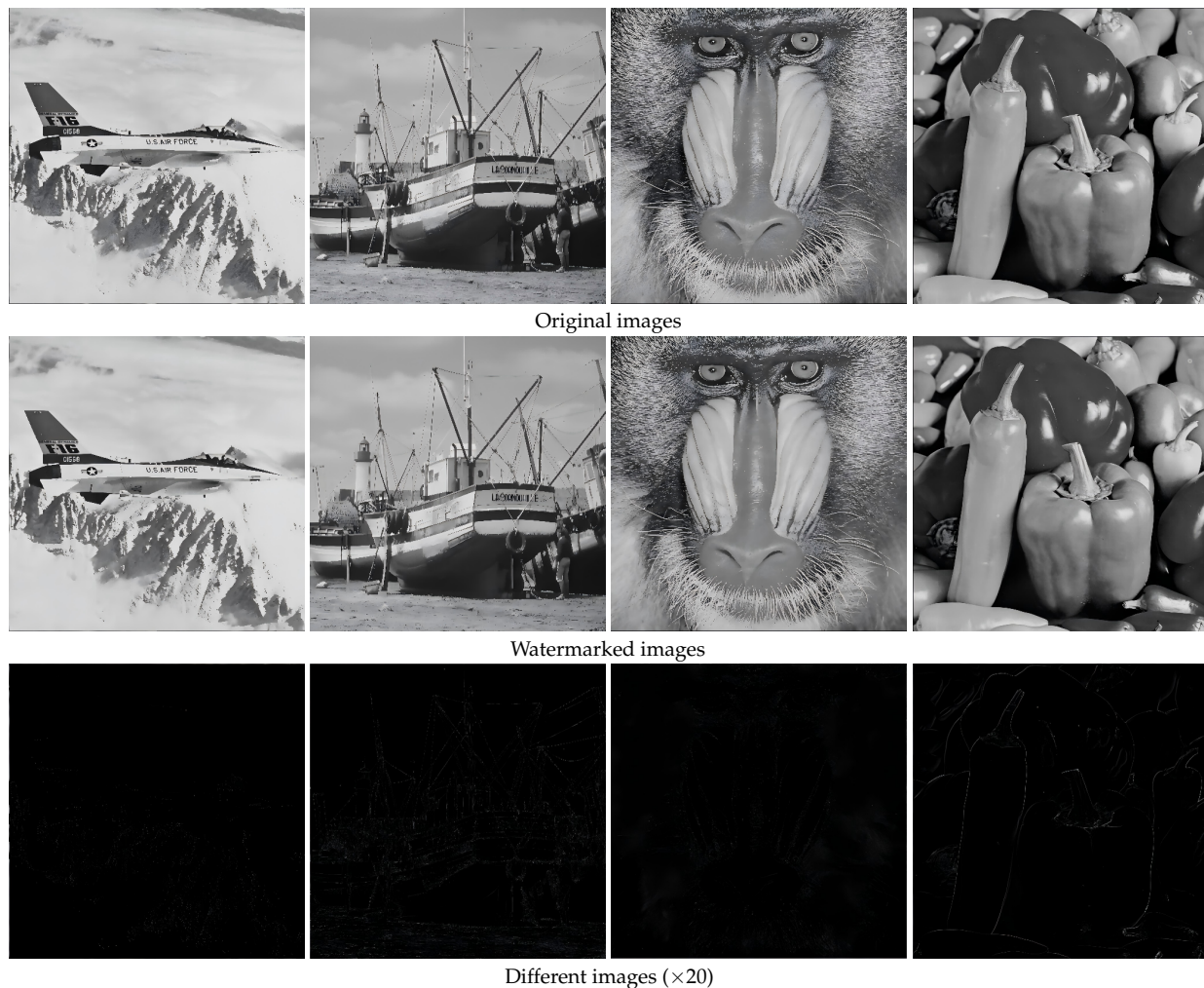
Imperceptibility is one of the core indicators for evaluating a digital watermarking algorithm. The PSNR is commonly used to quantitatively characterize the degree of image distortion after watermark embedding. In general, when the PSNR value exceeds 38 dB, the human eye cannot perceive the difference between images, indicating that the corresponding algorithm achieves satisfactory imperceptibility. In this study, experiments were carried out on several standard test images under the conditions of a watermark length of 1000 bits and watermark strength of  $\text{WDR} = -40$  dB. Figure 3 presents a comparison between the original images and watermarked images of Airplane, Boat, Mandrill and Peppers. Their PSNRs were 47.6349 dB, 48.2461 dB, 44.2462 dB, and 48.7182 dB. The experimental results show that there is no significant difference in appearance between the watermarked image and the original image, and the PSNRs of all test images are higher than 38 dB, which verifies the excellent imperceptibility of the algorithm.

