Robust Image Tampering Detection and Ownership Authentication Using Zero-Watermarking and Siamese Neural Networks

Rodrigo Eduardo Arevalo-Ancona[1], Manuel Cedillo-Hernandez[2], Francisco Javier Garcia-Ugalde[3]
Instituto Politecnico Nacional, SEPI-ESIME Culhuacan, Mexico City, Mexico[1, 2]
Universidad Nacional Autonoma de Mexico, Facultad de Ingenieria, Mexico City, Mexico[3]

***Abstract*—The development of advanced image editing tools has significantly increased the manipulation of digital images, creating a pressing need for robust tamper detection and ownership authentication systems. This paper presents a method that combines zero-watermarking with Siamese neural networks to detect image tampering and verify ownership. The approach utilizes features from the Discrete Wavelet Transform (DWT) and employs two halftone images as watermarks: one representing the owner's portrait and the other corresponding to the protected image. A feature matrix is generated from the owner's portrait using the Siamese network and securely linked to the image's halftone watermark through an XOR operation. Additionally, data augmentation enhances the model's robustness, ensuring effective learning of image features even under geometric and signal processing distortions. Experimental results demonstrate high accuracy in recovering halftone images, enabling precise tamper detection and ownership verification across different datasets and image distortions (geometric and image processing distortions).**

***Keywords*—Zero-watermarking; tampering detection; ownership authentication; neural network**

I. Introduction

In recent years, digital image manipulation has become a growing concern with the development of advanced editing tools, many of which are easily accessible to the general public. This issue poses significant challenges to the integrity and authenticity of digital images in applications such as copyright protection, digital security, and forensic investigations [1], [2]. Manipulation techniques such as copy-move and splicing can compromise the content and image ownership [3], [4], [5]. Furthermore, ownership authentication is crucial to prevent unauthorized use or distribution, ensuring their protection [6].

Traditional watermarking methods verify the authenticity and ownership of images. However, these approaches degrade the image quality by embedding watermark signals into the visual content. To address this issue, zero-watermarking techniques preserve image quality by creating a master share that is not embedded into the image content. This method creates a master share by linking image features with a watermark for authentication, avoiding distortions and preserving image quality [7], [8].

Additionally, deep neural networks enhance tamper detection and image authentication by learning patterns and semantic representations from images. This approach significantly improves the efficiency of detecting manipulations and verifying image content.

This paper presents a robust zero-watermarking methodology that integrates zero-watermarking with Siamese neural networks for tamper detection and ownership verification. The approach utilizes the Discrete Wavelet Transform (DWT) to extract key features from images, which are then used to train the Siamese network and generate a feature matrix. This matrix is linked to the watermark—a proprietary halftone image—through an XOR operation, enabling accurate watermark recovery even in the presence of geometric distortions and signal processing attacks. Additionally, tamper detection is achieved by analyzing image blocks' eigenvalues to identify image manipulations effectively. The main contributions of this work are as follows:

- Enhance ownership authentication and tamper detection using a zero-watermarking technique that verifies image ownership and detects manipulations by combining image features extracted by the Siamese neural network with the watermark, ensuring security and integrity.
- Create a robust feature representation with Siamese Neural Networks. Robust feature representation is achieved because the neural network learns unique and invariant image features. One branch is trained with undistorted images, while the other is trained with distorted images, providing an efficient method improving the accuracy of ownership authentication and increases the method's robustness against manipulated images.
- Data augmentation techniques prevent overfitting and improve generalization across different types of images. This is essential because overfitting is common in neural network models, especially when working with limited or highly distorted data. In addition, data augmentation enhances feature extraction since the model learns more robust image features.
- The Siamese neural network learns and compares features, analyzing similarities and differences between the original halftone image's and distorted image's DWT coefficients. This is critical for tamper detection and authenticity verification.

- Optimized feature extraction using the Low-Low (LL) sub-band coefficients of the DWT. These filters remove irrelevant high-frequency details, focusing on the most relevant information for watermarking. This step is necessary to increase the robustness of the feature extraction process for watermark recovery, ensuring that important features are preserved even when the image has geometric or image processing distortions.
- A precise tamper detection method that compares the image blocks eigenvalues between the original halftone image and a distorted version, providing an additional layer of accuracy for tampering detection.

The paper is organized as follows: Section II reviews existing methods. Section III describes the proposed technique. Section IV presents and analyzes experimental results. Finally, Section V concludes with the study's advantages and limitations.

II. Literature Review

This section presents an overview of recent techniques developed for image tampering detection and image authentication, providing the necessary context for the method proposed in this paper.

Several approaches are focused on detecting tampered regions by extracting specific image features. For example, Xing et al. [9] used a high-pass filter to capture edge information, concentrating on modified regions to identify manipulated areas. Alsughayer et al. [10] employed a U-Net architecture to extract residual noise from images, identifying tampered regions in remote sensing data. However, noise-based methods are susceptible to false positives when noise is introduced by natural image compression or processing.

With the rise of deep learning, some network-based solutions for tamper detection were developed, such as Priyadharsini et al. [11], who applied a modified GoogleNet model to extract features using a nearest neighbor algorithm for splicing detection. However, this method is computationally expensive and is prone to overfitting with small datasets. Ren et al. [12] introduced a neural network for low-level feature extraction, focusing on geometric details and semantic segmentation for copy-move detection. Goel et al. [13] used a dual neural network for inherent feature extraction, although their method requires large amounts of labeled data for accurate detection. Hosny et al. [14] proposed a method based on deep neural networks for extracting specific image features, while Chu [15] used a generative neural network to detect forged regions. Das et al. [16] leveraged MobileNet V2 to detect manipulated areas, highlighting the potential of lightweight models for real-time applications. Nikalie et al. [17] combined convolutional neural networks with Local Binary Pattern (LBP) analysis to detect texture inconsistencies. Mallick et al. [18] incorporated VGG16 and VGG19 models to identify manipulations across images with different compression levels, but their approach requires significant computational resources. Dai et al. [19] employed the Xception model to extract image edges and textures, offering improvements in detecting fine manipulations.

