Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SUPERGAN: SUPERPIXEL-GUIDED GAN FOR ROBUST AND IMPERCEPTIBLE IMAGE WATERMARKING

ABDELHAY HASSANI ALLAF[1], M'HAMED AÏT KBIR[2]

[1, 2] Intelligent Automation & BioMedGenomics Laboratory (IABL), STSM Doctoral Studies Center,

Abdelmalek Essaadi University, Tangier, Morocco

E-mail: [1]a.hassani@uae.ac.ma, [2]m.aitkbir@uae.ac.ma

ABSTRACT

This paper presents SuperGAN, a novel image watermarking framework that combines superpixel segmentation with a generative adversarial network (GAN) to achieve robust, imperceptible, and content-aware watermark embedding. By leveraging SLIC-based superpixel segmentation, regions with high texture complexity and sufficient size are selected for embedding to enhance both resilience and visual transparency. A 32-bit watermark is redundantly embedded across multiple regions to improve recovery under distortions. The embedding process integrates a region-aware attention mechanism to modulate watermark bits locally and imperceptibly. Extensive experiments on the COCO dataset assess the method's performance under common attacks, including compression, noise, cropping, scaling, and rotation. Comparative results show notable improvements in robustness and imperceptibility. These findings highlight SuperGAN's ability to serve as a reliable tool for image watermarking applications in real-world.

Keywords: *Image Watermarking, Superpixels, Generative Adversarial Networks (GANs), Robustness, Attention Mechanism.*

1. INTRODUCTION

The rapid growth of communication technologies has led to the widespread distribution of digital content, raising significant concerns related to unauthorized copying, tampering, and the loss of ownership attribution. Digital watermarking has emerged as a key technique to address these issues by embedding hidden information (known as a watermark) within media files, thereby supporting authentication, copyright enforcement, and traceability.

The embedded digital watermark must remain visually undetectable, withstand various transformations, and contain sufficient information. Finding the right balance between these properties is a considerable challenge, especially in real-world applications where attacks may impose compression, noise, or geometric transformations.

Traditional methods (often based on spatial or transform domain) have been widely explored for watermark embedding. While spatial techniques are lightweight and offer high capacity, they typically lack robustness [1], [2]. On the other hand, transform-domain approaches provide stronger resilience but at the cost of higher computational complexity and reduced imperceptibility [3], [4], [5], [6].

In recent years, deep learning has offered new opportunities for watermarking, particularly through the use of generative adversarial networks (GANs) [7]. These models can learn from sophisticated image distributions, giving them the ability to incorporate watermarks while maintaining excellent visual quality. GAN-based methods [8], [9], [10], [11], [12], [13], by training against a discriminator, are able to reduce perceptual distortion while incorporating beneficial hidden information. However, many of these methods continue to treat all image areas uniformly, implementing standardized embedding without considering texture, structure, or local visual importance. This can affect robustness and invisibility, especially in cases where some areas are more suitable for watermarking than others.

To bridge this gap, we propose SuperGAN, a hybrid watermarking framework that combines GAN-based adversarial learning with superpixel segmentation and entropy-guided region selection. Unlike conventional models that embed uniformly across the image, SuperGAN identifies the most structurally stable and information-rich regions (those with high entropy, sufficient size, and strong textures) for watermark insertion. This spatially aware strategy significantly improves the method's resilience to image distortions while preserving visual quality.

Digital watermarking techniques have diverse applications, including copyright protection [14], [15], content indexing, tamper detection, broadcast monitoring, and medical image security [16], [17]. While watermarking methods have been developed for various media types such as audio [18], video [19], and documents, this study focuses specifically on digital image watermarking, with a particular emphasis on methods that provide strong robustness and high visual fidelity.

The main contributions of this study are summarized as follows:

- ✓ A superpixel-guided embedding strategy is proposed to concentrate watermark insertion in visually stable regions.
- ✓ A GAN-based encoder-decoder architecture is developed to ensure both high imperceptibility and accurate watermark recovery.
- ✓ A multi-region redundancy mechanism is introduced to increase resilience under partial image degradation or occlusion.
- ✓ Extensive experiments are conducted under five common attack types (JPEG compression, noise, scaling, cropping, and rotation), with comparisons against conventional and GAN-based watermarking methods.

The remaining of this paper is structured as follows: Section 2 reviews the literature on traditional and learning-based watermarking. Section 3 presents the SuperGAN framework and its components. Section 4 details the experimental setup and results under various attack conditions. Section 5 concludes the work and outlines future research directions.

2. RELATED WORK

To situate the recent developments in digital watermarking on images, this section gathers and summarizes past research on classical, transformative, hybrid, deep learning-based, and superpixel-guided watermarking techniques.

2.1 Traditional Image Watermarking Methods

Traditional image watermarking methods can be categorized according to the integration domain: spatial domain, frequency domain (transformation), and hybrid methods. Each option has advantages and limitations, depending on the watermarking requirements.

2.1.1 Spatial Domain Watermarking Methods

Spatial domain techniques embed watermark bits directly into pixel values, offering high capacity and computational efficiency [1], [2]. The simplest example is the Least Significant Bit (LSB) method, first introduced by Tanaka et al. [20], while the Patchwork algorithm by Yeo and Kim [21] improves statistical robustness via Gaussian modulation. However, these approaches are generally vulnerable to compression and noise [22].

To improve robustness, variants such as the Dual Intermediate Significant Bit (DISB) [23] technique target more stable bit planes. A visual quality-preserving enhancement was later proposed in [24]. More advanced versions include Rinki and Verma's matrix-based LSB approach [25] and Mustaqim et al.'s luminance-channel quantization method [26]. Belferdi et al. [27] use segmented blocks and RGB redundancy, respectively. Bhalerao et al. [28] employed SHA-1 hashing with LSB for authentication. Thongkor and Amornraksa [29] addressed distortions from camera acquisition, while Pan-Pan et al. [30] improved desynchronization resilience using affine-invariant embedding.

Other spatial methods link watermarking to image features. Content-based and corner-based schemes, such as those using Harris corner detection with triangular [31] or rectangular decomposition [32], [33].

More recent approaches focus on structural adaptation. Su et al. [34] modified blue-channel pixels based on DC coefficients and key quantization. Nana et al. [35] used chaotic SVD for block-level embedding. Abraham et al. [36] applied masking in the blue channel to minimize distortion, distributing the watermark across specific regions.

In video watermarking, LSB techniques have been extended across frames [37], while spread-spectrum strategies convert frames into 1D signals for embedding [38]. Further refinements targeting compression robustness have been proposed in [39], [40], and [41].

Despite their advantages in simplicity and capacity, spatial techniques remain less robust to signal and geometric attacks, prompting interest in transform and hybrid domains.

2.1.2 Transform Domain Watermarking Methods

Transform domain watermarking techniques embed watermark data into the frequency coefficients of an image, offering improved robustness over spatial domain approaches. Common transforms include the Discrete Cosine Transform (DCT), Discrete Fourier

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Transform (DFT), and Discrete Wavelet Transform (DWT) [3], [4]. These methods are resilient to compression, noise, and geometric distortions, especially when embedding in mid-frequency bands to balance invisibility and durability [5], [6].

DCT-Based Methods

Wang et al. [42] use the Y channel in YCbCr space, incorporating a JND model and Arnold scrambling for robustness. Loan et al. [43] combine DCT and chaotic encryption to improve resistance against filtering and cropping. Moosazadeh et al. [44] apply TLBO optimization to identify optimal embedding blocks. Zhang et al. [45] propose a tri-directional correlation scheme to resist signal distortions. Vishwakarma and Sisaudia [46] use DE and KELM to optimize embedding strength, while Rangel-Espinoza et al. [47] propose a blind dual watermarking strategy using QIM-DM.

DFT-Based Methods

DFT watermarking excels in scenarios with geometric changes. Prajwalasimha et al. [48] introduce a bulk watermarking scheme with high imperceptibility. Liao et al. [49] apply compressive sensing for reversible embedding, and Riad et al. [50] enhance print-scan robustness through Hough-based preprocessing. Chen et al. [51] improve screen-cam resistance via local-max extraction. Hsu et al. [52] combine QDFT, PSO, and EMMQ modulation to counter compression, while Cedillo-Hernandez et al. [53] balance quality and robustness using PSO and VIF metrics.

