Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz

ROBUST COLOR IMAGE WATERMARKING BASED ON DWT AND CNN

Aqeel H. Younus [1, 2]*, and Abdulhakeem O. Mohammed [3]

[1] Department of Information Technology, Technical College of Duhok, Duhok Polytechnic University, Iraq.
[2] Department of Information Technology, Technical College of Informatics - Akre, Akre University for Applied Sciences, Iraq.
[3] Department of Computer Science, College of Science, University of Zakho, Zakho, Kurdistan Region, Iraq.

*Corresponding author, E-mail: aqeel.younus@dpu.edu.krd. (Tel.: +964-7518087765)

Received: 28, Jul, 2025
Accepted: 14, Sep, 2025
Published: 09, Apr, 2026

ABSTRACT

Digital watermarking is one of the most important technologies used today for copyright protection and authentication as well as data security in the digital domain. While many existing techniques perform well under certain conditions, designing a method that is both imperceptible and robust against various attacks remains a key challenge, especially for color images. This paper proposes robust color image watermarking based on Discrete Wavelet Transform (DWT) and Convolutional Neural Networks (CNNs). The method incorporates a U-Net architecture enhanced with Squeeze-and-Excitation (SE) blocks and residual learning. The watermark is embedded in the High-Low (HL1) sub-band of the blue-difference chrominance (Cb) channel in YCbCr color space leveraging its lower perceptual sensitivity. A parallel extraction network is jointly trained using a hybrid loss function combining Mean Squared Error (MSE) and Structural Similarity Index (SSIM) to ensure visual quality and extraction reliability. The experiment conducted on the COCO2017 dataset thus shows that the proposed method can achieve a good of imperceptibility with PSNR reaching 45.35 dB and SSIM attaining 0.996. Moreover, it can demonstrate against a variety of attacks such as noise, compression, filtering, rotation, and cropping.

KEYWORDS: Watermarking, Copyright Protection, DWT, CNN, U-Net.

1. INTRODUCTION

In recent years, digital content has become highly vulnerable due to widespread editing tools and the rapid growth of online multimedia platforms, making images, audio, and video easy to copy, modify, or misuse, raising serious copyright and ownership concerns (Rai *et al.* 2023). A digital watermark is a hidden signal within an image that verifies ownership or detects tampering (Aberna & Agilandeeswari 2024; Darwish *et al.* 2024). It must remain imperceptible to the human eye while being robust enough to withstand common attacks like compression, noise, or filtering (Juarez-Sandoval *et al.* 2021; Zhou *et al.* 2021). Traditional methods such as LSB substitution or DCT embedding are simple and effective but often fail to balance invisibility and resilience against modern image processing or intentional attacks (Mohammed *et al*. 2023; Gull & Parah 2024; Boujerfaoui *et al.* 2022), motivating the development of smarter, adaptive solutions.

Consequently, recent research is increasingly adopting deep learning-based methods (Luo *et al.* 2024; Zhong *et al.* 2023), particularly CNNs, which can adaptively learn hierarchical image features and capture complex patterns (Ben Jabra & Ben Farah 2024; Hosny *et al.* 2024; Lee *et al*. 2020). When combined with DWT, which decomposes images into sub-bands (LL, LH, HL, HH) for multi-resolution frequency analysis (Lu *et al.* 2022), this hybrid approach leverages CNNs for spatial feature learning and DWT for precise frequency localization (Zhao *et al.* 2022; Boujerfaoui *et al.* 2022). Embedding watermarks in DWT sub-bands like HL1 (capturing high-frequency edges) enhances robustness while maintaining invisibility, producing watermarks that are harder to detect and more resilient to attacks (Tavakoli *et al.* 2023; Benoraira *et al.* 2015; Ye *et al*. 2023)

Despite advances in digital watermarking, achieving both imperceptibility and robustness remains challenging, especially for color images, which are more sensitive to distortions and have higher-dimensional data. Often, enhancing one property compromises the other. To address this trade-off, this paper proposes a robust color image watermarking method based on Discrete Wavelet Transform (DWT) and Convolutional Neural Networks (CNNs), using an improved U-Net architecture with SE blocks and residual connections. The watermark is embedded

Access this article online
https://doi.org/10.25271/sjuoz.2026.14.2.1740
Printed ISSN 2663-628X;
Electronic ISSN 2663-6298
Science Journal of University of Zakho
Vol. 14, No. 02, pp. 198 –208 April -2026

in the HL1 sub-band of the Cb component within the YCbCr color model, taking advantage of the human visual system's lower sensitivity to chrominance to balance invisibility and robustness.

Contributions of the suggested method are as follows:

- A blind and robust color image watermarking method is proposed that combines Discrete Wavelet Transform (DWT) with Convolutional Neural Networks (CNNs). The watermark is embedded in the HL1 sub-band of the Cb channel in the YCbCr color space to enhance imperceptibility and robustness against distortions.
- An enhanced U-Net architecture is employed, incorporating Squeeze-and-Excitation (SE) blocks to focus on important feature channels and residual connections to preserve visual details.
- The watermark embedding and extraction are handled by two U-Net-based networks trained jointly. During training, the extraction network uses a hybrid loss function that combines Mean Squared Error (MSE) and Structural Similarity Index (SSIM), while the embedding network uses MSE to ensure accurate watermark placement.

This paper is organized as follows. Section 2 presents a literature review on digital watermarking. Section 3 explains the theoretical foundation of the proposed method, including the roles of CNN and DWT. Section 4 details the watermark embedding and extraction procedures. Section 5 presents experimental results on invisibility, robustness against attacks, and comparisons with other methods. Finally, Section 6 concludes the paper and suggests potential directions for future work.

2. RELATED WORKS

This section reviews related watermarking methods grouped into three main categories, traditional transform-based, deep CNN-based, and hybrid/advanced architectures. The grouping facilitates clearer comparison and highlights the differences with the proposed method.

Traditional Methods:

Early watermarking research primarily used transform-domain techniques such as DWT, DCT, and SVD to balance invisibility and robustness. For instance, Benoraira *et al.* embedded watermark bits by modifying mid-frequency DCT coefficients within DWT sub-bands, and Abdullatif *et* al. used pseudorandom sequences in specific DWT coefficients, achieving high PSNR (38–40 dB) and strong resistance to compression, noise, and filtering without requiring the original image (Benoraira *et al.* 2015; Abdullatif *et* al. *2014*). Later works enhanced these methods by combining multiple transforms and adding optimization or encryption; for example, Mohammed *et al*. integrated DWT, DCT, and chaotic logistic encryption for RGB images, while Singh and Singh, Kang *et al*. and Zhang and Wei combined DWT, DCT, and SVD, sometimes using PSO for embedding strength, reaching PSNR above 40 dB and robustness against geometric and signal processing attacks, though with limited rotation resilience(Mohammed *et al*. 2023; Singh and Singh 2017; Kang *et al*. 2018; Zhang & Wei 2019). Other studies like Zear *et al. and* Zermi *et al*. addressed medical and dual watermarking, embedding multiple watermarks in DWT–DCT or DWT–SVD sub-bands to protect sensitive data while maintaining high PSNR and SSIM and ensuring resistance to noise, compression, and filtering (Zear *et al.* 2018; Zermi *et al*. 2021).