On the other hand, recent zero-watermarking methods for image authentication integrated neural networks to improve robustness. Xiang et al. [20] combined watermarking with features generated by a ResNet-based neural network, which relied on correlations between neural network layer responses to verify authenticity. However, these models are susceptible to geometric distortions. Dong et al. [21] proposed a method using NasNet-Mobile features and Discrete Cosine Transform (DCT) coefficients to create a master share, allowing watermark recovery without degrading the image. Similarly, Li et al. [22] introduced an image authentication method based on ConvNeXt layers and Swin Transformer optimization for feature extraction, although the approach lacks comprehensive evaluation against geometric attacks. He et al. [23] further advanced the field by integrating convolutional layers based on the Swin Transformer to optimize feature extraction for zero-watermarking, yet the robustness under signal processing distortions remains underexplored.

While significant progress has been made in both neural network-based tamper detection and watermarking techniques, these methods still need to improve their robustness, especially under geometric distortions and signal processing attacks. In contrast, the method proposed in this paper addresses these limitations for image ownership authentication and detects image tampering by combining Siamese neural networks with zero-watermarking. This ensures high accuracy in tamper detection and strong resilience to various distortions without compromising image quality. Additionally, the Discrete Wavelet Transform (DWT) allows for more efficient feature extraction, focusing on relevant low-frequency components, further enhancing robustness.

III. Proposed Image Tampering Detection and Ownership Authentication Method

The proposed algorithm introduces a robust zero-watermarking technique for image protection, creating a master share, which is generated by logically linking the halftone image representation with features extracted from the watermark, represented by the owner's portrait, through a neural network (Fig. 1(a)). The preprocessing stage begins by dividing the input image into 32x32 pixel blocks and applying the Discrete Wavelet Transform (DWT) to extract the Low-Low (LL) sub-band coefficients, which enhance the system's robustness against geometric distortions and signal processing attacks.

The features extracted by the Siamese neural network are carefully selected because they capture invariant and distinctive representations of the watermark, which are crucial for an accurate ownership verification and tamper detection. The Floyd-Steinberg dithering algorithm generates the halftone effect [24], [25].

For the tamper detection and ownership verification stage, the master share is combined with the features extracted by the Siamese network to recover the halftone representation of the protected image. The eigenvalues from the feature matrices of the recovered and potentially manipulated images are compared to identify image forgery regions. (Fig. 1(b)).

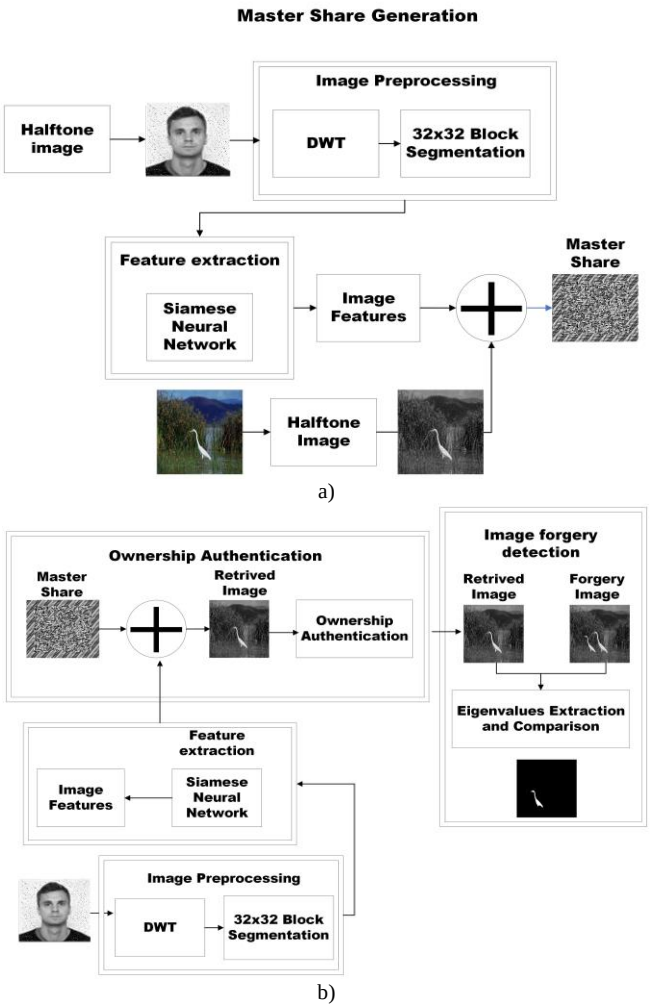


Fig. 1. (a) General diagram of the proposed master share generation algorithm (b) Ownership authentication and tamper detection procedures.

The proposed method consists of the following stages:

Preprocessing: Halftone representations of both the image and the watermark are created. The halftone watermark is decomposed using the Haar wavelet transform (DWT), focusing on the LL sub-band. This sub-band is divided into 32x32 blocks used to train the neural network. Dividing the image into smaller blocks identifies specific features from each image region more effectively, improving the accuracy of feature extraction and reducing computational complexity.

Feature extraction for the master share generation: The Siamese neural network extracts key features from the watermark. These features are combined with the image's halftone representation to generate the master share.

Image ownership authentication: The halftone image is retrieved by linking the watermark's extracted features with the master share. This step validates image ownership, offering robust intellectual property protection.

Image tampering detection: The eigenvalues from the halftone and retrieved images are compared to detect forgeries. Eigenvalues are sensitive to structural changes in the image, making them an effective tool to identify image tampering.

A. Preprocessing

In the preprocessing stage, the halftone image representations are generated for the image and the owner's portrait (watermark). The halftone representation simplifies the visual content with binary patterns, eliminating unnecessary details. Then, the owner's halftone image is transformed into the frequency domain using the DWT with Haar wavelets. Haar wavelets are effective for image processing tasks capturing spatial and frequency information. The DWT decomposes the image into frequency sub-bands, each capturing specific image characteristics. The low-low (LL) sub-band is selected for further processing and contains low frequency image components. These components represent the most important structural details, such as edges and large-scale patterns, which are less susceptible to distortion like noise or compression and eliminate redundant features. The LL sub-band coefficients are divided into 32x32 pixel blocks to optimize the system's robustness. This segmentation enhanced feature extraction because each block represents an image region, allowing the neural network to learn specific features. In addition, when working with smaller blocks, the overall processing time is reduced, and memory usage is optimized, making the algorithm more scalable and suitable for real-time applications, highlighting the system's adaptability.