#### 5.1.3. Watermark Capacity

As the watermark capacity increases, the visual quality of the watermarked image tends to degrade. Table 1 presents the average PSNR values of the proposed watermarking algorithm under different watermark capacities. In the experiments, watermark sequences with varying lengths were embedded into 24 standard  $512 \times 512$  grayscale test images, with  $\text{WDR} = -40$  dB. The results demonstrate that the proposed scheme maintains a high level of imperceptibility even at high watermark capacities.

**Table 1.** Watermark capacity and average PSNR value.

| Watermark capacity | 500 bits   | 1000 bits  | 2000 bits  |
|--------------------|------------|------------|------------|
| Average PSNR       | 49.3763 dB | 47.9542 dB | 46.9411 dB |



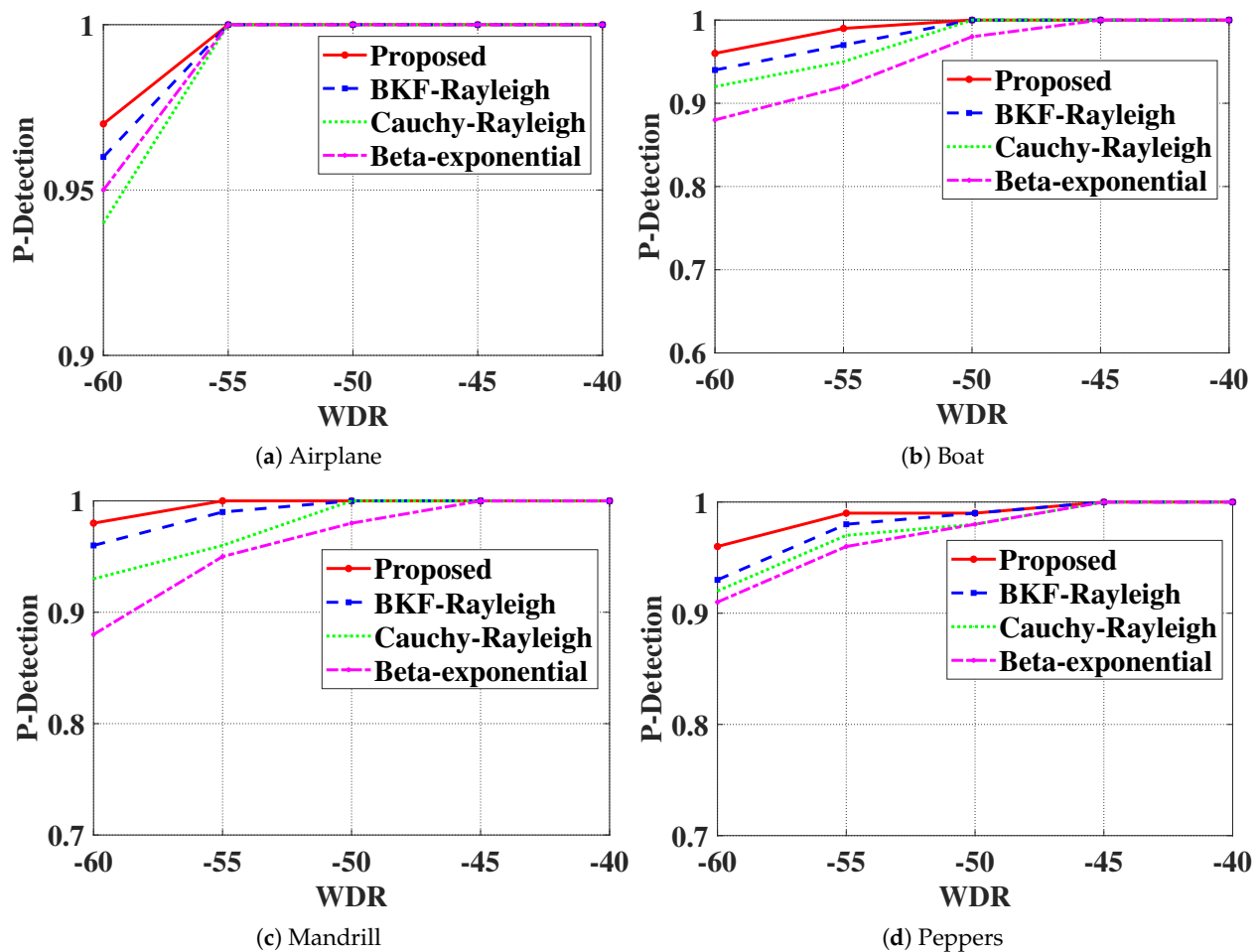
**Figure 3.** Imperceptibility analysis.