DWT-Based Methods

DWT techniques are valued for multiresolution embedding and perceptual fidelity. Yuan et al. [54] combine DWT and CGI to resist common noise types. Kumar and Singh [55] embed in YCbCr channels using entropy-guided alpha blending. For secure medical data transmission, Anand and Singh [56] use DWT with turbo coding and Paillier encryption. Al-Otum [57] enhances signal robustness using WPT and chaotic scrambling. Tan et al. [58] apply JND thresholds to embed in multiple YCbCr components, and Zhang and Wei [59] integrate RIIWT, QR, and SVD for improved geometric attack resistance.

Overall, transform domain techniques offer stronger robustness than spatial methods, particularly against compression and signal distortion. Yet, their higher complexity and risk of visual degradation in low-frequency bands often require careful design or hybrid integration for optimal results.

2.1.3 Hybrid Watermarking Methods

Hybrid watermarking methods combine the strengths of spatial and transformation domains, with the aim of increasing resilience, invisibility, and adaptability in complex situations. Commonly used combinations include DWT, DCT, SVD, Schur decomposition, and evolutionary optimizers.

Darwish et al. [60] introduced a DWT–WHT–SVD framework using genetic algorithms to optimize embedding in color images. Alzahrani et al. [61] proposed a medical image watermarking system using DWT, DCT, and SVD on RONI regions with HVS-guided block selection. A dual watermarking scheme by [62] combines DWT and Schur, enhanced with LZW compression and PSBFO optimization, achieving strong capacity and robustness at the cost of computational complexity. Further improvement was explored by [63], combining DWT-DCT-SVD with chaotic maps and PSO for adaptive embedding, although geometric resilience remains limited. A rotation-invariant variant in [64] embeds SVD-based watermarks into DWT-DCT blocks but lacks encryption security.

Yuan et al. [65] proposed a blind scheme that manipulates DC coefficients across RGB channels to balance imperceptibility and robustness without full DCT use. Ernawan et al. [66] applied DWT-DCT on variance-selected blocks with adaptive scaling, showing good results against filtering and compression but sensitivity to noise.

In conclusion, although traditional, transform, and hybrid methods have laid the groundwork for digital watermarking, their reliance on handcrafted designs limits adaptability to modern attack scenarios. This has motivated the shift toward deep learning–based approaches, as well as the exploration of superpixel-guided techniques for more structure-aware embedding, which are discussed in Sections 2.2 and 2.3.

2.2 Deep Learning-Based Watermarking Methods

With the increasing need for robustness, invisibility, and automation in digital image watermarking, deep learning-based approaches have emerged as a powerful alternative to conventional algorithms. These techniques leverage neural network architectures capable of learning abstract feature representations from image data to perform embedding and extraction in a secure and adaptive manner.

2.2.1 DNN-Based Watermarking Methods

Deep neural networks consist of multiple layers of interconnected neurons that perform nonlinear transformations on incoming data. DNNs are capable of learning complex correspondences and hierarchical attributes. This capability makes them ideal for watermarking systems, feature selection

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

optimization, classification-based extraction, or supporting secure and reversible watermarking.

In watermarking (Figure 1), the host image is typically passed to the input layer (and sometimes the watermark signal), while the hidden layers learn embedding rules that balance invisibility and robustness. The output layer is either the inserted watermarked image or the watermark obtained during extraction. This embedded learning model substitutes manual rules with adaptive data-driven strategies.

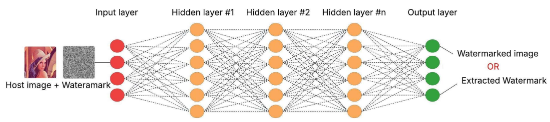


Figure 1 : DNN architecture.

Ge et al. [67] proposed a DNN-based framework for watermarking document images, integrating a noise simulation layer and text-sensitive loss function to limit distortions in textual regions. To address quality degradation, they introduced a strength-adjustment mechanism for preserving visual fidelity. Jaiswal and Pandey [68] developed a DNN model for blind color image watermarking, incorporating integer wavelet transform (IWT) and principal component analysis (PCA) for preprocessing, followed by binary classification during extraction. Sinhal and Ansari [69] applied DNNs to medical images, combining IWT, SHA-256 hashing, and LZW compression to protect region-of-interest (ROI) data while enabling blind watermark recovery.

Battarusetty et al. [70] introduced an optimization-driven method where Squirrel Search and Grey Wolf Optimization algorithms tuned a DNN to identify embedding zones. Wang et al. [71] further expanded the applications of DNNs by developing a reversible watermarking scheme that incorporates copyright protection watermarks directly into DNN models, while preserving their operational performance on diverse datasets.

These contributions illustrate the adaptability of DNNs in diverse watermarking scenarios, offering intelligent and resilient embedding strategies.

DNN-based watermarking typically relies on pre-computed features or transformed representations fed into fully connected layers, whereas CNNs process raw image blocks or pixel grids directly, extracting spatial features through convolution. This makes CNNs more effective for image watermarking, where local structures and texture details are essential for robustness and imperceptibility.

2.2.2 CNN-Based Watermarking Methods

Convolutional Neural Networks are deep learning architectures (Figure 2) specifically tailored for processing image data. Their layered design (comprising convolutional filters, pooling operations, and fully connected neurons) enables robust feature extraction and spatial reasoning in watermarking systems.

Liu et al. [72] introduced a zero-watermarking approach based on CNNs that stylizes host images with logo information using style transfer, followed by Arnold transform encryption. Chacko et al. [73] proposed a watermarking scheme combining CNNs, DCT, superpixel segmentation, and Harris Hawks Optimization to achieve superior PSNR and robustness.

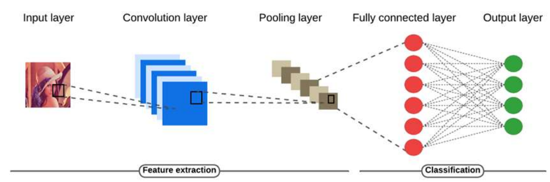


Figure 2: CNN architecture.

Wu et al. [74] embedded fragile watermarks into DWT coefficients and used CNNs for tampering detection and recovery. Singh et al. [75] and Darwish et al. [76] explored encoder–decoder and VGG-based architectures for blind watermarking and chaotic encryption. Other works like Hsu et al. [77] and Mahto et al. [78] improved watermark recovery through PSO-tuned super-resolution CNNs and hybrid-domain embedding.

Advanced CNN strategies include channel and frequency domain attention mechanisms [79], zero watermarking for medical imaging [80], [81], [82], and residual/dense network-based extraction modules [83], [84]. Lightweight models such as MobileNetV2 and NasNet-Mobile have also been adapted for resource-constrained scenarios [85].

These diverse studies reflect CNNs' role as versatile engines for deep watermarking tasks, balancing invisibility, speed, and robustness across practical image applications.

2.2.3 GAN-Based Watermarking Methods

Generative adversarial networks, presented by Goodfellow [7], are a pair of neural networks (generator and discriminator) that challenge each other in a zero-sum game (Figure 3). The generator is trained to insert watermarks into source images in such a way that they are undetectable from the original data, while the discriminator tries to identify irregularities.

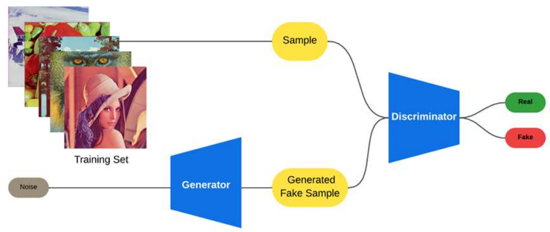


Figure 3: GAN architecture.