B. Feature Extraction for the Master Share Generation

In this stage, the Siamese neural network extracts relevant features from the halftone watermark. The Siamese network is designed to compare two images and learn invariant features robust to distortions and manipulations (Fig. 2). This neural network model consists of two identical subnetworks, ensuring the feature matrix contains specific representations from the watermark. In this case, one subnetwork processes the original DWT watermark blocks while the other processes a distorted version of the watermark DWT blocks.

The features extracted from the watermark by the Siamese network are used to construct a feature matrix, which encodes the unique characteristics of the watermark. This feature matrix contains patterns from the watermark and the distorted watermark version with geometric and image processing attacks. Then, the feature matrix is combined with the halftone representation of the image. The combination is achieved using an XOR operation.

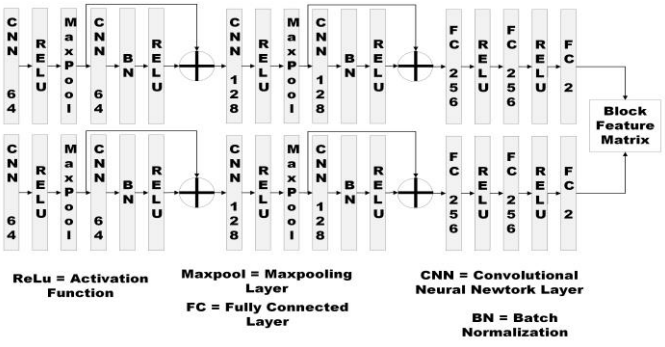


Fig. 2. Siamese neural network architecture.

The resulting master share contains a visual structure that encrypts the image and the semantic features of the watermark. This technique preserves the image quality since no watermark is embedded directly into the visual content of the image, maintaining its integrity.

The parameters and configurations used in the training process to increase the robustness and accuracy in the system are presented in Table I. The Siamese neural network architecture consists of two identical branches. To train the neural network one branch processes the DWT watermark blocks, and the second branch processes the DWT distorted watermark blocks. This architecture is especially useful when the data is limited. The purpose of the network is to compare image features from both input images and identify unique patterns in each region. The network creates a feature matrix (f_{vm}) (2) based on the vectors generated from the output of the Siamese neural network (v_1, v_2) related to each image region, representing the unique characteristics.

$$v_1 = f_\theta(I_1)$$
$$v_2 = f_\theta(I_2) \quad (1)$$

$$f_{vm} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (2)$$

Then feature matrix with the Siamese neural network output (f_{vm}) has a size of (3) and is restructured to the image halftone size (f_m) (4).

$$size = (2, (\frac{m\ x\ n}{block\ size})) \quad (3)$$

$$fm = \begin{bmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{m,1} & \cdots & c_{m,n} \end{bmatrix} \quad (4)$$

m is the height of the image and *n* the width of the image.

1) Neural network training: One key aspect is data augmentation, which enhances the model's feature detection. Data augmentation is important when the amount of data is limited, as it artificially increases the dataset by applying distortions to the input images. The image processing attacks applied during augmentation include different image filters for image blurring (average, median, Gaussian, motion blurring), JPEG compressions (quality factor = 90, 60, 10), noise addition (Gaussian noise σ = 0.09, 0.009 and salt and pepper), rotations (random rotations from 10° to 350°), translations (image pixels shift in x, y and xy axis = 50, 100, 150), cropping (random cropping with sizes in x and y= 50, 100, 150), and scaling (scaling factors = 2, 0.5, 0.25).

The application of these augmentation techniques ensures that the network learns robust, invariant features to obtain high ownership authentication and image tampering detection accuracy even in the presence of noise, distortions, or other forms of manipulation.

TABLE I. SIAMESE NEURAL NETWORK PARAMETERS

Epochs	Learning rate	Momentum	Optimazer	Batch size
3	0.0001	0.9	Stochastic Gradient Descent	64

2) Master share generation: The features extracted from the watermark by the Siamese network are used to construct a f_m, which encodes the unique characteristics of the watermark (W).

This feature matrix is then combined with the halftone representation of the watermark using an XOR operation ($\oplus$) to create the master share (*MS*) (5) (see Fig. 3). This operation effectively encrypts the image while preserving its quality, as no watermark is embedded directly into the visual content of the image. This XOR-based encryption method securely. The resulting master share (see Fig. 4(d)) is then stored and distributed. To recover the original image, the master share must be used together with the watermark.

$$ms = f_m \oplus H \quad (5)$$

Once the master share has been generated and stored, image ownership authentication can be carried out. The following sections detail how this ownership authentication process works and how the master share is utilized for both authentication and tamper detection.

C. Image Ownership Authentication and Tamper Detection

To verify ownership, the master share used watermark during recovery even if it is distorted. This element ensures that only the owner, who possesses the watermark, can authenticate the image. The authentication process involves recovering the halftone image from the master share using both the owner's portrait and the master share. This procedure, illustrated in Fig. 5, confirms the image's ownership. The ownership is confirmed by reconstructing the halftone image and verifying its authenticity.

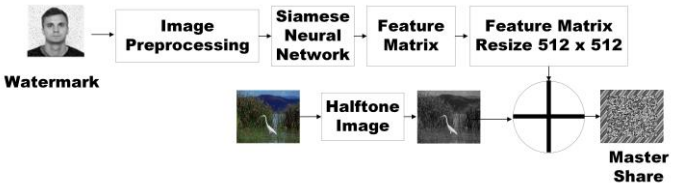


Fig. 3. Master share generation process.

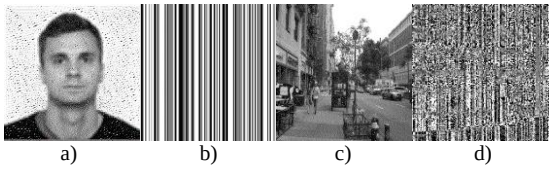


Fig. 4. a) Halftone watermark, b) Feature matrix generated from the Siamese neural network, c) Halftone image, d) Master share.