### 5.2. Comparison with the Latest Methods

To further verify the performance of the proposed watermark detection algorithm, a comparative analysis was conducted with several existing methods presented in references [13,23–26]. For fairness, identical test images and experimental parameters were adopted in all comparisons.

#### 5.2.1. Detection Probability of Different Watermark Strengths

This paper compares and analyzes the performance of multiple detectors under different watermark embedding strengths. In the experiment,  $P_{fa} = 10^{-1}$  was set in advance, and on this basis,  $P_{det}$  was calculated under different embedding strengths; the experimental results are visualized in the form of a line graph. The comparison algorithms selected in this study include the Cauchy–Rayleigh model [23], beta-exponential model [24] and BKF–Rayleigh model [25]. The test dataset contains four standard test images: Airplane, Boat, Mandrill and Peppers. As can be seen in Figure 4, the detection probability of each detector increases with the increase in watermark embedding strength. Further comparative analysis shows that compared with the existing contrast detectors, the proposed detector has better detection performance in the full range of watermark embedding strength, which verifies the effectiveness and robustness of the algorithm.



**Figure 4.**  $p_{det}$  under different watermark strengths.

### 5.2.2. AUROC Under Various Attacks

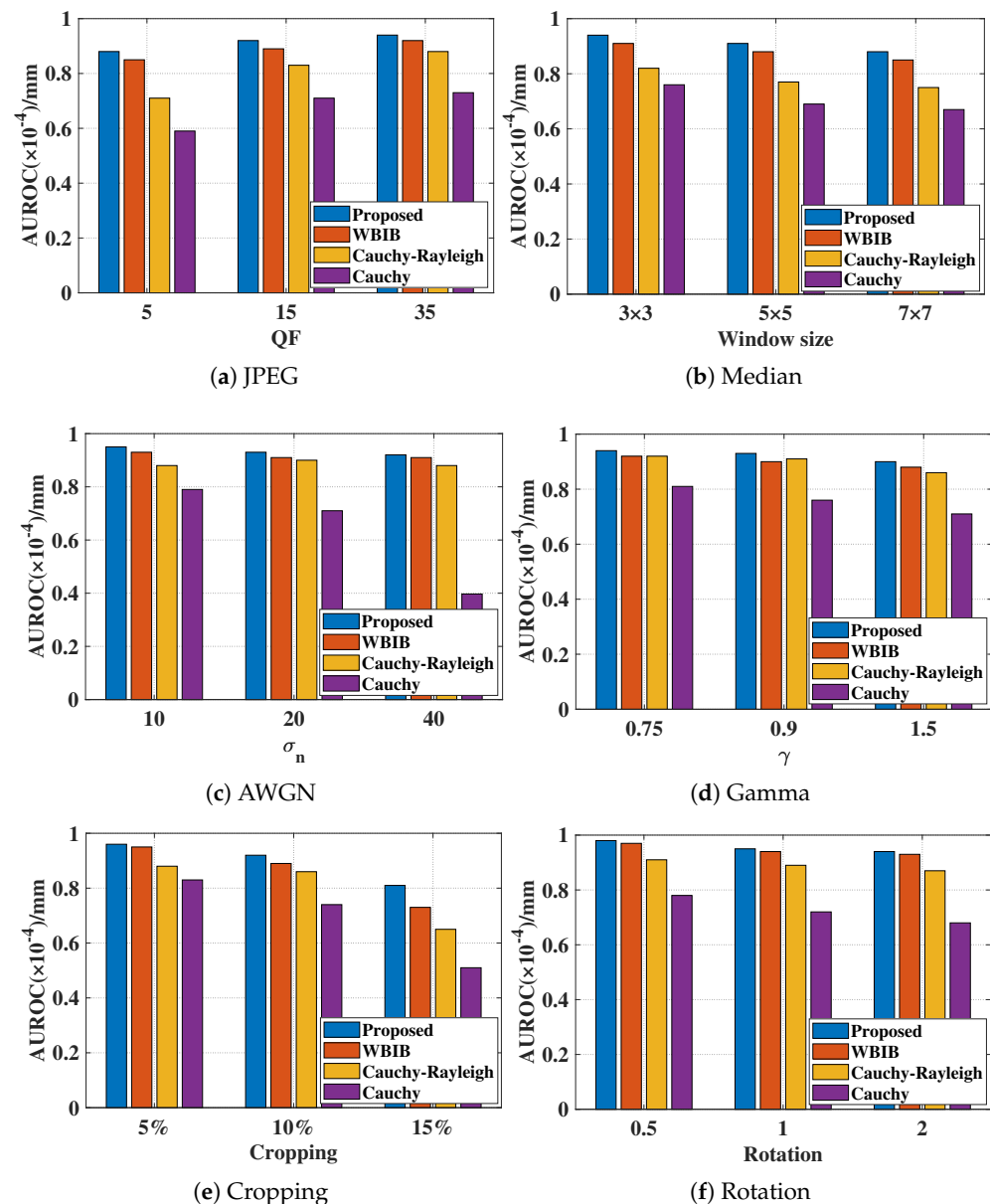
Secondly, the robustness of different detectors was tested and compared. The area under the receiver operating characteristic curve (AUROC) is employed as the quantitative metric for algorithm performance evaluation, with its value ranging from 0 to 1. Under a fixed WDR condition, a higher AUROC value indicates better detection performance of the corresponding detector. The methods proposed in references [13,23,26] were selected as the comparison benchmarks in this work. The experiments were conducted on 24 grayscale test images with a resolution of  $256 \times 256$  pixels. The length of the watermark sequence was set to 1000 bits,  $P_{fa}$  was defined in the range  $[0, 10^{-4}]$ , and the WDR =  $-42$  dB. Table 2 shows the average AUROC results of five detectors without attack. The results demonstrate that the proposed detector achieves a significantly higher AUROC value than the other methods.