This opposing process results in the creation of highly undetectable and robust watermarked images. In watermarking applications, GANs have enabled end-to-end learning of embedding and extraction pipelines. Singh et al. [86] applied GANs in a healthcare context to encode patient identity within medical images, using randomized SVD and RDWT for layered protection. Annadurai et al. [87] integrated DWT, SVD, and GAN segmentation to achieve high resistance to re-encoding and signal processing attacks.

Pang et al. [8] employed GANs and U-Net to enable blind watermarking in open-source datasets, while Huang et al. [88] enhanced watermark visibility using attention-guided fusion layers. Their system also utilized discriminators to differentiate between watermarked and original images.

In addition to classic GANs, recent variants such as CycleGAN [9], WGAN [10], InfoGAN [11], and RivaGAN [12], [13] have also been studied. These structures include elements such as attention masks, feature consistency losses, and simulation layers to train the model against real-world distortions, including JPEG compression, rotation, and cropping.

GANs continue to push the boundaries of image watermarking by facilitating robust, dynamic, and invisible embedding mechanisms adaptable to various threat models.

2.3 Superpixel-Based Watermarking Methods

The concept of superpixels was first introduced by Ren and Malik [89]. A superpixel is a collection of pixels that share similar features such as intensity, texture, color, brightness, and other characteristics (Figure 4). With superpixel segmentation, the image can be divided into a number of similar regions. Unlike fixed-size blocks, superpixels preserve semantic boundaries and local coherence, offering a more intelligent subdivision of the image for region-based processing tasks. Early techniques such as TurboPixels [90], Lattice-Cut [91], and Entropy Rate Superpixel Segmentation [92] were developed to generate compact and uniform superpixel representations. Among these, entropy-based segmentation has proven particularly effective in ensuring spatial uniformity and preserving meaningful structure in natural images.

Figure 4: Images segmented into superpixels using SLIC algorithm.

In digital watermarking, superpixels have been used to improve imperceptibility and robustness, by enabling adaptive embedding aligned with image structures. One prominent approach is the fragile watermarking scheme proposed by Qiao et al. [93], which embeds two types of watermarks per superpixel: (1) an authentication watermark generated via a feedback-based chaotic system embedded within the same superpixel, and (2) a compressed recovery watermark embedded into another region.

Concerning robust watermarking, Niu et al. [94] proposed a method combining Entropy Rate Superpixel Segmentation with the Quaternion Wavelet Transform (QWT). They first segment the image into homogeneous superpixels, then extract feature points using the SIFER detector, and define affine-invariant regions for embedding. Watermark information is inserted by modulating QWT modulus coefficients, resulting in strong resilience against common signal processing operations and local desynchronization attacks such as cropping and geometric transformations.

In another direction, Singh et al. have explored multiple designs for embedding data into superpixel-segmented color images. In [95], they introduce a robust watermarking scheme tailored for compressed images using a combination of superpixel segmentation, DCT, and the Arnold Transform. The image is first converted to YCbCr space, and superpixel segmentation is applied to the Y component. Candidate embedding blocks are selected based on heterogeneity testing, followed by DCT-based quantization for watermark insertion. This method showed strong performance in terms of robustness and imperceptibility under compression.

In a follow-up work, Singh et al. [96] proposed a robust reversible data hiding scheme, again using superpixel segmentation in the Y channel of YCbCr images. Based on block

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classification into homogeneous and heterogeneous types, Cellular Automata (CA) is used for spatial embedding in homogeneous regions, while DCT is applied to heterogeneous regions. The Arnold transform is again employed for random block selection. This hybrid domain method ensures a balance of security, reversibility, and visual quality.

Additional work by Jani and Jani [97] investigates superpixel-based SVD embedding for image authentication. The authors compare segment-based and block-based methods, concluding that superpixels offer lower computational cost and better visual quality. Their method showed potential for use in medical image protection and forensic applications, thanks to its tamper resistance and acceptable PSNR performance.

While promising, superpixel-based methods remain underexplored and often rely on static heuristics. Their integration with adaptive learning remains limited, highlighting a gap addressed in this work.

To facilitate a comparative understanding of the surveyed techniques, and to lay the groundwork for the approach presented later in this paper, Table 1 summarizes the key characteristics of representative watermarking methods across various domains. It highlights the core techniques, strengths, and limitations associated with each approach.

Table 1: Comparative summary of watermarking methods.

Category	Key References (Year)	Techniques	Advantages	Limitations
Spatial Domain	Tanaka (1990) [20], Yeo (2003) [21], Mohammed (2014) [23], Rinki (2022) [25]	LSB, Patchwork, DISB, matrix-based LSB	High capacity, fast	Low robustness, vulnerable to attacks
Transform Domain	Wang (2020) [42], Prajwalasimha (2019) [48], Hsu (2020) [52], Yuan (2021) [54], Zhang (2020) [59]	DCT, DFT, QDFT, DWT, SVD	Robust to compression, good multiresolution	Computational complexity, possible visual distortion
Hybrid Methods	Darwish (2020) [60], Alzahrani, (2021) [61], Ernawan (2021) [66]	DWT-DCT-SVD, Schur, PSO	Balance between robustness and imperceptibility	Complex design, higher processing time
DNN-Based	Ge (2023) [67], Jaiswal (2023) [68], Ansari (2023) [69], Wang (2023) [71]	IWT + PCA, Feature learning, ROI encoding	Adaptive embedding, visual fidelity	Overfitting risk, training data dependency
CNN-Based	Liu (2022) [72], Chacko (2022) [73], Wu (2022) [74], Singh (2024) [75]	CNN + DCT/DWT, Style transfer, Encoder–Decoder	Robust feature extraction, tamper detection	High resource usage, parameter tuning needed
GAN-Based	Singh (2023) [86], Annadurai (2023) [87], Pang (2023) [8], Huang (2023) [88]	GAN, RDWT, U-Net, Attention fusion	High imperceptibility, robust to distortions	Training instability, need for large datasets
Superpixel-Based	Qiao (2014) [93], Niu (2020) [94], Singh (2022) [96], Jani (2017) [97]	Superpixel + DCT/QWT/CA/SVD	Local adaptivity, boundary preservation	Mostly heuristic, lacks deep learning integration

3. PROPOSED SUPERGAN FRAMEWORK

SuperGAN is a hybrid image watermarking framework that integrates superpixel segmentation with GAN-based encoding to embed data within semantically meaningful and visually textured regions of an image.

The overall pipeline comprises four main stages:

1. Superpixel segmentation and region selection, based on entropy and size.
2. GAN-based local embedding, applied selectively within selected regions.
3. Robustness evaluation, through various simulated distortions.
4. Blind watermark extraction and performance assessment, based on standard recovery metrics.

The framework is illustrated in Figure 5. Each step details are discussed in the following subsections.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