Deep Learning CNN-Based Methods:

Deep learning has greatly advanced watermarking by enabling models to learn embedding and extraction directly from data. Singh and Singh used a CNN-based encoder–decoder with a denoising autoencoder, improving robustness and imperceptibility (Singh & Singh 2024). Jamali *et al*. incorporated a CNN with an attack simulation layer, achieving PSNR above 40 dB and high SSIM (Jamali *et al*. 2023). Padhi *et al*. proposed a dual invisible watermarking scheme with perceptual and cryptographic hashes for strong robustness while maintaining image quality (Padhi *et al*. 2024). Rai *et al.* combined CNN with ECOA, Octave Convolution, and Pyramid Feature Extraction, balancing accuracy and simplicity with high PSNR (54.64 dB), SSIM (0.97), and NC (0.98), resilient to most attacks except histogram equalization and content-based manipulations (Rai *et al.* 2023). Lightweight CNNs have also been explored. Lee *et al.* built a blind system for grayscale images with high imperceptibility, Mahapatra *et al.* used an autoencoder–CNN hybrid with low computational cost, and Subramanian *et al.* developed a lightweight autoencoder–decoder architecture, achieving PSNR 34.55 dB with strong imperceptibility and resilience (Lee *et al.* 2020; Mahapatra *et al.* 2023; Subramanian *et al.* 2021). To handle complex distortions, Ma *et al.* combined Swin Transformers with deformable CNNs and a distortion-style-ensemble noise layer, achieving near-perfect extraction under most geometric attacks (Ma *et al.* 2024).

Hybrid Methods:

Several studies have enhanced watermarking by combining traditional transform techniques with deep learning to improve robustness, imperceptibility, and capacity. Sy *et al.* integrated DWT with CNNs for image decomposition, embedding, and extraction, including simulated attacks during training, achieving high PSNR (>39 dB), near-perfect SSIM, and strong robustness (Sy *et al.* 2020). Similarly, Lu *et al*. and Tavakoli *et al*. used DWT within CNNs with residual learning and attack simulation, reaching PSNR around 40 dB and strong resistance to compression, noise, and cropping (Lu *et al*. 2022; Tavakoli *et al*. 2023). Ahmadi *et al.* proposed ReDMark, combining FCNN with DCT layers and differentiable attack layers, achieving PSNR up to 40.24 dB, SSIM 0.987, and very low bit error rates (Ahmadi *et al.* 2020). Mahto *et al.* presented a hybrid framework embedding multiple marks in spatial and transform domains with hybrid optimization, encryption, and DnCNN, reaching PSNR 57.71 dB with high robustness (Mahto *et al.* 2022). Finally, Hsu and Hu applied QDCT with Grey Wolf Optimizer for optimal embedding, followed by DnCNN for quality restoration, maintaining high PSNR and SSIM while resisting compression, noise, and filtering (Hsu & Hu 2021).

Traditional transform-based methods are fast and easy to interpret but often require manual tuning and provide limited robustness against complex distortions. CNN-based methods are more adaptive and robust, yet they demand large datasets and high computational cost. Therefore, hybrid approaches combine both advantages. In this study, the proposed method uses DWT preprocessing and embeds the watermark in the Cb channel of the YCbCr space with robustness-oriented training, achieving better invisibility and higher PSNR/SSIM than earlier methods.

3. PRELIMINARIES

In the next subsections, a brief theoretical background on the methods employed by the proposed methodology is presented to the reader.

Convolutional Neural Networks (CNNs) in Watermarking:

Currently, Convolutional Neural Networks (CNNs) are widely used as the main deep learning architecture for image processing tasks, including digital watermarking (Ben Jabra & Ben Farah 2024), because they automatically learn hierarchical image features with high accuracy, improving watermark embedding and extraction performance (Hosny *et al.* 2024). Unlike traditional methods that depend on manual feature

engineering, CNN-based watermarking learns features directly from data, which increases robustness against common attacks such as compression, noise, and geometric transformations (Luo *et al.*2024). Moreover, CNNs can be trained using simulated attacks or adversarial strategies to remain robust under severe distortions (Kandi *et al*. 2017; Liu *et al*. 2019; Zhang *et al*.2018). A typical CNN watermarking model includes convolutional layers for feature extraction, pooling layers for reducing complexity and overfitting, and fully connected layers for producing outputs such as the watermarked image or extracted watermark (Rouhani *et al*. 2019; Lu *et al.* 2022), as illustrated in Figure (1). In addition, CNNs can learn resistance to distortions like JPEG compression, Gaussian noise, rotation, and cropping by training on attacked samples (Luo *et al.* 2020), making them practical for real-world applications (Ma *et al.* 2024). Finally, CNNs can also be combined with transform-domain techniques such as DWT to exploit both spatial and frequency information, improving imperceptibility and robustness by embedding in less perceptually sensitive frequency bands (Jaiswal & Pandey 2023; Huang *et al.* 2019).

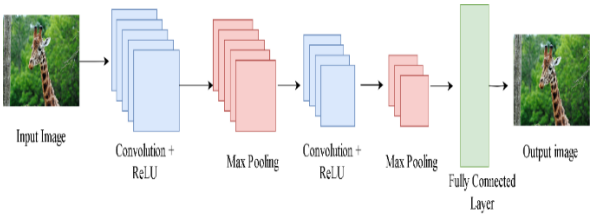
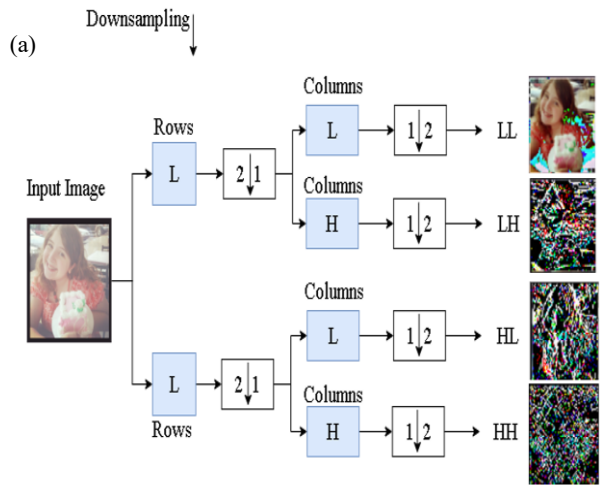


Figure 1: Diagram illustrating the architecture of CNN.