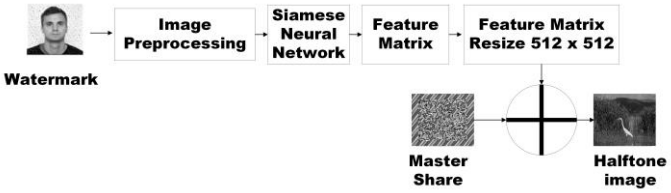


Fig. 5. Image halftone recovery process.

This process involves decoding the master share (*ms*) to reconstruct the halftone image (H_r) by combining the feature matrix from the Siamese neural network (f_m) with the master share using the XOR ($\oplus$) logical operation (6).

$$H_r = f_m \oplus ms \tag{6}$$

Once ownership has been confirmed through the successful reconstruction of the halftone image, it is important to ensure that the image has not been tampered. While the ownership authentication process guarantees that only the owner can verify the image, the tampering detection process compares the original and recovered halftone images to detect unauthorized modifications. The following section details how the system uses eigenvalues to detect manipulations by comparing the structural integrity of the original and recovered images, ensuring that the image remains unchanged and free from tampering.

D. Tampering Detection

The manipulation detection used the recovered and manipulated halftone images to identify any potential alterations. In this stage, eigenvalues are calculated for both the recovered halftone and potentially manipulated images. The images are divided into smaller blocks, and the eigenvalues for each block are calculated. These eigenvalues represent distinct structural characteristics of each image and are compared to identify discrepancies. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are computed by (7).

$$v_I = \det(I - \lambda) \text{ and } v_m = \det(I_m - \lambda) \tag{7}$$

where I is the retrieved halftone image, I_m is the manipulated image, v_I is the vector with the eigenvalues from the retrieved image and $v_m$is the vector with the eigenvalues from the manipulated image. The eigenvalues comparison is realized by the Euclidean distance (*d*) (8).

$$di = \sqrt{(\lambda_I - \lambda_m)^2} \tag{8}$$

For each block, if the distance exceeds a predefined threshold (*th*), the block (b_d) is considered manipulated (9).

$$b_d = \begin{cases} 0 & if \quad d \le th \\ 1 & if \quad d > th \end{cases} \tag{9}$$

The results of this process, illustrating the detection of manipulations, are shown in Fig. 6. This method enhances the precision of manipulation detection. Block-based eigenvalue comparisons allow localized forgery detection, ensuring that even subtle tampering can be identified, thus offering a more detailed and accurate analysis of image manipulations.

a) b) c)

Fig. 6. a) Original image, b) Tampered image, c) Forgery detection.

In the following section, the experimental results obtained from this methodology demonstrating its effectiveness for ownership authentication and image tampering detection.

IV. Experimental Results

This section presents the experimental results of the proposed algorithm for both ownership authentication and image tampering detection. The image watermark was subjected to a wide range of image processing techniques and geometric distortions, simulating real scenarios to assess the method's robustness. The experimental results demonstrate the algorithm's effectiveness in accurately verifying ownership and detecting manipulations, even when the watermark has been significantly altered.

The algorithm was implemented in a system equipped with an NVIDIA GTX 960 graphics card and an Intel Core i7-6700 processor running at 3.4 GHz. This hardware configuration provided a suitable environment for efficiently executing the algorithm within the PyTorch Python framework, ensuring smooth performance during training and testing. A 512 x 512 pixel halftone representation of the author's face was used as the watermark for the experiments. The proposed algorithm's performance was evaluated using three widely used image datasets, showcasing its ability to handle diverse scenarios and conditions:

MICC-F220: This dataset contains 220 RGB images, equally divided into 110 manipulated and 110 non-manipulated images. It evaluates owner authentication and tamper detection accuracy in a balanced setting [26].

MICC-F2000: This dataset is comprised of 2,000 RGB images (700 manipulated and 1,300 non-manipulated), which is used for a more extensive assessment of the algorithm's performance across diverse scenarios and conditions [26].

CASIA V2: This larger dataset includes 12,613 RGB images, 5,123 manipulated, and 7,490 non-manipulated images. Due to its size and diversity, CASIA V2 provides a robust evaluation framework, testing the algorithm's scalability and effectiveness across a wide range of image manipulations [27], [28].

These datasets cover different image manipulation scenarios to evaluate the algorithm's robustness and accuracy. The experiments using these datasets quantitatively assess the algorithm's ability to recover the halftone images for ownership authentication and image forgery detection. The results offer valuable insights into the algorithm's effectiveness in different conditions, showcasing its ability to detect subtle manipulations and authenticate image ownership.

A. Assessment of Image Ownership Authentication

The proposed method's performance is assessed in terms of ownership authentication, emphasizing its robustness against different image processing and geometric distortions. To ensure this robustness, two metrics are employed: Bit Error Rate (BER) and Normalized Cross-Correlation (NC). These metrics determine the algorithm's performance under image processing and geometric distortions applied to the watermark.

The BER measures the error bits between the original and recovered halftone images. A lower BER indicates that the algorithm successfully retrieves the watermark (10).

$$BER = \frac{Total\ Incorrect\ pixels}{Total\ pixels} \tag{10}$$

The NC metric measures the similarity between the original and recovered halftone images. The NC value ranges from 0 to 1, where a value of 1 indicates similarity between the images, and a value of 0 indicates no similarity (11).

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j)W_r(i,j))}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W(i,j))^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(W_r(i,j))^2}} \tag{11}$$

where, $H(i,j)$ is the halftone image, H_r is the retrieved halftone image M and N are the dimensions of the images. These metrics ensure a comprehensive evaluation of the algorithm's watermark recovery stage performance for ownership authentication, even under image distortion and manipulation.

TABLE II. HALFTONE IMAGE RECOVERY WITH WATERMARK GEOMETRIC DISTORTIONS

	Rotation 80 °	Rotation with cropping 250 °	Translation X = 90 Y = 90	Translation X = 160
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0
	Cropping upper left	Center cropping	Scale 64 x 64	Scale 1064 x 1064
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0

Tables II and III demonstrate the proposed method's robustness in recovering the halftone image even when the watermark has geometric distortions. Under distortions like rotation without cropping, translation, and scaling, the recovery performance remains nearly perfect, with BER = 0 and NC = 0.9999. Minor recovery errors occur with rotations involving cropping and downscaling, where BER reaches 0.0078 and NC drops slightly to 0.9914. These results demonstrated the method's robustness against geometric distortions, ensuring an accurate watermark recovery for ownership verification.