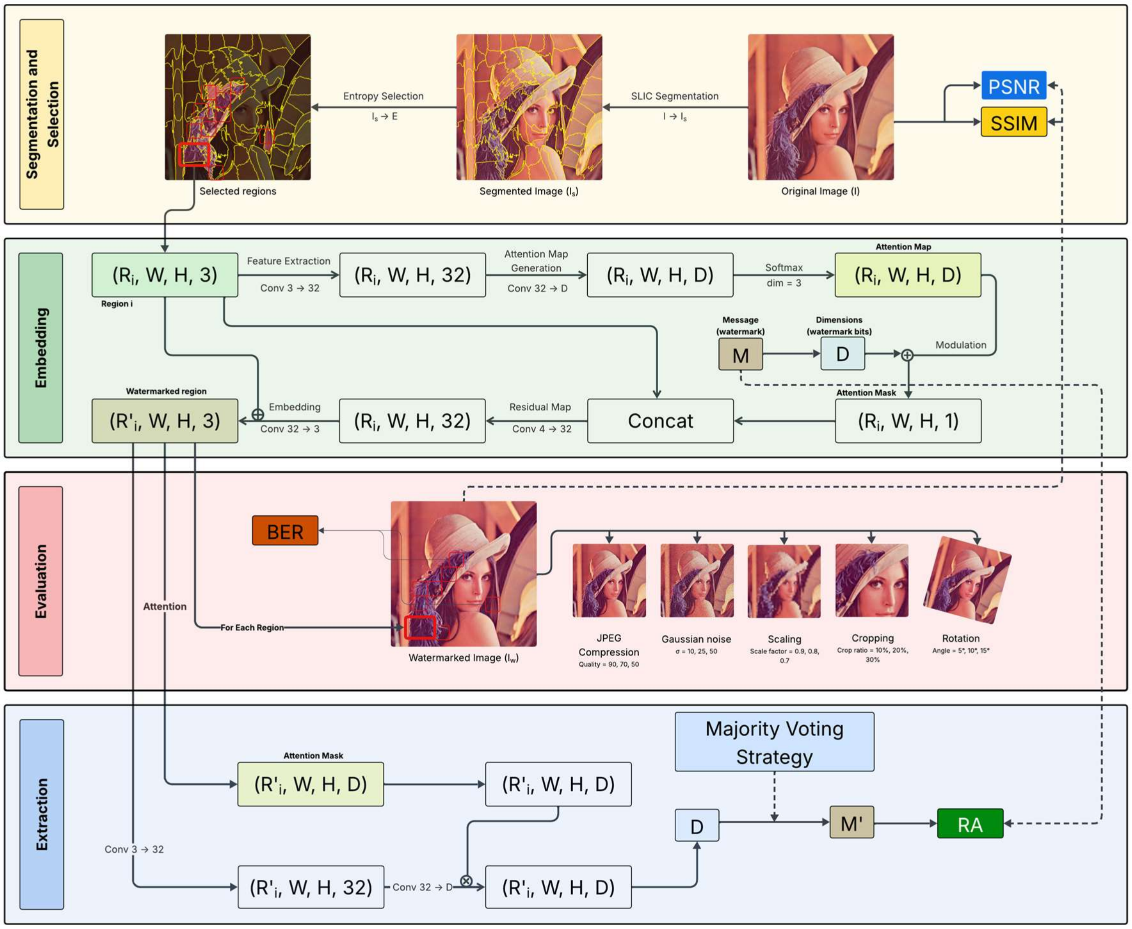


Figure 5: Overview of the SuperGAN framework.

3.1 Content-Aware Superpixel Segmentation and Selection

To guide the watermark embedding to robust and imperceptible areas, SuperGAN begins by segmenting the image into superpixels using the SLIC algorithm. Then, it selects the most suitable regions based on entropy and region size, favoring those with high texture complexity.

3.1.1 Superpixel Segmentation

Compared to fixed grid-based blocks, superpixels more effectively preserve object contours and semantic structures. This makes them highly suitable for watermarking, as changes are less noticeable and more resilient to common image processing operations.

The Simple Linear Iterative Clustering (SLIC) algorithm is employed [98], which clusters pixels in a five-dimensional space (Lab color + spatial coordinates) to form compact, boundary-respecting segments. Each superpixel S_i is defined as:

$$S_i = \{p_j \mid D(p_j, C_i) < D(p_j, C_k),\ \forall k \neq i\} \quad (1)$$

Where:

- p_j is a pixel.
- C_i is the cluster center of superpixel S_i.
- $D(p_j, C_i)$ represents the distance between pixel p_j and the cluster center C_i in the combined Lab color and spatial space.

In order to achieve an adequate balance between segmentation accuracy and embedding efficiency, we conducted empirical tests using three different values for the number of superpixels: 50, 100 and 200. For an image with dimensions H × W, the average area of a superpixel is calculated as follows:

$$A_{sp} = \frac{H \times W}{N} \quad (2)$$

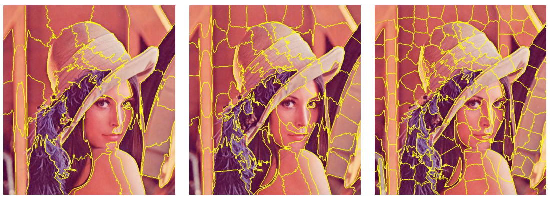
For an image with 512×512 dimensions:
50 superpixels ⟶ $A_{sp} \approx$ 5,242 pixels (≈ 72×72).

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

100 superpixels → A_{sp} ≈ 2,621 pixels (≈ 51×51).
200 superpixels → A_{sp} ≈ 1,310 pixels (≈ 36×36).

A configuration of 100 superpixels is selected as a trade-off between spatial granularity, robustness, and computational efficiency. Figure 6 illustrates an example output of the SLIC algorithm.



(a) (b) (c)

Figure 6: Original image Lena segmented using SLIC with different numbers of superpixels: (a) 50 segments, (b) 100 segments, and (c) 200 segments.

3.1.2 Entropy-Based Region Selection

Once segmentation is complete, SuperGAN evaluates the texture richness and structural stability of each superpixel using two complementary criteria:

- Shannon entropy (equation 3) [99], which measures local intensity variation and favors textured areas,
- Region size, which reflects spatial extent and improves decoding robustness by reducing the risk of deletion during attacks like cropping or compression.

$$E = -\sum_{k=1}^{K} P_k \, log_2\,(P_k) \quad (3)$$

Where:

- P_k is the probability of the intensity level k in the superpixel.
- K is the number of grayscale levels.

SuperGAN adopts a dual-scoring strategy that assigns equal importance (50% weight) to both entropy and size (presented in Section 4.2.3). For each superpixel, a combined score is computed by normalizing both features and averaging them. The top 10 superpixels with the highest combined scores are selected for embedding. This selection provides two benefits:

Imperceptibility, by avoiding flat, sensitive regions;

Robustness, by embedding in textured zones that resist degradation.

This selection strategy naturally categorizes image regions into:

Textured (high-entropy) zones, ideal for mark embedding due to their visual complexity. **Low-entropy regions**, which are avoided to preserve visual quality.

Figure 7. Illustration of region selection prior to embedding. The image shows the original input (Lena) with superpixel boundaries highlighted in yellow. The 10 selected regions, chosen based on a combined score of entropy and region size, are shown at full opacity to emphasize their suitability for robust embedding. Non-selected regions are dimmed to enhance visual focus on the embedding zones.

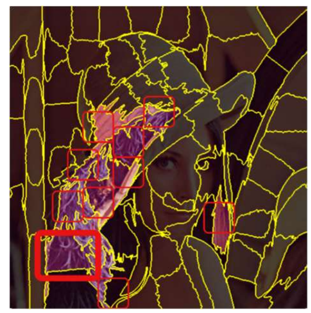


Figure 7: Watermark embedding regions in the Lena image. Yellow contours show SLIC segmentation; red boxes mark the top 10 superpixels selected based on combined entropy and size.

To enhance recovery reliability, the same 32-bit watermark is embedded in all 10 selected regions, forming a redundancy strategy.

3.2 Watermark Embedding

In the SuperGAN framework, the watermark embedding process is inspired by the RivaGAN model, originally developed for robust video watermarking. To adapt this approach to still images, SuperGAN replaces the temporal dimension (frames) with spatially localized regions. Each selected region is treated as a standalone embedding unit, and the watermark is independently embedded using a region-aware encoder.

The objective is to embed a multi-32-bit binary watermark $M \in \{0,1\}^{32}$ into an image. The embedding process follows four main stages: feature extraction, attention map generation, watermark modulation, and residual embedding.

3.2.1 Feature Extraction

Each selected region $R \in \mathbb{R}^{H\times W\times 3}$ is passed through a convolutional block to extract low-level visual features:

$$F = Conv_{3\to 32}(R)$$

This operation outputs a feature tensor $F \in \mathbb{R}^{H\times W\times 32}$, where each channel captures specific texture or structure characteristics of the region.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

These features guide the embedding process by highlighting areas more tolerant to visual alterations.

3.2.2 Attention Map Generation

The attention mechanism enables the encoder to focus on different watermark bits depending on the local image content. It generates an attention tensor $A \in \mathbb{R}^{H \times W \times D}$, where $D = 32$ corresponds to the length of the watermark. An illustrative example of the attention tensor structure is shown in Figure 8.