**Table 2.** The AUROC values under no attack.

| Reference | [26]   | [23]   | [13]   | Proposed |
|-----------|--------|--------|--------|----------|
| AUROC     | 0.9953 | 0.9967 | 0.9984 | 0.9991   |

### 5.2.3. Average AUROC Value Under Different Attacks

Figure 5 presents the average AUROC under different attacks in the form of a bar graph. The employed attacks include JPEG compression, median filtering, AWGN, Gamma correction, cropping, and rotation.



**Figure 5.** Average AUROC values of different detectors under different attacks.

It can be observed that the blue bar is higher than the bars of other colors, indicating that the proposed detection method exhibits stronger robustness than the other compared methods. The experimental results demonstrate that, under the same embedding load, the proposed algorithm achieves superior robustness compared with the other algorithms and obtains satisfactory experimental performance. This validates the effectiveness of employing NSST-FAPHFs and their magnitude characteristics to enhance the robustness of the algorithm.

#### 5.2.4. Comparison with Other Watermarking Algorithms

In addition, this study conducted comparative experiments between the proposed algorithm and existing deep learning and neural network watermarking algorithms. All experiments were performed in a unified environment with the same dataset type, consistent data splitting and the same preprocessing rules to ensure experimental fairness. PSNR, Structural Similarity Index (SSIM), Bit Error Rate (BER), and AUROC are adopted as core evaluation metrics. Reference [13] adopts DTML and QPHFM watermarking for active Deepfake detection, achieving both robustness and high detection accuracy.

Reference [14] uses dual watermarking and RDWT-SVD for deep learning model copyright protection, which can resist pruning and fine-tuning attacks. Reference [15] adopts an RDWT-SVD-based dual-watermark copyright protection method for deep learning models combined with weight shuffling, which is robust against pruning and fine-tuning attacks. Reference [16] uses a convolutional autoencoder-based dual-network watermarking method with CNN feature embedding and transposed convolution extraction, featuring strong robustness and high imperceptibility.