TABLE III. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK GEOMETRIC DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
No attack	0	0.9999	Translation x = 80 pixels	0	0.9999
Rotation 15° with no cropping	0	0.9999	Translation x = 150 pixels	0	0.9999
Rotation 55° with no cropping	0	0.9999	Translation x = 20, y = 20 pixels	0	0.9999
Rotation 110°	0	0.9999	Translation x = 80, y= 80 pixels	0	0.9999
Rotation 135° with no cropping	0	0.9999	Translation x = 150, y = 150 pixels	0	0.9999
Rotation 20° with cropping	0.0078	0.9914	Upper left cropping 100 x 100	0	0.9999
Rotation 100° with cropping	0.0078	0.9923	Upper right cropping 100 x 100	0	0.9999
Rotation 265° with cropping	0.0078	0.9901	Botton right cropping 100 x 100	0	0.9999
Rotation 320° with cropping	0.0078	0.9914	Center cropping 100 x 100	0	0.9999
Translation y = 20 pixels	0	0.9999	Scale 256 x 256	0	0.9999
Translation y = 80 pixels	0	0.9999	Scale 64 x 64	0.0078	0.9914
Translation y = 150 pixels	0	0.9999	Scale 640 x 640	0	0.9999
Translation x = 20 pixels	0	0.9999	Affine transform $\begin{bmatrix} 5 & 10 & 0 \\ 10 & 5 & 0 \end{bmatrix}$	0	0.9999

Tables IV and V illustrate the effectiveness of the proposed method for recovering halftone images when the watermark is modified with image processing distortions, such as JPEG compression, affine transformations, histogram equalization, and noise addition. In most cases, the NC is 0.999, and the BER is 0, indicating high accuracy. Even under more challenging conditions, such as JPEG compression at a quality factor of 30, the method still maintains a high accuracy with NC = 0.992 and BER = 0.007. The proposed technique remains robust against distortions like salt-and-pepper noise and Gaussian filtering, achieving NC = 0.999 and BER = 0,

demonstrating its robustness in the halftone image recovery stage for ownership authentication.

TABLE IV. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK IMAGE PROCESSING DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
JPEG Quality Factor = 70	0	0.9999	Gaussian Filter kernel = 7 x 7	0	0.9999
JPEG Quality Factor = 50	0	0.9999	Median filter kernel = 7 x 7	0.0078	0.9916
JPEG Quality Factor = 30	0	0.9999	Gaussian noise $\mu = 0, \sigma = 0.009$	0	0.9999
Blurring kernel = 5 x 5	0.0078	0.9920	Gaussian noise $\mu = 0, \sigma = 0.09$	0	0.9999
Gaussian Filter kernel = 5 x 5	0.0078	0.9914	Salt and pepper noise 0.005	0.0078	0.9915
Median filter kernel = 5 x 5	0.0078	0.9911	Salt and pepper noise 0.05	0.0078	0.9914
Average filter kernel = 5 x 5	0.0078	0.9910	Gamma correction $\gamma = 1.25$	0.0078	0.9914
Blurring kernel = 7 x 7	0	0.9999	Gamma correction $\gamma = 0.80$	0.0078	0.9919
Histogram equalization	0	0.9999	Bright adjust	0.0078	0.9916

TABLE V. HALFTONE IMAGE RECOVERY WITH WATERMARK IMAGE PROCESSING DISTORTIONS

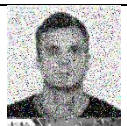
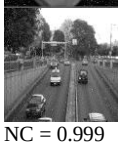
	JPEG 70	JPEG 30	Affine Transform	Histogram
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.992 BER = 0.007	NC = 0.999 BER = 0	NC = 0.999 BER = 0
	Gamma 1.5	Salt and pepper noise 0.09	Gaussian filter kernel = 7 x 7	Blurring kernel = 5x5
Distorted watermark				
Retrieved halftone image				
	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0	NC = 0.999 BER = 0

TABLE VI. HALFTONE IMAGE RECOVERY ASSESSMENT WITH WATERMARK COMBINED DISTORTIONS

Distortion	BER	NC	Distortion	BER	NC
$\gamma = 1.25$ and salt and pepper noise 0.005	0	0.9999	JPEG Quality Factor=30 and Blurring kernel=5x5	0	0.9999
$\gamma = 1.8$ and salt and pepper noise 0.005	0	0.9999	Bright adjust and scaling 1024x1024	0	0.9999
Rotation 35° and bright adjust	0	0.9999	JPEG Quality Factor=30 and Scaling 64x64	0	0.9999

Table VI evaluates the halftone image recovery under combined distortions applied to the watermark.

For all tested combinations, including gamma correction with salt-and-pepper noise, JPEG compression with blurring, brightness adjustment with scaling, and rotation with brightness adjustment, the method achieves NC = 0.9999 and BER = 0. These results indicate that the recovery process is highly robust even when multiple distortions are applied simultaneously to the watermark, ensuring the integrity and authenticity of the halftone image.

B. Tampering Detection Assessment

Following the evaluation of image ownership authentication, assessing the algorithm's effectiveness in detecting image tampering is essential, as both tasks are closely related. The tampering detection accuracy relies on the halftone image's successful recovery during the ownership authentication stage.

Table VII presents the results, demonstrating the robustness of the proposed method in detecting image forgery by analyzing the recovered and manipulated halftone images. These results highlight the system's effectiveness by accurately detecting any image manipulations. The proposed image tampering detection and localization algorithm is evaluated using five key metrics: accuracy, precision, recall, F1 score, and mean squared error (MSE).