The attention map is computed using two convolutional layers:

$$a = Conv_{32 \to D}(F)$$

A $softmax$ function is then applied to the raw attention scores across the bit dimension, converting them into a probability distribution. This gives the ability for each pixel (x,y) to assign a weight to each bit of the watermark, showing how much attention, it pays to bit d compared to the total D=32:

$$A(x, y, d) = \frac{e^{a(x,y,d)}}{\sum_{k=1}^{D} e^{a(x,y,k)}} \quad (4)$$

$A\,(x\,, y\,, d)$ represents the attention weight assigned by pixel $(x\,, y)$ to the d^{th} watermark bit.

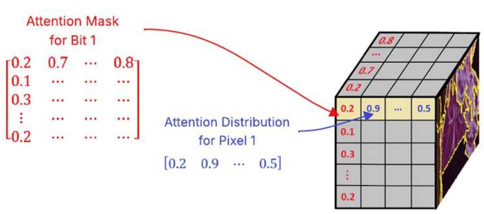


Figure 8: Attention tensor showing bit-wise probability distribution per pixel.

3.2.3 Watermark Modulation

Given a binary watermark $M \in \{0, 1\}^{32}$, we compute a compact watermark value at each pixel using:

$$W_c(x, y) = \sum_{d=1}^{D} A(x, y, d)\,.\,M[d] \quad (5)$$

This yields a smooth watermark tensor $W_c \in \mathbb{R}^{H \times W \times 1}$ adapted to the visual structure of the region. The compact watermark signal is derived by weighting each bit of the binary watermark according to the learned attention map, yielding a smooth, region-specific intensity profile per pixel, as formalized in Equation (5).

3.2.4 Residual Generation and Embedding

The compact watermark W_c is concatenated with the original region R, and the combined tensor is processed by a second convolutional block to generate a residual map $\Delta R \in \mathbb{R}^{H \times W \times 3}$:

$$\Delta R = \tanh\Big(ConvLayers\big(Concat(R, W_c)\big)\Big) \quad (6)$$

To maintain imperceptibility, the residual is scaled by a small constant (0.01):

$$\Delta R'(x, y) = 0.01\,.\,\Delta R(x, y) \quad (7)$$

Finally, the watermarked region R’ is obtained by adding the scaled residue to the initial region by adjusting its RGB channels.

$$R'(x, y) = R(x, y) + \Delta R'(x, y) \quad (8)$$

3.3 Attacks and Transformations

To evaluate the robustness of our SuperGAN method, we apply a series of common image processing and geometric attacks to the watermarked images. These transformations aim to simulate real-world degradation scenarios and test the decoder's ability to correctly recover the watermarked image under different degrees of distortion.

3.3.1 JPEG Compression

Compression is applied using the JPEG algorithm at three quality levels: 90, 70, and 50. Lower quality levels introduce stronger quantization and lead to loss of high-frequency details, making watermark recovery more challenging.

3.3.2 Additive Gaussian Noise

Gaussian white noise is added to simulate sensor or transmission noise. We test with zero-mean noise and standard deviations $\sigma \in \{10, 25, 50\}$ added independently to each RGB channel. The noisy image I_n is generated as:

$$I_n(x, y) = I_w(x, y) + N(0, \sigma^2) \quad (9)$$

Where:

- $I_w(x, y)$ is the value of the watermarked image at pixel (x, y).
- $I_n(x, y)$ is the resulting pixel value after noise is added.
- $N(0, \sigma^2)$ A random value from a normal (Gaussian) distribution with: Mean = 0 and $Variance = \sigma^2$.

3.3.3 Geometric Scaling

This attack evaluates the robustness of the watermark against changes in image resolution. The watermarked image is first downscaled by a factor $s \in \{0.9, 0.8, 0.7\}$, then rescaled back to its original size using bicubic interpolation. This process

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

introduces interpolation artifacts and potential signal degradation:

$$I_s = Resize\left(Resize(I_w, s), \frac{1}{s}\right) \tag{10}$$

I_w is the watermarked image, and I_s is the result after scaling. The test simulates real-world scenarios like display resizing or transmission rescaling.

3.3.4 Cropping

Center cropping is used to simulate partial occlusion. We adjust the image to keep only a central part, preserving 90%, 80%, or 70% of the initial dimensions. This method could potentially omit some embedded superpixels, thus evaluating the proficiency of the redundant strategy in preserving the integrity of the watermark despite partial obstruction.

3.3.5 Rotation

Mild geometric distortion is tested using image rotation with angles $\theta \in \{5°, 10°, 15°\}$. The image is rotated around its center using affine transformation and interpolated back to the original size. This tests both spatial robustness and decoder tolerance to misalignment.

Table 2: Summary of Attack Scenarios.

Attack Type	Parameter Values
JPEG Compression	90, 70 and 50
Gaussian Noise	10, 25 and 50
Scaling	0.9, 0.8 and 0.7
Cropping	10%, 20% and 30%
Rotation	5°, 10° and 15°

Each attack is applied independently, and the watermark is decoded afterward using the trained decoder.

3.4 Watermark Extraction and Evaluation

The final stage of the SuperGAN pipeline involves recovering the embedded watermark from the watermarked image, possibly after transformations, and evaluating the performance of the overall system. This section outlines the decoding strategy and defines the performance metrics used to assess both imperceptibility and robustness. All evaluations are performed in a blind setting, where the original image is not required during extraction.

3.4.1 Decoder Architecture and Attention-Pooling Mechanism

The decoder architecture is adapted from the RivaGAN framework and mirrors the encoder structure in reverse. It is designed to recover a 32-bit binary watermark vector $M \in \{0,1\}^{32}$ from a watermarked superpixel region $R' \in \mathbb{R}^{H \times W \times 3}$. Convolutional layers are used to extract visual features from the region, followed by an attention-guided pooling mechanism that facilitates the reconstruction of the embedded watermark.

The decoding process begins with two convolutional layers that generate pixel-level scores for each watermark bit. These scores are modulated using the same attention map structure learned during embedding. The final step aggregates the modulated predictions spatially to form a compact watermark estimate:

$$\begin{gathered} a = Conv_{3\to 32}(R') \\ b = Conv_{32\to D}(a) \\ T = Attention(R') \in \mathbb{R}^{H \times W \times D} \\ c = T \odot b \\ M' = AvgPool(c) \end{gathered} \tag{11}$$

Where, T is the attention tensor reused from the encoder, $\odot$ denotes element-wise multiplication, and $M' \in \{0,1\}^{32}$ is the soft prediction of the watermark bits.

3.4.2 Binary Decision and Redundancy Fusion

Each superpixel processed by the decoder provides a continuous-valued prediction for each bit. To obtain a binary watermark vector, a threshold is applied:

$$M'[i] = \begin{cases} 1, & if\ M'[i] \geq 0.5 \\ 0, & otherwise \end{cases} \tag{12}$$

Since the watermark is redundantly embedded across the top $K = 10$ selected superpixels, the decoding process is repeated independently for each region. This results in 10 predicted binary watermark vectors $M'_1, M'_2, \dots, M'_{10}$ (Figure 9).

To enhance reliability and compensate for potential errors in some regions (due to noise, cropping, etc.), a bit-wise majority voting strategy is employed. The final decoded watermark $M'_{final}[i]$ is computed as:

$$M'_{final}[i] = \begin{cases} 1, & if\ \sum_{k=1}^{K} M'_k[i] \geq \frac{K}{2} \\ 0, & otherwise \end{cases} \tag{13}$$

This redundancy-based fusion improves robustness by aggregating predictions across multiple spatial locations.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

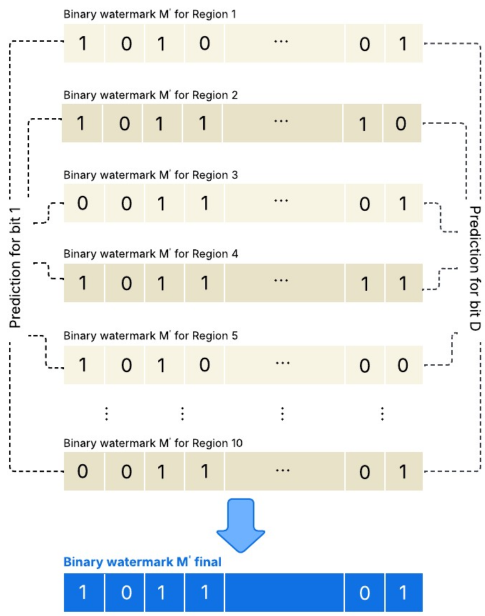


Figure 9: Redundancy fusion from 10 superpixel-based predictions. Rows show thresholded binary vectors; bottom row (blue) illustrates the final majority-voted watermark.