Discrete Wavelet Transform (DWT) in Watermarking:

The Discrete Wavelet Transform (DWT) (Mstafa 2022) is a multi-resolution method that converts an image from the spatial domain into the time-frequency domain, and it is widely used in image processing due to its strong energy compaction property, especially in denoising and compression. Since DWT provides both spatial and frequency information, it is highly suitable for digital watermarking (Giri *et al.* 2020). It is implemented using wavelet filter banks, and the Haar wavelet is one of the simplest and most popular choices because it is efficient and computationally inexpensive (Thakral & Manhas 2019). As is evident in Figure (2), DWT decomposes the image into four sub-bands (LL, LH, HL, HH), where LL contains most of the image energy, making it unsuitable for watermark embedding due to noticeable distortion, while HH is highly sensitive to compression and attacks. Therefore, watermark embedding is typically performed in the HL and LH sub-bands to achieve a good balance between imperceptibility and robustness (Guo *et al.* 2017; Yasmeen & Uddin 2021).



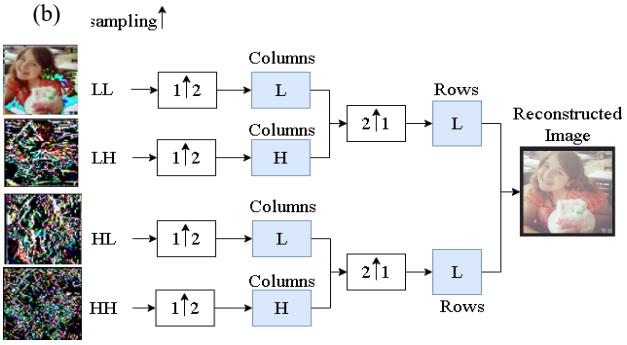


Figure 1: Visualization of Single-Level 2D DWT: (a) Sub-Band Decomposition and (b) Image Reconstruction.

4. PROPOSED METHOD

The proposed method combines DWT with CNNs via a U-Net-based architecture to achieve effective and imperceptible watermarking. Images are converted to YCbCr, and embedding is performed only in the Cb channel to reduce visible distortion. The channel is decomposed into frequency sub-bands using DWT, and the colored logo watermark is embedded into the HL1 coefficients through the U-Net model while preserving visual quality. Watermark extraction uses the same jointly trained architecture, making the method blind without requiring the original image. The CNN is enhanced with SE blocks to emphasize informative features and residual connections to retain image details. Training uses MSE for embedding accuracy and a hybrid loss (MSE + SSIM) to improve the quality of the extracted watermark. The overall framework is shown in Figure (3).

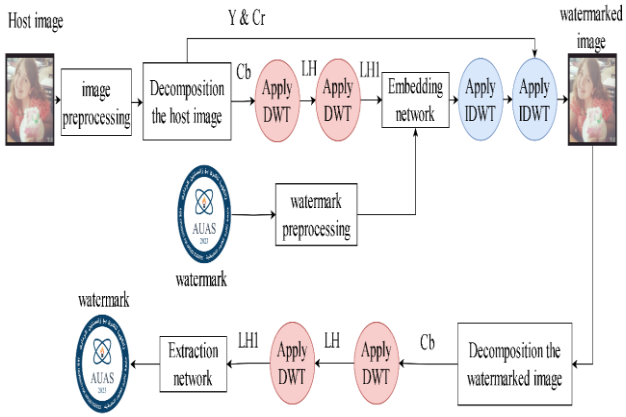


Figure 2: Schematic representation of the suggested watermarking method utilizing deep learning.

Preprocessing:

Before embedding, host images are resized to 128×128 pixels and normalized to [0,1]. RGB images are converted to the YCbCr color space, and the Cb channel is used for embedding to minimize visual distortion. A two-level DWT with the bior4.4 wavelet is applied, generating LL, LH, HL, and HH sub-bands at each level, and the HL1 sub-band is selected as the embedding domain for a balanced trade-off between imperceptibility and robustness.

Watermark Embedding Method:

After preprocessing, the HL1 sub-band from the Cb channel is used to embed the watermark using a modified U-Net with Squeeze-and-Excitation (SE) blocks and residual connections to enhance feature selection and training stability. The network has four encoder and decoder blocks with a bottleneck; encoder blocks extract deep features from HL1 using 3×3 convolutions (64–1024 filters), batch normalization, Leaky ReLU, and max

pooling, while decoder blocks up-sample with transposed convolutions and skip connections.

The watermark is resized and encoded via a 1×1 convolution, then concatenated with HL1 features at the bottleneck for deep embedding. The decoder reconstructs the HL1 sub-band containing the watermark, supervised by MSE loss, and the final watermarked image is obtained via IDWT and conversion back to RGB. This design integrates the watermark into the feature structure, preserving visual quality and enhancing robustness against distortions, with SE blocks ensuring the network focuses on relevant features (Figure 4).

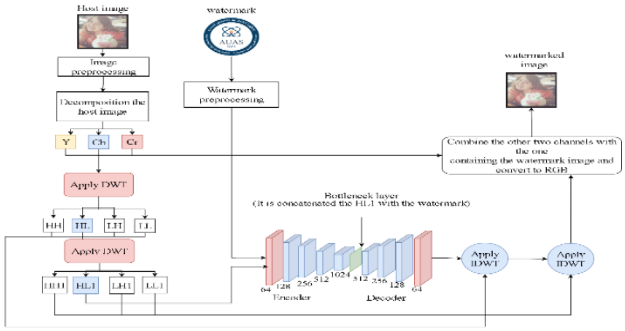


Figure 4: The watermark integration mechanism.

Watermark Extraction Method:

In the proposed method, the extraction process starts from the HL1 sub-band of the Cb channel, obtained after applying the preprocessing steps on the watermarked images as described in Section 4.1. This sub-band is then fed into an extractor network based on the same U-Net architecture used for embedding, with Squeeze-and-Excitation (SE) blocks and residual connections to enhance feature learning. Gaussian noise is injected during training to improve robustness.

The extractor is trained to recover the watermark using a hybrid loss function that combines Mean Squared Error (MSE) and Structural Similarity Index (SSIM), ensuring both numerical accuracy and perceptual quality. By operating directly in the frequency domain, the model achieves efficient and reliable watermark retrieval. Operations can sequentially be seen graphically in Figure (5).

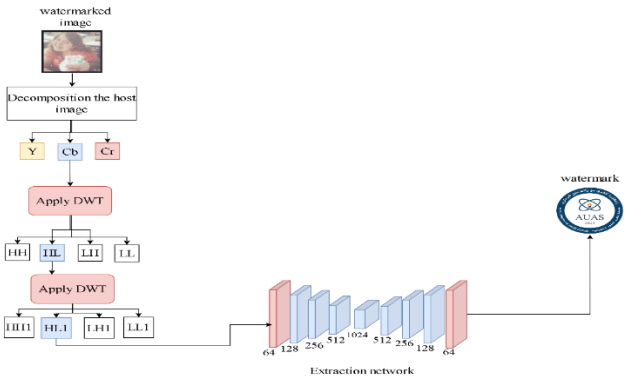


Figure 5: Watermark recovery mechanism

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section discusses an extensive description of the training procedure, including the dataset used, the computing resources deployed, experimental findings achieved, and evaluation metrics adopted.