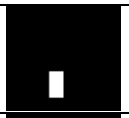
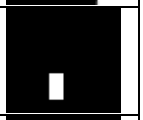
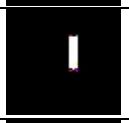
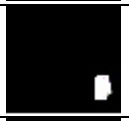
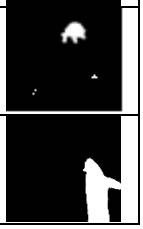
The accuracy represents the proportion of correctly detected pixels relative to the total number of pixels, indicating how well the algorithm identifies manipulated areas (12).

$$acc = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{12}$$

where T_p = true positives, T_n = true negatives, F_p = false positives, F_n = False Negatives. The precision measures the ratio of correctly detected pixels from the manipulated region (13).

$$precision = \frac{Tp}{Tp + Fp} \tag{13}$$

TABLE VII. IMAGE FORGERY DETECTION

Database	Original Image	Tampered Image	Ground Truth	Tampering Detection
MIC-F220				
				
MIC-F2000				
				
CASIA V2				
				

The recall measures the proportion of correctly detected pixels in the manipulated region (14).

$$recall = \frac{Tp}{Tp + Fn} \tag{14}$$

The F1 score provides a measure of the algorithm's performance. A high F1 score indicates that the algorithm achieves minimize false positives and false negatives (15)

$$F1 = \frac{2x(Precision\ x\ Recall)}{Precision + Recall} \tag{15}$$

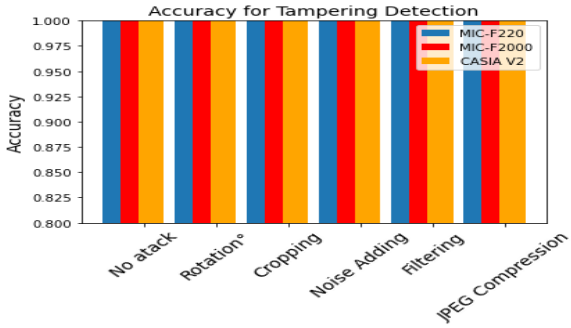


Fig. 7. Accuracy for watermark tampering with geometric distortions and image processing distortions.

Fig. 7 to Fig. 10 provide a detailed evaluation of the proposed method's effectiveness in detecting image manipulations. The results show high efficiency in detecting tampered areas, even when the watermark has been distorted. However, some loss of features is observed in the halftone image recovery when image processing distortions are applied to the watermark, which introduces errors in the recovery process. This error can impact the accuracy of tamper detection. Despite these challenges, the method identifies manipulations efficiently. High accuracy ensures that most manipulations are correctly identified, while precision and recall measure the method's ability to distinguish tampered areas from untampered ones. The F1 score, a balance between precision and recall, ensures a high performance even when the image is distorted.

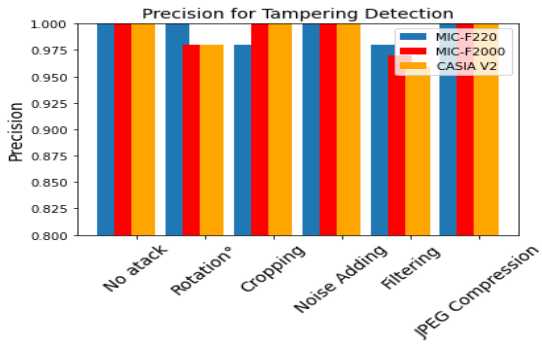


Fig. 8. Precision for watermark tampering with geometric distortions and image processing distortions.

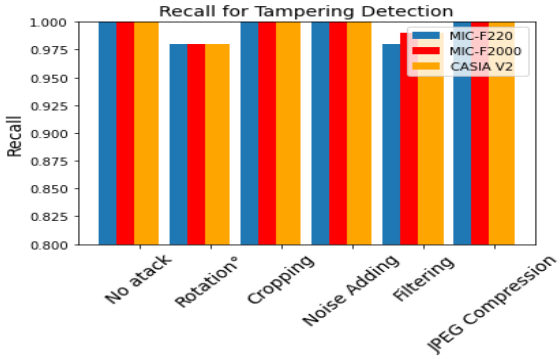


Fig. 9. Recall for for watermark tampering with geometric distortions and image processing distortions.

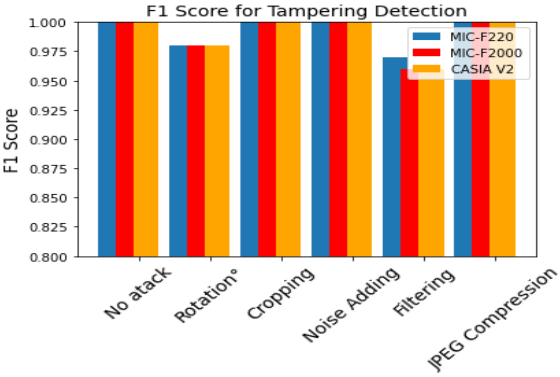


Fig. 10. F1 score for watermark tampering with geometric distortions and image processing distortions.

C. Ownership Authentication Performance Comparison

In this section, the effectiveness of the proposed image ownership authentication algorithm is compared to existing methodologies. This comparison highlights the proposed method's robustness against different image distortions relative to existing solutions in terms of accuracy, robustness, and efficiency.

Table VIII compares the proposed method with other existing techniques. One of the main advantages of the proposed method is block segmentation and the neural network model feature extraction for each region of the image. This processing provides the neural network with recognition of specific patterns related to each block, which generates a robust feature matrix. This methodology allows for specific feature analysis to improve the algorithm's capacity in image authentication. The comparison depicted that the method is more robust against most distortions; however, image recovery can generate errors in the case of filtration distortions. The results from the zero-watermarking comparison show that the proposed method outperforms existing approaches. This method uses LL coefficients from the DWT to train a Siamese network, generating a robust feature matrix. This generates higher accuracy in the recovery halftone stage. The results from Table VIII indicate that the proposed method is highly resistant to image quality loss, crucial for scenarios involving image storage or transmission. Also, the present technique shows robustness to rotations, scaling, and translations, while other methods obtain a BER of 0.02 for rotations of only 10°.

D. Image Forgery Detection Comparison

Table IX presents a comparison between the proposed method and other techniques developed for tamper detection. The results demonstrate the effectiveness of the proposed method in detecting and localizing image tampering across image databases. Furthermore, the results show that the proposed method achieves higher efficiency than other approaches, effectively identifying whether an image has been manipulated and accurately pinpointing the manipulated areas.