3.4.3 Evaluation Metrics

To assess the performance of SuperGAN, we adopt four widely used metrics that jointly evaluate visual fidelity and the accuracy of watermark extraction. These metrics are computed under two distinct conditions: without any attacks, to establish a performance baseline, and after applying various distortions, to test robustness.

Peak Signal-to-Noise Ratio (PSNR)

PSNR [100] is a classical metric used to quantify the visual distortion introduced during watermark embedding. It is defined as:

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right) \qquad (14)$$

Where:

- MAX_I is the maximum possible pixel value.
- MSE is the mean squared error between the original and watermarked image.

A higher PSNR indicates lower distortion, implying better imperceptibility. In this study, PSNR is calculated both before and after the image undergoes attacks.

Structural Similarity Index Measure (SSIM)

SSIM [100] provides a perceptual quality assessment by comparing structural information between the original image I and the watermarked version I_w . It is expressed as:

$$SSIM(I, I_w) = \frac{(2\mu_I\mu_{I_w} + c_1)(2\sigma_{I\,I_w} + c_2)}{(\mu_I^2 + \mu_{I_w}^2 + c_1)(\sigma_I^2 + \sigma_{I_w}^2 + c_2)} \qquad (15)$$

Where:

- μ_I, μ_{I_w}: mean intensities of the original and watermarked images.
- $\sigma_I^2 + \sigma_{I_w}^2$: their variances.
- $\sigma_{I\,I_w}$: their covariance.
- c_1, c_2: small constants to stabilize the division.

Bit Error Rate (BER)

BER quantifies the proportion of incorrect bits in the recovered watermark compared to the original. It is defined as:

$$BER = \frac{1}{D}\sum_{i=1}^{D}\left|M'_{final}[i] - M[i]\right| \qquad (16)$$

In our case, D represents the watermark dimension (32 bits), M refers to the original laughable thread, and M'_{final} is the decrypted version following post-processing or an attack.

A lower BER indicates greater resilience and decoding accuracy.
A lower BER indicates better resilience and decoding accuracy.

Recovery Accuracy (RA)

To provide a more intuitive interpretation of BER, we also report the recovery accuracy, calculated as:

$$RA = (1 - BER) \times 100\% \qquad (17)$$

Collectively, these metrics provide a balanced evaluation of both visual fidelity and watermark robustness, under pristine and distorted conditions.

The experimental results and comparative analysis under various attack conditions are discussed in the following section.

4. EXPERIMENTS AND RESULTS

This section presents a comprehensive experimental evaluation of the proposed SuperGAN watermarking framework. The analysis is divided into three parts: (1) baseline evaluation under ideal conditions, (2) evaluation under attack scenarios, and (3) comparative analysis with benchmark methods. Each part provides both quantitative metrics and interpretive insights to show system performance.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1 Dataset and Experimental Setup

All experiments in this paper were conducted using images from the COCO dataset, which consists of 5,000 images. This dataset was chosen due to its diversity of everyday scenes, reflecting the variety and complexity of real-world applications.

For baseline evaluation under ideal conditions (i.e., without attacks), all 5,000 images were normalized to 512 × 512 pixels and segmented into 10 superpixel regions.

To evaluate robustness under different attack scenarios, a representative subset of 500 images was randomly drawn from the same pool. These were used for comparative analysis against three reference methods, ensuring that all approaches were tested on identical data for fairness and consistency.

All experiments were performed on a regular workstation equipped with a mid-range NVIDIA GPU and between 16 to 32 GB of RAM. The implementation was developed in Python 3.11. The model inference was handled using TensorFlow, while OpenCV and scikit-image supported preprocessing and transformations. Data handling and statistical computations were performed using NumPy and pandas.

To provide a visual sense of the data used, Figure 10 displays a sample of 30 randomly selected images from the COCO dataset. This selection highlights the range of scene types and image characteristics present throughout the evaluation.

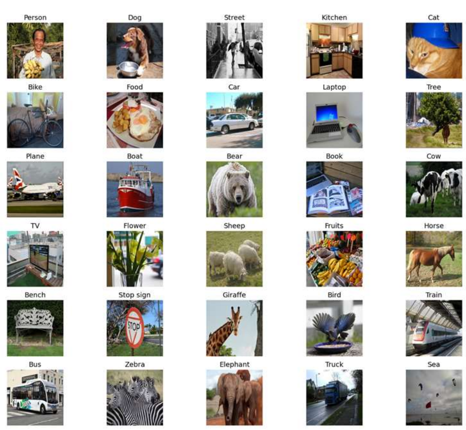


Figure 10: Examples of randomly selected COCO val2017 images used in the experiments.

4.2 Baseline Evaluation

Before introducing image distortions, SuperGAN was tested under ideal conditions using 5,000 images from the COCO dataset. Table 3 summarizes the baseline results obtained on 5,000 images, showing consistently high recovery accuracy alongside strong imperceptibility. The following subsections analyze these results in detail, focusing first on imperceptibility and then on robustness.

Table 3: Baseline performance metrics.

	RA (%)	BER	PSNR (dB)	SSIM
Mean	98.98	17.65	40.47	0.98
Std Dev	6.74	17.55	0.22	0.01
Min	0	0	39.93	0.92
Q1(25%)	100	4.17	40.39	0.97
Median	100	12.5	40.43	0.98
Q3(75%)	100	25	40.49	0.99
Max	100	100	45.81	1

4.2.1 Imperceptibility Analysis

Perceptually, SuperGAN embedding has a negligible impact on visual quality. Figure 11 shows that PSNR values are highly clustered around 40.47 dB, with minimal fluctuation within the data. Furthermore, SSIM values (Figure 12) are around 0.98, indicating that structure and contrast are well maintained across all images.

Most images scored above 0.97 in SSIM and fell within a tight PSNR band (40.39–40.49 dB), This confirms the suitability of the SuperGAN for real-world use cases where imperceptibility is essential.

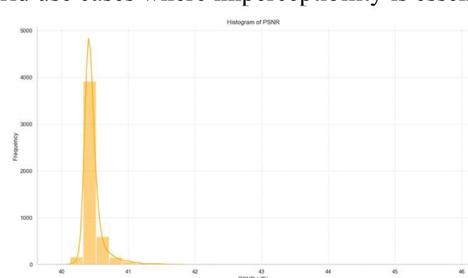


Figure 11: PSNR histogram for baseline performance.

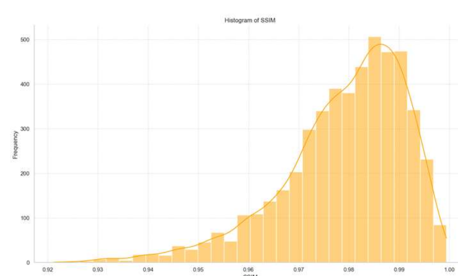


Figure 12: SSIM histogram for baseline performance.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2.2 Robustness Analysis

On the global scale, the RA metric reveals excellent performance: over 98% of the images achieved full recovery (RA = 100%), as seen in Figure 13. Only a few images showed partial recovery, suggesting rare edge cases rather than systemic flaws.

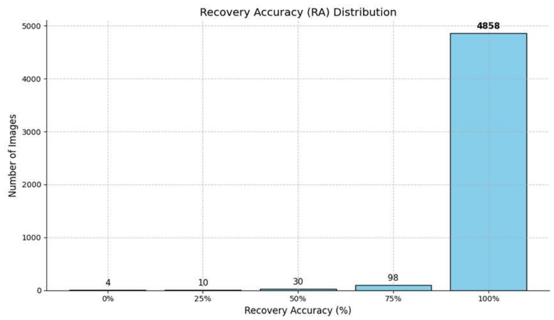


Figure 13: Distribution of RA values computed over 5,000 watermarked images.