Training and Dataset:

The publicly available COCO2017 dataset from Kaggle was used to train and evaluate the proposed watermarking method, as it provides diverse natural images with more than 160,000 samples across 80 categories (Lin *et al.* 2014). A subset of 10,000 images was randomly selected for training, with 250 images for validation and 250 for testing, then resized to 128×128 and normalized to [0,1], while a single-color 128×128 logo was used as the watermark. To improve robustness and reduce overfitting, on-the-fly augmentations such as random flips, brightness variation, and contrast adjustment were applied. The images were converted from RGB to YCbCr, and watermark embedding was performed only in the Cb channel using a two-level DWT to extract the HL1 sub-band, where a U-Net embedder inserted the watermark using MSE loss, and a U-Net extractor recovered it using a hybrid MSE + SSIM loss to maintain visual quality. The model was trained in TensorFlow with GPU acceleration using Adam, along with early stopping and checkpointing, and performance was evaluated using PSNR, SSIM, NC, and BER, while the main training settings are summarized in Table 1.

Table 1: Training settings for the proposed model.

Setting	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	4
Epochs	40
Regularization	Early stopping, checkpoints

Evaluation Metrics:

Image quality is a key criterion for evaluating any digital image watermarking system, providing objective evidence of its efficiency and reliability, while ensuring the watermark remains imperceptible and robust against attacks. In this study, watermarked image quality is assessed using Peak Signal-to-Noise Ratio (PSNR) (Abraham & Paul 2019) and Structural Similarity Index Measure (SSIM) (Rajani and Kumar 2020). PSNR reflects the distortion introduced during embedding, where lower values indicate higher distortion and higher values indicate better transparency (Chopra *et al.* 2018). Typically, a PSNR above 30 dB is considered good, meaning the watermarked image closely resembles the original visually, with PSNR calculated as shown in Eq.1 and expressed in decibels (dB).

$$PSNR = 20 \times \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \quad (1)$$

Where:

- MSE is the Mean Squared Error between the two images.
- MAX I is the maximum possible pixel value in the image (which is set to 1.0 in this case).
- PSNR is the Peak Signal-to-Noise Ratio.
- Steps for calculation:

The MSE is computed as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(I_1(i) - I_2(i))^2 \quad (2)$$

where $I_1(i)$ and $I_2(i)$ are the pixel values in images 1 and 2, respectively, and N is the number of pixels. Then, PSNR is calculated using the formula above (Eq.1).

SSIM is used to evaluate the visual similarity between the original and watermarked images (Kumar & Singh2021). Based on the human visual system, SSIM ranges from -1 to 1, where 1 indicates a perfect match. In practice, distortions from

watermark embedding prevent a perfect score, but values closer to 1 indicate higher visual resemblance. The SSIM equation is given in (Eq.3).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

Where:

- x and y are the two image patches that are compared.
- μ_x and μ_y are the mean values of x and y, respectively.
- σ_x^2 and σ_y^2 are the variances of x and y, respectively
- σ_{xy} s the covariance between x and y.
- $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ are small constants to avoid instability when the denominator is close to zero.
- L is the dynamic range of the pixel values (usually 255 for 8-bit images).
- K_1 and K_2 are constants, typically set to 0.01 and 0.03.

In addition to the previous evaluation metrics, Normalized Correlation (NC) (Zheng & Zhang 2020) and Bit Error Rate (BER) (Kang *et al.* 2018) were used to assess the robustness of the proposed method against image processing attacks. NC measures the similarity between the extracted and original watermark, ranging from 0 to 1, where 1 indicates perfect similarity and minimal distortion, and values near 0 indicate significant differences. Higher NC values reflect greater watermark robustness. NC is computed as in (Eq.4).

$$NC = \frac{\sum_i (wm_{\text{original}}(i) \cdot wm_{\text{extracted}}(i))}{\sqrt{\sum_i (wm_{\text{original}}(i)^2) \cdot \sum_i (wm_{\text{extracted}}(i)^2)}} \quad (4)$$

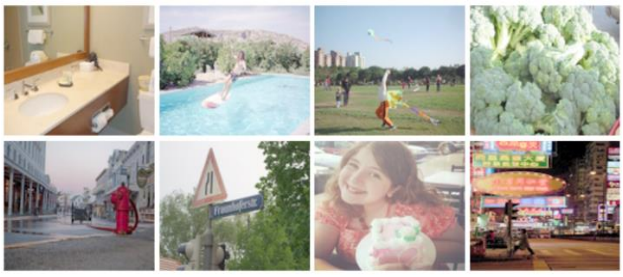
Where:

- $wm_{\text{original}}(i)$ and $wm_{\text{extracted}}(i)$ are the pixel values of the original watermark and the extracted watermark at pixel i, respectively.
- The numerator is the sum of the element-wise product of the original and extracted watermark values.
- The denominator is the square root of the sums of the original and extracted watermark pixels' squared values.

The BER is another measure used to evaluate the robustness of the proposed method, calculated as shown in (Eq.5). BER is defined as the number of incorrectly retrieved bits divided by the total embedded bits, ranging from 0 to 1. Unlike NC, it works inversely: a BER of 0 indicates perfect watermark retrieval, while values close to 1 indicate high dissimilarity between the original and extracted watermarks. Therefore, lower BER values correspond to higher system robustness.

$$BER(W,W') = \frac{\sum_{l=1}^{L_w} X\,OR(W(l),W'(l))}{L_w} \quad (5)$$

where W and W' refer for the extracted and original binary watermarks, respectively. The watermark's length is indicated by the parameter Lw.



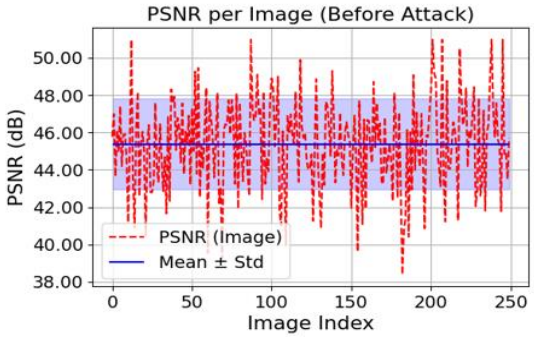
(a)

(b)

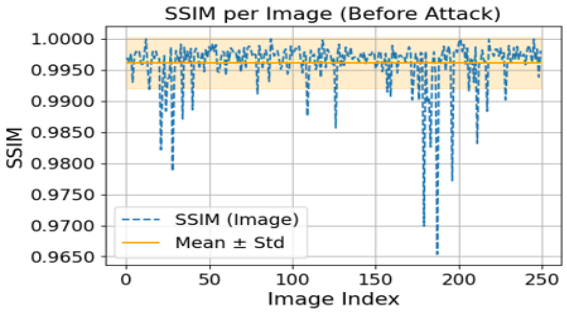
Figure 6: (a) A selection of color host images used for testing, sourced from the COCO dataset, and (b) the watermark image employed in the experiments.