TABLE VIII. ZERO-WATERMARKING COMPARISON

	Xiang et al. [20]	Dong et al. [21]	Li et al. [22]	He et al. [23]	Proposed method
Methodology	Correlation between shallow and deep features. Use the ResNet model as a feature extractor.	DCT low frequencies from the features of the NasNet-Mobile model.	ConvNext blocks from the ZWnet	Redundant feature shrinkage and removal (SRFENet)	LL-DWT coefficients to train a Siamese Network and create a feature matrix from each image region.
Image type	Medical images	Medical images	Natural images	Natural images	Natural images
Watermark size	256 x 256	---	---	512 x 512	512 x 512
Gaussian Noise	σ = --- NC = 0.9836	σ = --- NC = 0.93	σ = 0.005 NC = 1	σ = 0.010 BER = 0.001	σ = 0.09 NC=0.9999 BER=0
Salt and Pepper noise	NC = 0.9836	---	Factor = 0.01 NC = 0.9922	Factor = 0.01 BER = 0.01	Factor = 0.05 NC=0.9999 BER=0
JPEG	Quality Factor = 10 NC = 0.9834	Quality Factor = 75 NC = 1	---	Quality Factor = 30 BER= .01	Quality Factor = 30 NC=0.9999 BER=0
Gaussian Filter	0.9834	---	NC = 1	BER = 0.015	NC=0.9920 BER=0 0078
Median Filter	Kernel 3x3 NC = 0.9833	Kernel 5x5 NC = 0.92	---	Kernel 3x3 BER= .01 Kernel 5x5 BER = 0.015	Kernel 3x3 NC=0.9920 BER=0.0078
Rotation	20° NC = 0.9835	---	15° NC = 0.9688	10° BER = 0.02	135° NC=0.9999 BER=0
Cropping	Size = 1/3 NC = 0.9834	Size = 10 % 0.94	Size = 1/8 NC = 0.9063	---	Size = 100 x 100 NC=0.9999 BER=0
Scaling	Factor 0.8 NC = 0.9835	Factor 0.3 NC = 0.88	---	Factor 0.8 BER = 0.01	Size = 64 x 64 NC=0.9999 BER=0
Translation	---	X = 15% NC = 0.94 Y = 10% NC = 0.86	---	---	X = 100, Y = 100 NC=0.9999 BER=0

TABLE IX. IMAGE FORGERY DETECTION COMPARISON

	Das et al [16]	Nikalje et al. [17]	Mallick et al. [18]	Dai et al. [19]	Proposed method
Tampering detection	Splicng	Splicing and copy-move	Splicing and copy-move	Splicing and copy-move	Splicing and copy-move
Technique	CNN based on transfer learning MobileNet	CNN and Local Binary Pattern	Pretrained VGG16 and VGG19	Dual-Net DeepLab V3	Siamese Neural Network and DWT LL coefficients
Accuracy	0.9301	0.9901	VGG16 = 0.944 VGG19 = 0.995	0.8725	MIC-F220 = 0.993 MIC-F2000 = 0.991 CASIA V2 = 0.997
Precision	0.926	0.9581	---	---	MIC-F220 = 0.986 MIC-F2000 =0.983 CASIA V2 =0.985
Recall	0.966	0.9661	---	---	MIC-F220 = 0.994 MIC-F2000 = 0.997 CASIA V2 = 0.985

V. CONCLUSION

The method proposed in this paper provides a robust solution for owner authentication and image manipulation detection. The use of halftone images for image authentication based on the Siamese neural network features implemented in the zero-watermarking technique increased the efficiency of the watermark recovery from the master share. Furthermore, this technique provides an additional security stage by encrypting the image in the master share. In addition, the results show robustness in the image halftone recovery process when the watermark is distorted with geometric and image processing attacks. This is reflected in a low error and high similarity between the recovered and original halftone images. The coefficients belonging to the LL sub-band of the DWT and their segmentation into blocks allow the neural network to recognize unique patterns from each image region even though the watermark is distorted. On the other hand, the architecture of two branches from the Siamese neural network detects unique and invariant characteristics related to the watermark.

However, the generated error in the image recovery process increases when image processing distortions are applied to the watermark because some image features are deleted; despite this, a low error can be observed in the halftone image recovery. Furthermore, the recovered halftone image process enhances the effectiveness of the proposed image tampering detection method. The retrieved halftone image with minimal error serves as a reference for detecting discrepancies between the original and potentially manipulated images. The comparison with existing methods highlights the efficiency of forgery detection. The proposed methodology's main contribution focuses on its double function since it performs image owner authentication without distorting the image and effectively detects image tampering. In addition, the retrieved image analysis from the owner authentication process allows the detection and localization of image tampering.

ACKNOWLEDGMENT

This work was supported in part by the Instituto Politecnico Nacional (IPN), Universidad Nacional Autonoma de Mexico (UNAM) under the Direccion General de Asuntos del Personal Academico (DGAPA) research project under grant PAPIITIT100123, and in part by Consejo Nacional de Humanidades, Ciencias y Tecnologias (CONAHCYT).

REFERENCES

[1] S. Tyagi, D. Yadav, “A detailed analysis of image and video forgery detection techniques”. Vis Comput vol. 39, pp. 813–833, 2023, doi; https://doi.org/10.1007/s00371-021-02347-4.

[2] K. B. Meena and V. Tyagi, “Image Forgery Detection: Survey and Future Directions”, Data Engineering and Applications, vol. 2, pp. 163-194, 2019, doi: https://doi.org/.

[3] M. Maashi, H. Alamro, H. Mohsen, N. Negm, G. P. Mohammed, N. A. Ahmed, S. S. Ibrahim and M. I. Alsaid, “Modeling of Reptile Search Algorithm With Deep Learning Approach for Copy Move Image Forgery Detection”, IEEE Access, vol. 11, pp. 87297 - 87304, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3304237.

[4] T. Pomari, G. Ruppert, E. Rezende, A. Rocha and T. Carvalho, “Image Splicing Detection Through Illumination Inconsistencies and Deep Learning”, IEEE International Conference on Image Processing, Athens, Greece, 2018, doi: https://doi.org/10.1109/ICIP.2018.8451227.

[5] Chen, H., Han, Q., Li, Q. et al., “Digital image manipulation detection with weak feature stream”. Vis Comput, vol. 38, pp. 2675–268, 2022, doi: https://doi.org/10.1007/s00371-021-02146-x.