Locally, the BER analysis provides finer insight. Out of 50,000 embedded superpixel regions (10 per image), almost half failed extraction (BER = -1), but more than 30% were decoded perfectly (BER = 0), as shown in Figure 14. This duality highlights a critical property of SuperGAN: the method is globally robust, but recovery varies at the superpixel level.

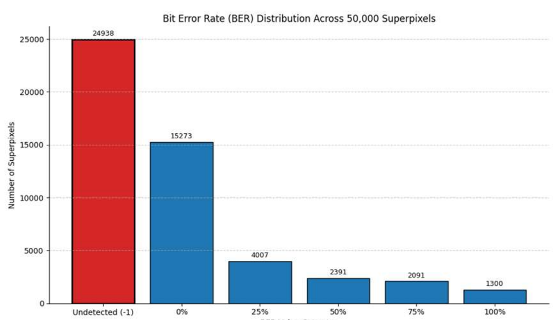


Figure 14: Distribution of BER values across all superpixel regions (10 per image).

Figure 15 offers spatial insight with a BER heatmap over 25 sampled images. It becomes evident that certain regions are more extraction-friendly (likely those with high entropy or stable texture) while others are more sensitive. This spatial variability reinforces the importance of strategic superpixel selection and possibly calls for added redundancy in future versions.

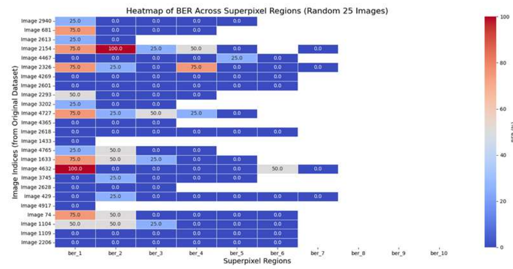


Figure 15: Heatmap of BER across the 10 embedded superpixels in 25 randomly sampled images.

4.2.3 Analysis of Entropy and Region Size Impact on BER

To better understand the factors that influence local decoding success, we conducted a statistical correlation analysis between the Bit Error Rate (BER) and two key regional characteristics: entropy and size. This investigation is restricted to successfully decoded superpixel regions (i.e., those with BER values in the [0–100%] range), totaling 1,500 regions extracted from 150 images. Failed regions with BER = –1 were excluded to ensure a focused and meaningful interpretation.

Figure 16 presents a scatter plot of entropy versus BER, with color-mapped points and a linear regression line superimposed. A weak but observable negative correlation is evident: higher-entropy regions tend to produce lower BER values. The fitted regression yields an R^2 of 0.18, indicating that entropy accounts for a modest portion of the variance in BER. Nevertheless, the trend suggests that regions with richer texture or variability are more suitable for robust watermark embedding.

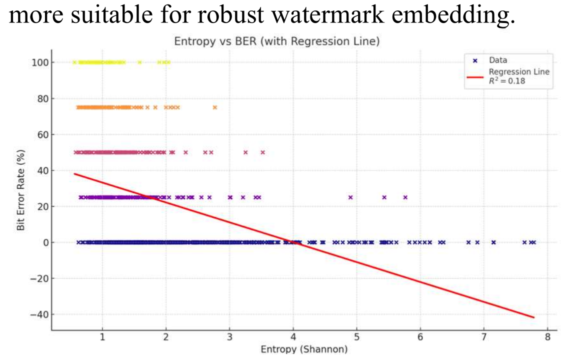


Figure 16: Entropy cs BER (with Regression Line).

Similarly, Figure 17 explores the relationship between region size and BER. While the overall spread is wider and the R^2 value is slightly lower ($R^2 = 0.14$), the trend remains consistent: larger regions are generally more resilient, exhibiting lower BER values. This is expected, as larger areas provide more capacity and tolerance to distortion or encoding

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

errors.

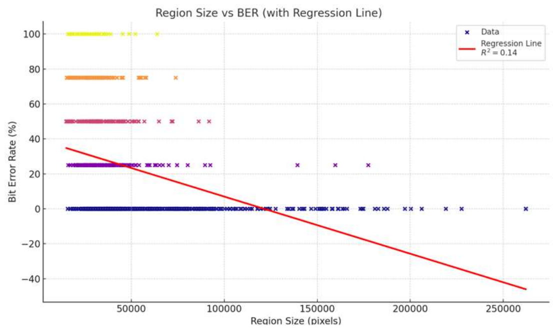


Figure 17: Region Size vs BER (with Regression Line).

Finally, to capture the joint effect of entropy and size on BER, a 3D scatter plot was constructed (Figure 18). Each point in this space represents a superpixel region, plotted with its corresponding entropy, size, and BER. The color scale encodes the BER, with cooler tones indicating low error and warmer tones indicating higher error. The distribution suggests that optimal embedding conditions (low BER) are more likely when both entropy and region size are relatively high. Conversely, compact and low-entropy regions are more vulnerable to decoding errors.

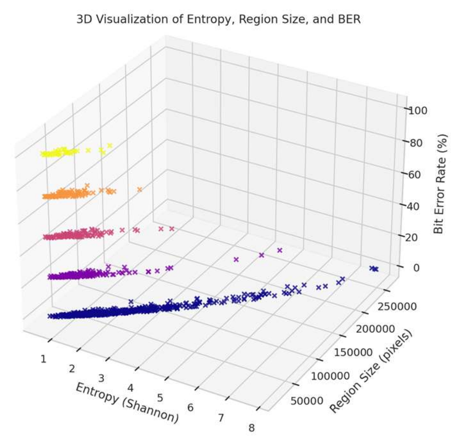


Figure 18: 3D Visualization of Entropy, Region Size and BER.

In summary, SuperGAN delivers impressive performance under optimal conditions. It maintains high visual fidelity while ensuring near-perfect recovery at the image scale. However, localized failures in some regions with the difficulty of finding at least ten robust regions in the image suggest that it may be possible to refine the embedding strategy by improving region diversity or adapting redundancy to the variability of image textures.

4.3 Evaluation Under Attack Scenarios

To assess SuperGAN's resilience, we subjected it to five commonly encountered distortions (JPEG compression, Gaussian noise, scaling, cropping, and rotation) each tested at three severity levels. For each case, we analyze recovery accuracy (RA), PSNR, and SSIM, while referring to both summary tables and the corresponding figures for graphical insight.

4.3.1 JPEG Compression

As shown in Table 4, SuperGAN maintained strong robustness under JPEG compression, with a near-perfect median RA (100%) up to quality 70. Even at quality 50, watermark recovery was still feasible in over half of the cases (median RA = 50%). This stability is mirrored in Figure 19, where performance degrades gradually.

In terms of visual quality, the PSNR and SSIM medians decline modestly as compression increases (Figures 20 and 21) but remain within acceptable limits (PSNR > 33 dB, SSIM > 0.93) for quality levels 90 and 70. The wide value distribution at quality 50 highlights variability in how compression artifacts impact embedding regions.

These results suggest that the entropy-based superpixel embedding preserves watermark integrity even when quantization removes high-frequency image details.

Table 4: JPEG Compression Results Summary.

Metric	Statistic	Quality 90	Quality 70	Quality 50
RA (%)	**Mean**	**97.75**	**93.25**	**75.55**
	Min	0.0	0.0	0.0
	Median	100.0	100.0	50.0
	Max	100.0	100.0	100.0
PSNR (dB)	**Mean**	**36.15**	**33.06**	**31.65**
	Min	27.46	25.34	23.95
	Median	36.23	33.12	31.65
	Max	42.57	39.93	38.76
SSIM	**Mean**	**0.965**	**0.935**	**0.914**
	Min	0.869	0.815	0.784
	Median	0.966	0.934	0.914
	Max	0.998	0.997	0.996

4.3.2 Gaussian Noise

Table 5 reveals that SuperGAN remains resilient under light noise (variance 10), maintaining high RA (mean = 97.85%) and acceptable PSNR/SSIM scores (Figure 20 and 21). However, the system's robustness drops sharply at variance 25 and collapses at 50, where the median RA falls to just 25%, and visual metrics deteriorate (Figure 19).