As shown in Table 3, to ensure a fair and unbiased comparison, all experiments in this work were conducted under strictly identical conditions. All color test images from the CIFAR-10 dataset were converted to grayscale while retaining their native  $32 \times 32$  pixel resolution, consistent with the standard dataset configuration. Evaluated using the same metrics, the proposed scheme achieves higher PSNR and SSIM values than all competing methods. These results demonstrate that our approach preserves superior image quality after watermark embedding, exhibiting clear performance advantages over existing solutions.

**Table 3.** Comparison of PSNRs and SSIMs between existing schemes and our method.

| Method                  | Dataset Cover Image | Watermarked Image | PSNR (dB) | SSIM   |
|-------------------------|---------------------|-------------------|-----------|--------|
| ConvAE + CNN [16]       | CIFAR10             | Random            | 39.3415   | 0.9725 |
| CNN Denoising AE [14]   | CIFAR10             | Random            | 42.9122   | 0.9931 |
| Deep learning (DL) [15] | CIFAR10             | Random            | 45.3674   | 0.9962 |
| Proposed                | CIFAR10             | Random            | 45.6242   | 0.9980 |

For a fair comparison of robustness under various attacks, we included three representative deep learning-based watermarking methods [14–16]. Table 4 compares the AUROC performance of different watermarking algorithms under various common image processing attacks. The results show that the proposed method obtains higher AUROC values in most attack cases and possesses better robustness.

**Table 4.** AUROC comparison of different watermarking algorithms under various attacks.

| Attack Type                         | DTML [13] | ConvAE + CNN [16] | CNN Denoising AE [14] | Proposed |
|-------------------------------------|-----------|-------------------|-----------------------|----------|
| JPEG Compression (QF = 15)          | 0.9955    | 0.9935            | 0.9949                | 0.9963   |
| Median Filtering ( $5 \times 5$ )   | 0.9920    | 0.9940            | 0.9962                | 0.9951   |
| Gaussian Filtering ( $5 \times 5$ ) | 0.9216    | 0.9371            | 0.9366                | 0.9499   |
| Salt-and-Pepper Noise (P = 20%)     | 0.9928    | 0.9934            | 0.9950                | 0.9947   |
| Rotation ( $\theta = 2^\circ$ )     | 0.9158    | 0.9145            | 0.9268                | 0.9355   |
| Scaling (SF = 1.2)                  | 0.8642    | 0.8739            | 0.8858                | 0.8731   |
| Cropping (10%)                      | 0.9029    | 0.9012            | 0.9184                | 0.9254   |

## 6. Conclusions

This paper proposes a robust blind watermarking detection algorithm based on NSST-FAPHFMs magnitude coefficients and the W-PC distribution. Firstly, the NSST is used to decompose the carrier image in a multi-scale and multi-directional manner. By combining the NSST with FAPHFMs, a robust NSST-FAPHFMs magnitude coefficient feature is constructed to provide a reliable feature carrier for watermark detection. Secondly, the W-PC distribution is introduced to accurately model the non-Gaussian statistical characteristics of the magnitude coefficients of NSST-FAPHFMs, which can effectively capture the strong correlation between image edge features and coefficients and improve the representation ability of the model for the inherent characteristics of the image. Finally, a robust blind watermarking detector is designed based on the W-PC statistical model and the Local Optimal Detector with the Neyman–Pearson criterion. Experimental results show that the

proposed algorithm has excellent watermark imperceptibility and shows strong robustness under various common image processing attacks. Its watermark detection accuracy is significantly better than that of existing mainstream blind watermark detection methods, and it can effectively balance the inherent conflict between imperceptibility, robustness and watermark capacity. Subsequent research will further expand the application scenarios of the algorithm in the field of color image and video watermarking and further optimize and improve the detection performance and robustness of the algorithm combined with deep learning technology.

**Author Contributions:** Conceptualization, S.Y.; methodology, S.Y.; software, S.Y. and H.Z.; validation, S.Y. and Y.G.; formal analysis, Y.G. and H.Z.; investigation, Y.G. and H.Z.; data curation, S.Y. and Y.G.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y. and Y.G.; visualization, H.Z.; supervision, S.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Scientific Research Fund for Doctor of Dalian Polytechnic University (7261041040).