Analysis of Imperceptibility:

Imperceptibility is a key criterion for secure watermarking, ensuring that embedded information remains invisible to the human eye. To evaluate the proposed method, the logo image (Figure 6b) was embedded into COCO dataset images (Figure 6a) following the process in Section 4.1, and PSNR and SSIM values were calculated using Eq. (1) and (3). The results (Figure 7) show an average PSNR above 45.35 dB and SSIM values close to 1, indicating excellent watermark invisibility, as confirmed by visual examples (Figure 8). The nearly constant PSNR and SSIM across 250 host images also demonstrate the method's robustness and reliability in consistently preserving imperceptibility, confirming that the proposed approach meets this essential criterion for effective watermarking.



(a)



(b)

Figure 7: Results of imperceptibility (a) PSNR values and (b) SSIM values.

(a)

(b)

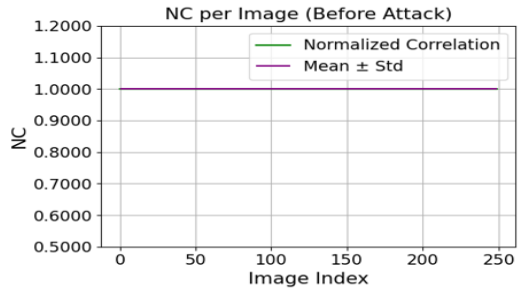
Figure 8: Depiction of Watermark Embedding: (a) Original Host Images and (b) Watermarked Outputs Produced by the Proposed Method.

Analysis of Robustness Against Attacks:

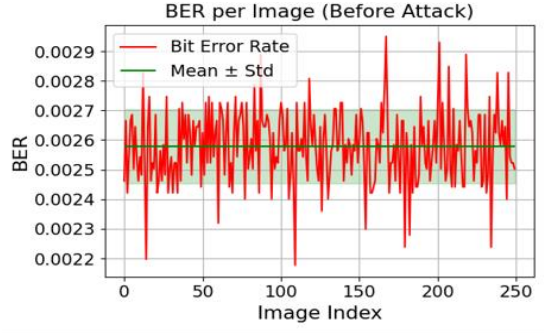
Robustness ensures that a watermarking algorithm can accurately retrieve the embedded watermark even after the watermarked images undergo various attacks, making it a key criterion for evaluating performance under adverse conditions (Abdallah *et al.* 2009). In this study, NC and BER were used to test robustness. Watermarked images were subjected to a range of standard attacks, and the watermark was extracted following the method in Section 4.2. The extracted watermark was then compared with the original using Eq. (4) and (5), providing a comprehensive evaluation of the algorithm's performance under typical distortions.

No attack:

In this study, the watermarked images were evaluated under ideal, attack-free conditions. The average NC and BER values reflect the robustness of the proposed method. As shown in Figure (9), the method achieves an average NC of 0.9999 and a BER of 0.0026 across all tested color images, indicating excellent watermark decoding while preserving image quality without interference.



(a)



(b)

Figure 9: Results of robustness without attack (a) NC values and (b) BER values.

Noise Attack:

In this experiment, Gaussian and salt-and-pepper noise were applied to the watermarked images to test robustness. As reported in Tables 2 and 3, the method achieved NC values close to 1 and consistently low BER values (often near zero), indicating that the extracted watermarks remained highly similar to the original. These results demonstrate that the proposed approach is strongly robust to noise, especially Gaussian noise.

Tables 2: *NC and BER values under Gaussian noise attack.*

Strength	0.1	0.2	0.3	0.4
NC	0.9999	0.9999	0.9999	0.9999
BER	0.0026	0.0024	0.0023	0.0022

Table 3: Performance results (NC and BER) under Salt and Pepper noise disturbance.

Strength	0.06	0.08	0.1	0.2
NC	0.9995	0.9992	0.9987	0.9970
BER	0.0058	0.0077	0.0100	0.0191

JPEG Compression Attack:

JPEG compression is widely used to reduce image file size while maintaining some visual quality, making it an important factor in digital watermarking where storage and transmission matter. In this study, watermarked images were subjected to JPEG compression at different quality factors (QFs), with resulting NC and BER values reported in Table 4. For QFs between 70 and 90, NC remained near 1 and BER near 0, showing that retrieved watermarks closely match the originals. Lower QFs (50 and 30) caused noticeable degradation, especially at QF = 30, confirming the method's strong resistance to JPEG compression at moderate to high quality levels.

Table 4: Performance metrics for different JPEG quality factors.

Strength-QF	90	70	50	30
NC	0.9999	0.9999	0.9998	0.9998
BER	0.0027	0.0028	0.0028	0.0029

Filtering Attack:

In the proposed experiment, watermarked images were subjected to various filtering attacks, including Gaussian, average, and median filtering. The corresponding NC and BER values are presented in Tables 5, 6, and 7. Results indicate that increasing the mask size does not significantly affect NC and BER, and the proposed method demonstrates excellent robustness against all filtering attacks, particularly Gaussian, average, and median filtering, validating its strength

Tables 5: Performance of NC and BER metrics after applying Gaussian filtering.

Filter size	3*3	5*5	7*7	9*9
NC	0.9999	0.9999	0.9999	0.9999
BER	0.0026	0.0026	0.0027	0.0027

Tables 6: Performance evaluation of watermark extraction after average filtering (NC and BER).

Filter size	3*3	5*5	7*7	9*9
NC	0.9999	0.9999	0.9999	0.9999
BER	0.0026	0.0026	0.0026	0.0027

Tables 7: NC and BER performance after applying median filtering to watermarked images.

Filter size	3*3	5*5	7*7	9*9
NC	0.9990	0.9990	0.9998	0.9999
BER	0.0054	0.0054	0.0029	0.0028

Geometric Attacks:

The watermarked images were tested against geometric attacks including rotation, cropping, and dropout. As shown in Tables 8, 9, and 10, cropping had minimal effect, with NC values close to 1 and BER near 0, indicating strong resistance. Similarly, rotation and dropout attacks produced NC and BER results close to ideal, demonstrating the proposed method's reliability in extracting watermarks under geometric distortions.

Tables 8: NC and BER values under Cropping attack.

Strength	0.2	0.3	0.4	0.5
NC	0.9999	0.9999	0.9999	0.9999
BER	0.0026	0.0026	0.0027	0.0029

Tables 9: NC and BER performance under rotational distortion.

Strength	30%	45%	60%	90%
NC	0.9999	0.9999	0.9999	0.9998
BER	0.0023	0.0024	0.0024	0.0028

Tables 10: NC and BER values under Dropout attack.

Strength	0.1	0.3	0.5	0.7
NC	0.9988	0.9958	0.9954	0.9954
BER	0.0081	0.0210	0.0230	0.0216

Visual Examples of Extracted Watermarks Under Attacks:

To complement the quantitative analysis, Table 11 presents visual samples of the extracted watermarks after applying various attacks. These examples include Salt and Pepper noise, JPEG compression, median filtering, rotation, and dropout. As observed, the proposed method successfully preserves the watermark's visual integrity under most attack conditions, consistent with the high NC and low BER values reported in Tables 2–10.