[6] R. Sinhal and I. A. Ansari, “Machine learning based multipurpose medical image watermarking”, Neural Computing and Applications, vol. 35, 2023, doi: https://doi.org/10.1007/s00521-023-08457-5.

[7] Y. Li, J. Li, U. A. Bhatti, J. Ma, D. Li and F. Dong, “Robust Zero-watermarking Algorithm for Medical Images Based on ORB and DCT”, ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Taiyuan, China, 2023, doi: https://doi.org/10.1109/SNPD-Winter57765.2023.10223992.

[8] G. Sun, J. Li, U. A. Bhatti, J. Ma, F. Dong and Y. Li, “Robust Zero-Watermarking Algorithm for Medical Images Based on AGAST-LATCH and DCT”, ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Taiyuan, China, 2023, doi: https://doi.org/10.1109/SNPD-Winter57765.2023.10224055.

[9] J. Xing, X. Tian and Y. Han, “A Dual-channel Augmented Attentive Dense-convolutional Network for power image splicing tamper detection”, Neural Computing and Applications, 2024, doi: https://doi.org/10.1007/s00521-024-09511-6.

[10] R. Alsughayer, M. Hussain, F. Saeed and H. AboalSamh, “Detection and localization of splicing on remote sensing images using image-to-image transformation”, Applied intelligence, vol. 53, pp. 13275–13292, 2023, doi: https://doi.org/10.1007/s10489-022-04126-7.

[11] S. Priyadharsini and K. K. Devi, “Effective image splicing detection using deep neural network, International Journal of Wavelets”, Multiresolution, vol. 21, nº 2, pp. 2250051-2250079, 2022, doi: https://doi.org/10.1142/S0219691322500515.

[12] R. Ren, S. Niu, J. Jin, K. Xiong and H. Ren, “ERINet: efficient and robust identification network for image copy-move forgery detection and localization”, Applied Intelligence, vol. 53, p. 16170–16191, 2023, doi: https://doi.org/10.1007/s10489-022-04104-z.

[13] N. Goel, S. Kaur and R. Bala, “Dual branch convolutional neural network for copy move forgery detection”, IET Image Processing, 2020, doi: https://doi.org/10.1049/ipr2.12051.

[14] K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin, “An Efficient CNN Model to Detect Copy-Move Image Forgery”, IEEE Access, pp. 48622 - 48632, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3172273.

[15] L. Chu, “Research on Image Tampering Detection and Localization Based on Iterative GAN”, IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE), Jinzhou, China, 2023, doi: https://doi.org/ 10.1109/ICSECE58870.2023.10263546.

[16] D. Das and R. Naskar, “Image Splicing Detection based on Deep Convolutional Neural Network and Transfer Learning”, IEEE 19th India Council International Conference (INDICON), Kochi, India, 2023, doi: https://doi.org/10.1109/INDICON56171.2022.10039789.

[17] S. Nikalje and M. V. Mane, “Copy-Move and Image Splicing Forgery Detection based on Convolution Neural Network”, International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, doi: https://doi.org/ 10.1109/ICICICT54557.2022.9917679.

[18] D. Mallick, M. Shaikh, A. Gulhane and T. Maktum, Copy Move and Splicing Image Forgery Detection using CNN, ITM Web Conferences, vol. 44, pp. 03052-03058, 2022, doi: https://doi.org/10.1051/itmconf/20224403052.

[19] C. Dai, L. Su, B. Wu and J. Chen, DS-Net: Dual supervision neural network for image manipulation localization, IET Image Processing, 2023, doi: https://doi.org/10.1049/ipr2.12885.

[20] R. Xiang, G. Liu, K. Li, J. Liu, Z. Zhang and M. Dang,” Zero-watermark scheme for medical image protection based on style feature and ResNet”, Biomedical Signal Processing and Control, vol. 86, nº A, pp. 105127, 2023, doi: https://doi.org/10.1016/j.bspc.2023.105127.

[21] F. Dong, J. Li, U. A. Bhatti, J. Liu, Y.-W. Chen and D. Li, “Robust Zero Watermarking Algorithm for Medical Images Based on Improved NasNet-Mobile and DCT”, Electronics, vol. 16, nº 16, pp. 3444, 2023, doi: https://doi.org/10.3390/electronics12163444.

[22] C. Li, H. Sun, C. Wang, S. Chen, X. Liu, Y. Zhang, N. Ren and D. Tong, "ZWNet: A Deep-Learning-Powered Zero-Watermarking Scheme", Applied Sciences, vol. 14, pp. 435, 2024.

[23] L. He, Z. He, T. Luo and Y. Song, "Shrinkage and Redundant Feature Elimination Network-Based", Symmetry, vol. 15, nº 5, pp. 964, 2023, doi: https://doi.org/10.3390/sym15050964.

[24] Q. Dong, L. Feng and T. Lu, "Reversible Watermarking Algorithm for Halftone Images Based on Overlapping Blocks Scanning and Central Pixels Flipping", International Conference on Computer Graphics, Image and Virtualization (ICCGIV), Chongqing, China, 2023, doi: https://doi.org/10.1109/ICCGIV57403.2022.00054.

[25] W. Lu, Y. Xue, Y. Yeung, H. Liu, J. Huang and Y.-Q. Shi, "Secure Halftone Image Steganography Based on Pixel Density Transition", IEEE Transactions on Dependable and Secure Computing, vol. 18, nº 3, pp. 1137 - 1149, 2021, doi: https://doi.org/10.1109/TDSC.2019.2933621.

[26] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo and G. Serra, A SIFT-based forensic method for copy-move attack detection and transformation recovery, IEEE Transactions on Information Forensics and Security, vol. 6, nº 3, pp. 1099-1110, 2011, doi: https://doi.org/ 10.1109/TIFS.2011.2129512.

[27] J. Dong, W. Wang and T. Tan. "CASIA Image Tampering Detection Evaluation Database", IEEE China Summit and International Conference on Signal and Information Processing, Beijing,China, 2013, doi: https://doi.org/10.1109/ChinaSIP.2013.6625374.

[28] N. T. Pham, J.-W. Lee, G.-R. Kwon and C.-S. Park, "Hybrid Retrieval Method for Image Splicing Validation", Symmetry, vol. 11, n° 1, pp. 83, 2019, doi: https://doi.org/10.3390/sym11010083.