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

## Appendix A

To further verify the superiority of the adopted NSST framework, we conducted comprehensive comparative experiments on the NSST, conventional DWT and contourlet transform under identical experimental setups, watermark embedding configurations and test datasets. Table A1 compares the performance of different transform-domain watermarking schemes.

**Table A1.** Performance comparison of different transform-domain watermarking schemes.

| Transform                            | PSNR (dB) | SSIM   | BER    |
|--------------------------------------|-----------|--------|--------|
| 2-level NSST + 3rd-order FAPHF       | 47.25     | 0.9991 | 0.0063 |
| 2-level Contourlet + 3rd-order FAPHF | 44.69     | 0.9824 | 0.0106 |
| 2-level DWT + 3rd-order FAPHF        | 43.43     | 0.9691 | 0.0185 |

As the tabulated results reveal, the scheme combining two-level NSST decomposition and 3rd-order FAPHF yields optimal overall performance: it obtains superior imperceptibility and stronger robustness.

To validate the optimality of our hyperparameter selection, we performed ablation experiments to analyze the influence of NSST decomposition levels and FAPHF polynomial orders. Table A2 shows a performance comparison for different NSST levels and FAPHF orders.

**Table A2.** Performance comparison for different NSST decomposition levels and FAPHF orders.

| Combination                    | PSNR (dB) | SSIM   | BER    | Time (s) |
|--------------------------------|-----------|--------|--------|----------|
| 2-level NSST + 3rd-order FAPHF | 47.25     | 0.9991 | 0.0063 | 4.28     |
| 3-level NSST + 3rd-order FAPHF | 46.78     | 0.9824 | 0.0106 | 6.74     |
| 2-level NSST + 4th-order FAPHF | 44.89     | 0.9691 | 0.0185 | 6.15     |
| 3-level NSST + 4th-order FAPHF | 44.62     | 0.9685 | 0.0241 | 9.28     |

The results in Table A2 indicate that the combination of two-level NSST decomposition and third-order FAPHFM provides the best overall performance: it has excellent imperceptibility and robustness while also having the lowest computational overhead.