Tables 11: Examples of recovered watermarks under different attacks.

Attacks	Recovered watermark
Salt and Pepper noise (0.2)	
JPEG compression (QF=30)	
Median Filtering (3*3)	
Rotation (90%)	
Dropout (0.7)	

Comparison With Modern Techniques:

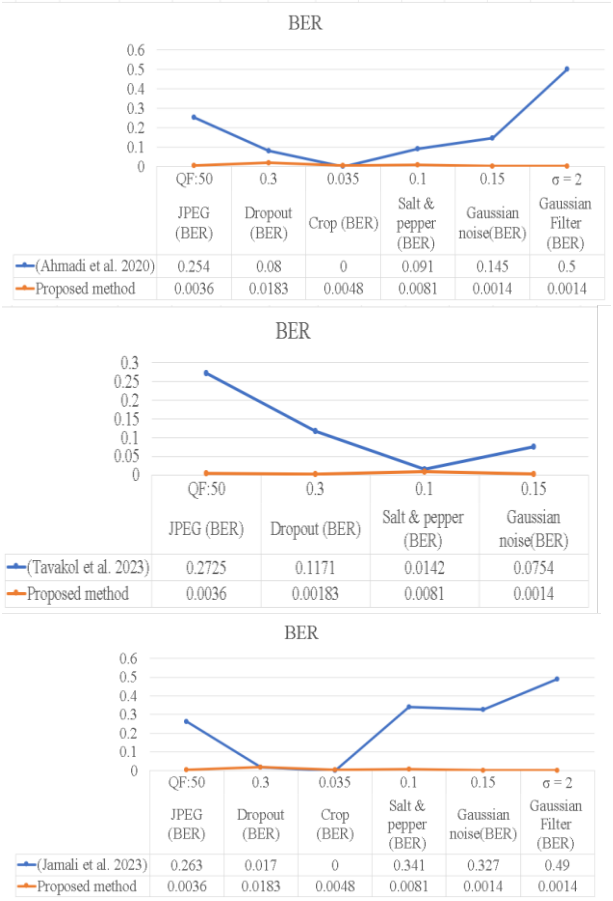
This section presents an empirical comparison between the proposed method and recent image watermarking techniques, focusing on imperceptibility and robustness. All methods were tested on the COCO dataset under identical conditions. As shown in Table 12, the proposed method achieves higher PSNR and SSIM values than alternative approaches, indicating superior visual quality and imperceptibility. Table 13 and Figure 10 demonstrate its robustness against various attacks, including JPEG compression, Gaussian filtering, Gaussian noise, and Salt and Pepper noise, where it generally outperforms methods by Ahmadi *et al.* (2020) Tavakoli *et al*. (2023) and Jamali *et al.* (2023). In Dropout attacks, our method surpasses those of Ahmadi et al. (2020) and Tavakoli et al. (2023), but is slightly less robust than Jamali et al. (2023), while in Cropping attacks, some existing methods show marginally better resilience. Overall, the results confirm that the proposed method effectively balances high imperceptibility with strong robustness, making it a competitive solution among current image watermarking techniques.

Tables 12: Assessment of visual quality metrics (PSNR and SSIM) for the suggested method versus other existing methods.

Method	PSNR	SSIM	Publication date
Ahmadi *et al.* 2020	40.24	0.987	2020
Rai *et al*. 2023	54.64	0.97	2023
Tavakoli *et al.* 2023	40.1	—	2023
Jamali *et al*. 2023	40.34	0.991	2023
Padhi *et al.* 2024	46.87	0.95	2024
Ma *et al.* 2024	35.973	0.992	2024
Proposed method	**45.35**	**0.996**	2025

Tables 13: Robustness Comparison (BER) Between the Suggested Method and Other Techniques.

Attack	Strength	(Ahmadi *et al.* 2020)	(Tavakoli *et al.* 2023)	(Jamali *et al.* 2023)	Proposed method
JPEG (BER)	QF: 50	0.254	0.2725	0.263	**0.0028**
Dropout (BER)	0.3	0.08	0.1171	**0.017**	0.0210
Crop (BER)	0.035	**0**	—	**0**	0.0024
Salt & pepper (BER)	0.1	0.0910	0.0142	0.3410	**0.0100**
Gaussian noise (BER)	0.15	0.145	0.0754	0.327	**0.0026**
Gaussian Filter (BER)	$\sigma = 2$	0.5	—	0.49	**0.0027**



6. CONCLUSION

This paper proposed a blind and robust watermarking method for color images using deep learning, combining discrete wavelet transform (DWT) with a U-Net-based CNN to leverage the multilayer capabilities of DWT and the learned features of U-Net for an effective balance between imperceptibility and robustness. Embedding was performed in the Cb channel of the YCbCr color space using a DWT, with the watermark inserted into the HL1 sub-band via a U-Net model enhanced with attention blocks. Evaluation on the COCO2017 dataset demonstrated high visual quality, with an average PSNR of 45.35 dB, SSIM of 0.996, and strong robustness against noise, JPEG compression, filtering, rotation, and cropping attacks.

Despite these results, limitations remain, including the lack of testing against resizing and histogram equalization, and the use of only a single colored watermark. Future work will address these attacks, explore multiple and diverse watermark designs, and optimize the model for deployment in resource-constrained environments.

Acknowledgements:

The authors sincerely thank Duhok Polytechnic University and Akre University for Applied Sciences, Technical College of Informatics (Akre) for providing support and research resources. They also express deep appreciation to their supervisor for guidance and continuous support. They send special thanks to their families for encouragement and motivation throughout the research period.

Ethics Statement:

Ethical approval was not required for this study because it did not involve human participants or animal experiments.

Author Contributions:

The MSc student A.H.Y, under the supervision of A.O.M, conducts this paper as a part of MSC research. Both authors have reviewed the final version to be published and have agreed to be accountable for all aspects of the work.

Funding:

This research received no external funding.

REFERENCES:

Abdallah, E. E., Ben Hamza, A., & Bhattacharya, P. (2009). Watermarking 3D models using spectral mesh compression. *Signal, image and video processing*, *3*(4), 375-389. DOI: https://doi.org/10.1007/s11760-008-0079-y.

Abdullatif, M., Khalifa, O. O., Olanrewaju, R. F., & Zeki, A. M. (2014, September). Robust image watermarking scheme by discrete wavelet transform. In *2014 International Conference on Computer and Communication Engineering* (pp. 316-319). IEEE. DOI: 10.1109/ICCCE.2014.95.

Aberna, P., & Agilandeeswari, L. (2024). Digital image and video watermarking: methodologies, attacks, applications, and future directions. *Multimedia Tools and Applications*, *83*(2), 5531-5591.DOI: https://doi.org/10.1007/s11042-023-15806-y.

Abraham, J., & Paul, V. (2019). An imperceptible spatial domain color image watermarking scheme. *Journal of King Saud University-Computer and Information Sciences*, *31*(1), 125-133. DOI: https://doi.org/10.1016/j.jksuci.2016.12.004.

Ahmadi, M., Norouzi, A., Karimi, N., Samavi, S., & ReDMark, A. E. (2019). Framework for residual diffusion watermarking based on deep networks., 2020, 146. *DOI:* https://doi. org/10.1016/j. eswa, 113157.

Ben Jabra, S., & Ben Farah, M. (2024). Deep learning-based watermarking techniques challenges: a review of current and future trends. *Circuits, Systems, and Signal Processing*, *43*(7), 4339-4368. DOI: https://doi.org/10.1007/s00034-024-02651-z.

Benoraira, A., Benmahammed, K., & Boucenna, N. (2015). Blind image watermarking technique based on differential embedding in DWT and DCT domains. *EURASIP Journal on Advances in Signal Processing*, *2015*(1), 55.DOI: https://doi.org/10.1186/s13634-015-0239-5.

Boujerfaoui, S., Riad, R., Douzi, H., Ros, F., & Harba, R. (2022). Image watermarking between conventional and learning-based techniques: a literature review. *Electronics*, *12*(1), 74. DOI: https://doi.org/10.3390/electronics12010074.

Chopra, J., Kumar, A., Aggarwal, A. K., & Marwaha, A. (2018, February). An efficient watermarking for protecting signature biometric template. In *2018 5th international conference on signal processing and integrated networks (SPIN)* (pp. 413-418). IEEE. DOI: 10.1109/SPIN.2018.8474269.

Darwish, M. M., Farhat, A. A., & El-Gindy, T. M. (2024). Convolutional neural network and 2D logistic-adjusted-Chebyshev-based zero-watermarking of color images. *Multimedia Tools and Applications*, *83*(10), 29969-29995. DOI: https://doi.org/10.1007/s11042-023-16649-3.

Giri, K. J., Quadri, S. M. K., Bashir, R., & Bhat, J. I. (2020). DWT based color image watermarking: a review. *Multimedia Tools and Applications*, *79*(43), 32881-32895. DOI: https://doi.org/10.1007/s11042-020-09716-6.

Gull, S., & Parah, S. A. (2024). Advances in medical image watermarking: a state of the art review. *Multimedia Tools and Applications*, *83*(1), 1407-1447.DOI: https://doi.org/10.1007/s11042-023-15396-9.

Guo, Y., Li, B. Z., & Goel, N. (2017). Optimised blind image watermarking method based on firefly algorithm in DWT-QR transform domain. *IET Image processing*, *11*(6), 406-415.DOI: https://doi.org/10.1049/iet-ipr.2016.0515.

Hosny, K. M., Magdi, A., ElKomy, O., & Hamza, H. M. (2024). Digital image watermarking using deep learning: A survey. *Computer Science Review*, *53*, 100662. DOI: https://doi.org/10.1016/j.cosrev.2024.100662.

Hsu, L. Y., & Hu, H. T. (2021). QDCT-based blind color image watermarking with aid of GWO and DnCNN for performance improvement. *IEEE Access*, *9*, 155138-155152. DOI: 10.1109/ACCESS.2021.3127917

Huang, Y., Niu, B., Guan, H., & Zhang, S. (2019). Enhancing image watermarking with adaptive embedding parameter and PSNR guarantee. *IEEE Transactions on Multimedia*, *21*(10), 2447-2460.DOI:10.1109/TMM.2019.2907475.

Jaiswal, S., & Pandey, M. K. (2023, February). Deep artificial neural network based blind color image watermarking. In *Doctoral Symposium on Human Centered Computing* (pp. 101-112). Singapore: Springer Nature Singapore. DOI: https://doi.org/10.1007/978-981-99-3478-2_10.

Jamali, M., Karimi, N., Khadivi, P., Shirani, S., & Samavi, S. (2023). Robust watermarking using diffusion of logo into auto-encoder feature maps. *Multimedia Tools and Applications*, *82*(29), 45175-45201. DOI: https://doi.org/10.1007/s11042-023-15371-4.

Juarez-Sandoval, O. U., Garcia-Ugalde, F. J., Cedillo-Hernandez, M., Ramirez-Hernandez, J., & Hernandez-Gonzalez, L. (2021). Imperceptible–visible watermarking to information security tasks in color imaging. *Mathematics*, *9*(19), 2374.DOI: https://doi.org/10.3390/math9192374.

Kandi, H., Mishra, D., & Gorthi, S. R. S. (2017). Exploring the learning capabilities of convolutional neural networks for robust image watermarking. *Computers & Security*, *65*, 247-268. DOI: https://doi.org/10.1016/j.cose.2016.11.016.

Kang, X. B., Zhao, F., Lin, G. F., & Chen, Y. J. (2018). A novel hybrid of DCT and SVD in DWT domain for robust and invisible blind image watermarking with optimal embedding strength. *Multimedia Tools and Applications*, *77*(11), 13197-13224. DOI: https://doi.org/10.1007/s11042-017-4941-1.

Kumar, S., & Singh, B. K. (2021). DWT based color image watermarking using maximum entropy. *Multimedia Tools and Applications*, *80*(10), 15487-15510. DOI: https://doi.org/10.1007/s11042-020-10322-9.

Lee, J. E., Seo, Y. H., & Kim, D. W. (2020). Convolutional neural network-based digital image watermarking adaptive to the resolution of image and watermark. *Applied Sciences*, *10*(19), 6854. DOI: https://doi.org/10.3390/app10196854.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, Y., Guo, M., Zhang, J., Zhu, Y., & Xie, X. (2019, October). A novel two-stage separable deep learning framework for practical blind watermarking. In *Proceedings of the 27th ACM International conference on multimedia* (pp. 1509-1517) DOI: https://doi.org/10.1145/3343031.3351025.

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1740

Lu, J., Ni, J., Su, W., & Xie, H. (2022, July). Wavelet-based CNN for robust and high-capacity image watermarking. In *2022 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE.DOI: 10.1109/ICME52920.2022.9859725.

Luo, X., Zhan, R., Chang, H., Yang, F., & Milanfar, P. (2020). Distortion agnostic deep watermarking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13548-13557).DOI: https://doi.org/10.48550/arXiv.2001.04580.

Luo, Y., Tan, X., & Cai, Z. (2024). Robust Deep Image Watermarking: A Survey. *Computers, Materials & Continua*, *81*(1). DOI: 10.32604/cmc.2024.055150.

Ma, L., Fang, H., Wei, T., Yang, Z., Ma, Z., Zhang, W., & Yu, N. (2024, September). A Geometric Distortion Immunized Deep Watermarking Framework with Robustness Generalizability. In *European Conference on Computer Vision* (pp. 268-285). Cham: Springer Nature Switzerland. DOI: https://doi.org/10.1007/ 978-3-031-72855-6_16.

Mahapatra, D., Amrit, P., Singh, O. P., Singh, A. K., & Agrawal, A. K. (2023). Autoencoder-convolutional neural network-based embedding and extraction model for image watermarking. *Journal of Electronic Imaging*, *32*(2), 021604-021604. DOI: https://doi.org/10.1117/1.JEI.32.2.021604.