## References

- Awasthi, D.; Tiwari, A.; Khare, P.; Srivastava, V.K. A comprehensive review on optimization-based image watermarking techniques for copyright protection. *Expert Syst. Appl.* **2024**, *242*, 122830. [\[CrossRef\]](#)
- Kumar, C.; Kumar, D. Recent security challenges and robust techniques in colour image watermarking. *Int. J. Comput. Vis. Robot.* **2025**, *15*, 312–329. [\[CrossRef\]](#)
- Melman, A.; Evsutin, O. Methods for countering attacks on image watermarking schemes: Overview. *J. Vis. Commun. Image Represent.* **2024**, *99*, 104073. [\[CrossRef\]](#)
- Panigrahi, R.; Padhy, N. An effective steganographic technique for hiding the image data using the LSB technique. *Cyber Secur. Appl.* **2025**, *3*, 100069. [\[CrossRef\]](#)
- Li, Z.; Yi, Q.P.; Ji, Z.C.; Lu, Y.J.; Li, Y.Q.; Xiao, K.Y.; Liang, H.L. Bimarker: Enhancing text watermark detection for large language models with bipolar watermarks. *arXiv* **2025**, arXiv:2501.12174.
- Amini, M.; Sadreazami, H.; Ahmad, M.O.; Swamy, M.N.S. A channel-dependent statistical watermark detector for color images. *IEEE Trans. Multimed.* **2018**, *21*, 65–73. [\[CrossRef\]](#)
- Wang, X.Y.; Shen, Y.X.; Song, L.; Niu, P.P. WMHMT-IWD: Weibull mixtures-HMT based image watermark detector. *Circuits Syst. Signal Process.* **2024**, *43*, 5624–5660. [\[CrossRef\]](#)
- Iqbal, N. Multi-Channel Statistical Framework for Robust and Reliable Watermark Detection in Color Image Processing. *Int. J. Innov. Sci. Res. Technol.* **2025**, *10*, 743–750.
- Stanković, D.; Draganić, A.; Ioana, C.; Orović, I. An Image Watermarking based on Rayleigh distribution detector. In Proceedings of the 2024 International Symposium ELMAR, Zadar, Croatia, 16–18 September 2024; pp. 31–35.
- Zhang, J.; Du, J.L.; Xi, X.; Yang, Z.H. Chebyshev chaotic mapping and DWT-SVD-based dual watermarking scheme for copyright and integrity authentication of remote sensing images. *Symmetry* **2024**, *16*, 969. [\[CrossRef\]](#)
- Yuan, Y.; Zhao, W.; Li, Z.; Liu, W. The Color Image Watermarking Algorithm Based on Quantum Discrete Wavelet Transform and Chaotic Mapping. *Symmetry* **2025**, *18*, 33. [\[CrossRef\]](#)
- Niu, P.P.; Wang, D.W.; Xu, X.H.; Zhang, X.H.; Wang, X.Y. Robust Domain Image Watermarking Using WBIB Distribution. *Circuits Syst. Signal Process.* **2025**, *44*, 8534–8548. [\[CrossRef\]](#)
- Wang, C.P.; Shi, C.Y.; Wang, S.M.; Xia, Z.Q.; Ma, B. Dual-Task Mutual Learning with QPHFM Watermarking for Deepfake Detection. *IEEE Signal Process. Lett.* **2024**, *31*, 2740–2744. [\[CrossRef\]](#)
- Singh, H.K.; Singh, A.K. Digital image watermarking using deep learning. *Multimed. Tools Appl.* **2024**, *83*, 2979–2994.
- Ahmed, M.I.; Singh, K.N.; Singh, H.K.; Singh, A.K.; Agrawal, A.K. Ownership protection of deep learning models using watermarking. *Comput. Electr. Eng.* **2025**, *125*, 110481. [\[CrossRef\]](#)
- Mahapatra, D.; Amrit, P.; Singh, O.P.; Singh, A.K.; Agrawal, A.K. Autoencoder-convolutional neural network-based embedding and extraction model for image watermarking. *J. Electron. Imaging* **2023**, *32*, 021604.
- Jose, J.; Gautam, N.; Tiwari, M.; Tiwari, T.; Suresh, A.; Sundararaj, V.; Rejeesh, M.R. An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion. *Biomed. Signal Process. Control* **2021**, *66*, 102480. [\[CrossRef\]](#)
- Yang, S.Y.; Deng, A.S. Fast and accurate computation of polar harmonic Fourier moments for image description. *J. Opt. Soc. Am. A* **2023**, *40*, 1714–1723. [\[CrossRef\]](#)
- Tahir, M.H.; Zubair, M.; Cordeiro, G.M.; Alzaatreh, A.; Mansoor, M. The Weibull-Power Cauchy distribution: Model, properties and applications. *Hacet. J. Math. Stat.* **2017**, *46*, 767–789.
- Amini, M.; Ahmad, M.O.; Swamy, M.N.S. A new locally optimum watermark detection using vector-based hidden Markov model in wavelet domain. *Signal Process.* **2017**, *137*, 213–222. [\[CrossRef\]](#)
- Deng, C.Z.; Zhu, H.S.; Wang, S.Q. Curvelet domain watermark detection using alpha-stable models. In *Proceedings of the 2009 Fifth International Conference on Information Assurance and Security*; IEEE Computer Society: Washington, DC, USA, 2009; Volume 1, pp. 313–316.
- CVG-USG-Image Database. 2014. Available online: <https://ccia.ugr.es/cvg/dbimages/index.php> (accessed on 8 July 2026).
- Niu, P.P.; Wang, L.; Tian, J.L.; Zhang, S.Y.; Wang, X.Y. A statistical color image watermarking scheme using local QPCET and Cauchy-Rayleigh distribution. *Circuits Syst. Signal Process.* **2021**, *40*, 4516–4545. [\[CrossRef\]](#)
- Wang, X.Y.; Niu, P.P.; Tian, J.; Tian, J.L. A new statistical image watermark detector in RHFM domain using beta-exponential distribution. *Soft Comput.* **2022**, *26*, 9707–9727. [\[CrossRef\]](#)

25. Yang, S.Y.; Deng, A.S.; Cui, H. Statistical Image Watermark Algorithm for FAPHFMs Domain Based on BKF–Rayleigh Distribution. *Mathematics* **2023**, *11*, 4720. [[CrossRef](#)]
26. Sadreazami, H.; Ahmad, M.O.; Swamy, M.N.S. A robust multiplicative watermark detector for color images in sparse domain. *IEEE Trans. Circuits Syst. II Express Briefs* **2015**, *62*, 1159–1163. [[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.