Mahto, D. K., Anand, A., & Singh, A. K. (2022). Hybrid optimisation-based robust watermarking using denoising convolutional neural network. *Soft Computing*, *26*(16), 8105-8116.DOI: https://doi.org/10.1007/s00500-022-07155-z.

Mohammed, A. O., Hussein, H. I., Mstafa, R. J., & Abdulazeez, A. M. (2023). A blind and robust color image watermarking scheme based on DCT and DWT domains. *Multimedia Tools and Applications*, *82*(21), 32855-32881. DOI: https://doi.org/10.1007/s11042-023-14797-0.

Mousavi, S. M., Naghsh, A., & Abu-Bakar, S. A. R. (2014). Watermarking techniques used in medical images: a survey. *Journal of digital imaging*, *27*(6), 714-729.DOI: https://doi.org/10.1007/s10278-014-9700-5.

Mstafa, R. J. (2025). Reversible video steganography using quick response codes and modified elgamal cryptosystem. *arXiv preprint arXiv:2508.07289*.. DOI: 10.32604/cmc.2022.025791.

Padhi, S. K., Tiwari, A., & Ali, S. S. (2024). Deep Learning-based Dual Watermarking for Image Copyright Protection and Authentication. *IEEE Transactions on Artificial Intelligence*. DOI: 10.1109/TAI.2024.3485519.

Rai, M., Goyal, S., & Pawar, M. (2023). An optimized deep fusion convolutional neural network-based digital color image watermarking scheme for copyright protection. *Circuits, Systems, and Signal Processing*, *42*(7), 4019-4050. DOI: https://doi.org/10.1007/s00034-023-02299-1.

Rajani, D., & Kumar, P. R. (2020). An optimized blind watermarking scheme based on principal component analysis in redundant discrete wavelet domain. *Signal Processing*, *172*, 107556. DOI: https://doi.org/10.1016/j.sigpro.2020.107556.

Rouhani, B. D., Chen, H., & Koushanfar, F. (2019, April). Deepsigns: an end-to-end watermarking framework for protecting the ownership of deep neural networks. In *ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (Vol. 3, p. 1). DOI: https://doi.org/10.1145/3297858.3304051.

Singh, D., & Singh, S. K. (2017). DWT-SVD and DCT based robust and blind watermarking scheme for copyright protection. *Multimedia Tools and Applications*, *76*(11), 13001-13024. DOI: https://doi.org/10.1007/s11042-016-3706-6.

Singh, R., Saraswat, M., Ashok, A., Mittal, H., Tripathi, A., Pandey, A. C., & Pal, R. (2023). From classical to soft computing based watermarking techniques: A comprehensive review. *Future Generation Computer Systems*, *141*, 738-754. DOI: https://doi.org/10.1016/j.future.2022.12.015.

Sinhal, R., Jain, D. K., & Ansari, I. A. (2021). Machine learning based blind color image watermarking scheme for copyright protection. *Pattern Recognition Letters*, *145*, 171-177. DOI: https://doi.org/10.1016/j.patrec.2021.02.011.

Subramanian, N., Cheheb, I., Elharrouss, O., Al-Maadeed, S., & Bouridane, A. (2021). End-to-end image steganography using deep convolutional autoencoders. *IEEE Access*, *9*, 135585-135593. DOI:10.1109/ACCESS.2021.3113953.

Sy, N. C., Kha, H. H., & Hoang, N. M. (2020). An efficient robust blind watermarking method based on convolution neural networks in wavelet transform domain. *Int. J. Mach. Learn. Comput*, *10*, 675-684. DOI: 10.18178/ijmlc.2020.10.5.990

Tavakoli, A., Honjani, Z., & Sajedi, H. (2023). Convolutional neural network-based image watermarking using discrete wavelet transform. *International Journal of Information Technology*, *15*(4), 2021-2029. DOI: https://doi.org/10.1007/s41870-023-01232-8.

Thakral, S., & Manhas, P. (2018, July). Image processing by using different types of discrete wavelet transform. In *International Conference on Advanced Informatics for Computing Research* (pp. 499-507). Singapore: Springer Singapore..DOI: https://doi.org/10.1007/978-981-13-3140-4_45.

Yasmeen, F., & Uddin, M. S. (2021). An efficient watermarking approach based on LL and HH edges of DWT–SVD. *SN Computer Science*, *2*(2), 82. DOI: https://doi.org/10.1007/s42979-021-00478-y.

Ye, G., Gao, J., Yin, B., Xie, W., & Wei, X. (2023, October). Deep boosting robustness of DNN-based image watermarking via Dbmark. In *2023 International Conference on Culture-*

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1740

Oriented Science and Technology (CoST) (pp. 186-191). IEEE. DOI: 10.1109/CoST60524.2023.00046.

Zear, A., Singh, A. K., & Kumar, P. (2018). A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine. *Multimedia tools and applications*, *77*(4), 4863-4882.DOI: https://doi.org/10.1007/s11042-016-3862-8 .

Zermi, N., Khaldi, A., Kafi, R., Kahlessenane, F., & Euschi, S. (2021). A DWT-SVD based robust digital watermarking for medical image security. *Forensic science international*, *320*, 110691.DOI: https://doi.org/10.1016/j.forsciint.2021.110691.

Zhang, L., & Wei, D. (2019). Dual DCT-DWT-SVD digital watermarking algorithm based on particle swarm optimization. *Multimedia Tools and Applications*, *78*(19), 28003-28023. DOI: https://doi.org/10.1007/s11042-019-07902-9.

Zhang, X., Peng, F., & Long, M. (2018). Robust coverless image steganography based on DCT and LDA topic classification. *IEEE Transactions on Multimedia*, *20*(12), 3223-3238.. DOI: 10.1109/TMM.2018.2838334.

Zhao, X., Huang, P., & Shu, X. (2022). Wavelet-Attention CNN for image classification. *Multimedia Systems*, *28*(3), 915-924. DOI: https://doi.org/10.1007/s00530-022-00889-8.

Zheng, P., & Zhang, Y. (2020). A robust image watermarking scheme in hybrid transform domains resisting to rotation attacks. *Multimedia Tools and Applications*, *79*(25), 18343-18365. DOI: https://doi.org/10.1007/s11042-019-08490-4.

Zhong, X., Das, A., Alrasheedi, F., & Tanvir, A. (2023). A brief, in-depth survey of deep learning-based image watermarking. *Applied Sciences*, *13*(21), 11852. DOI: https://doi.org/10.3390/app132111852.

Zhong, X., Huang, P. C., Mastorakis, S., & Shih, F. Y. (2020). An automated and robust image watermarking scheme based on deep neural networks. *IEEE Transactions on Multimedia*, *23*, 1951-1961. DOI: 10.1109/TMM.2020.3006415.

DOI: https://doi.org/10.25271/sjuoz.2026.14.2.1740