Notably, even with heavy noise, some images still retain detectable watermarks. This uneven degradation suggests the effectiveness of embedding in certain superpixels over others, likely those with strong entropy structures resistant to noise.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5: Gaussian Noise Results Summary.

Metric	Statistic	Variance 10	Variance 25	Variance 50
RA (%)	**Mean**	**97.85**	**80.95**	**34.70**
	Min	0.0	0.0	0.0
	Median	100.0	100.0	25.0
	Max	100.0	100.0	100.0
PSNR (dB)	**Mean**	**28.35**	**20.68**	**15.17**
	Min	28.11	20.18	14.37
	Median	28.25	20.60	15.12
	Max	30.88	22.96	16.98
SSIM	**Mean**	**0.784**	**0.490**	**0.270**
	Min	0.435	0.129	0.044
	Median	0.794	0.483	0.255
	Max	0.979	0.901	0.730

4.3.3 Scaling

Under geometric scaling (Table 6), SuperGAN shows remarkable consistency. Median RA remains at 100% across all levels, with PSNR and SSIM decreasing slightly but staying well above quality thresholds (Figures 19–21).

Table 6: Scaling Results Summary.

Metric	Statistic	Scale 0.9	Scale 0.8	Scale 0.7
RA (%)	**Mean**	**98.35**	**98.30**	**98.30**
	Min	0.0	0.0	0.0
	Median	100.0	100.0	100.0
	Max	100.0	100.0	100.0
PSNR (dB)	**Mean**	**34.29**	**33.39**	**32.40**
	Min	23.91	23.08	22.11
	Median	34.08	33.16	32.17
	Max	47.87	47.17	46.15
SSIM	**Mean**	**0.965**	**0.958**	**0.949**
	Min	0.835	0.804	0.766
	Median	0.972	0.966	0.959
	Max	0.999	0.999	0.999

This robustness can be attributed to two factors: the use of bicubic interpolation, which preserves local texture, and the spatial redundancy of superpixel-based embedding. Unlike pixel-based schemes, SuperGAN benefits from having watermark bits distributed across well-structured regions that scale uniformly.

4.3.4 Cropping

Cropping presents a harsher challenge. As shown in Table 7 and Figures 19–21, even a modest 10% crop drops the average RA to 91%, and aggressive cropping (30%) leads to a drastic fall (RA = 5.05%). PSNR and SSIM also plunge, especially as more of the imageand with it, embedded regions—are removed.

The sharp decline illustrates a key limitation: if a cropped region eliminates several high entropy superpixels, the decoder fails to reconstruct the watermark. This confirms that while entropy selection aids embedding, spatial redundancy across the image may need to be improved to boost cropping resilience.

Table 7: Cropping Results Summary.

Metric	Statistic	10% Cropping	20% Cropping	30% Cropping
RA (%)	**Mean**	**91**	**25.95**	**5.05**
	Min	0.0	0.0	0.0
	Median	100.0	25.0	0.0
	Max	100.0	100.0	50.0
PSNR (dB)	**Mean**	**12.95**	**11.23**	**10.39**
	Min	7.63	5.63	5.15
	Median	12.61	10.81	10.06
	Max	24.55	22.55	20.75
SSIM	**Mean**	**0.501**	**0.411**	**0.355**
	Min	0.137	0.103	0.013
	Median	0.497	0.393	0.328
	Max	0.959	0.937	0.905

4.3.5 Rotation

Performance under rotation follows a similar decline pattern. Table 8 and Figures 19–21 show that light rotations (5°) are well tolerated (RA = 95.8%), but at 15°, the recovery rate drops below 40%, with PSNR and SSIM showing the worst overall values across all transformations.

This suggests that even small geometric misalignments disrupt the attention-based decoder, likely due to altered spatial alignment of embedded superpixels. A promising direction for future work is to integrate rotation-invariant or alignment-aware modules during decoding.

Table 8: Rotation Results Summary.

Metric	Statistic	5° Rotation	10° Rotation	15° Rotation
RA (%)	**Mean**	**95.8**	**82.85**	**39.45**
	Min	0.0	0.0	0.0
	Median	100.0	100.0	50.0
	Max	100.0	100.0	100.0
PSNR (dB)	**Mean**	**11.76**	**9.93**	**9.05**
	Min	7.54	5.49	4.42
	Median	11.46	9.71	8.86
	Max	19.22	17.20	16.18
SSIM	**Mean**	**0.494**	**0.381**	**0.313**
	Min	0.134	0.100	0.083
	Median	0.502	0.382	0.309
	Max	0.835	0.758	0.733

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

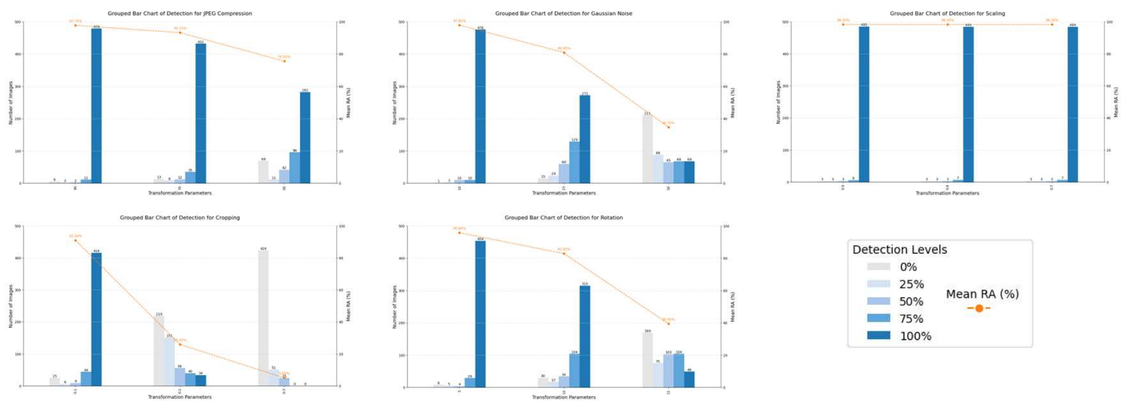


Figure 19: Detection levels and mean Recovery Accuracy (RA) under all image transformations.

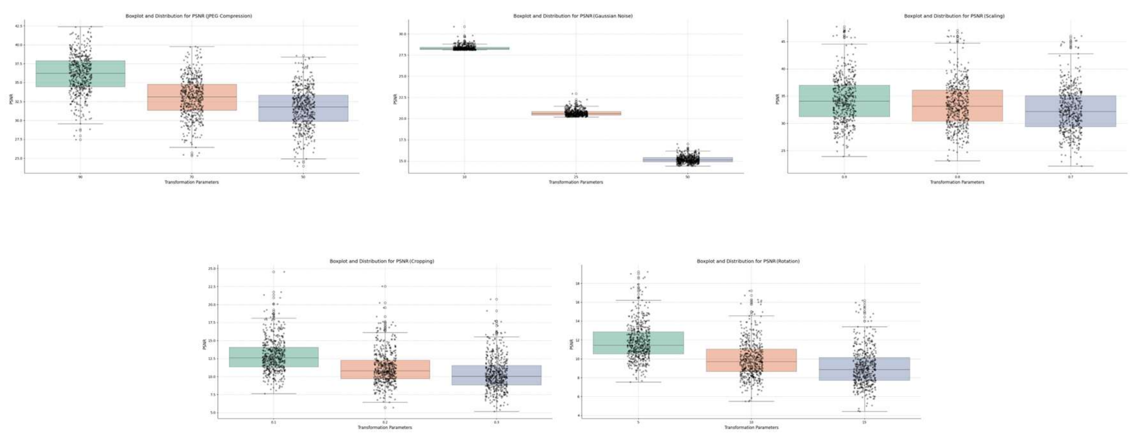


Figure 20: PSNR distribution and boxplots across all image transformations.

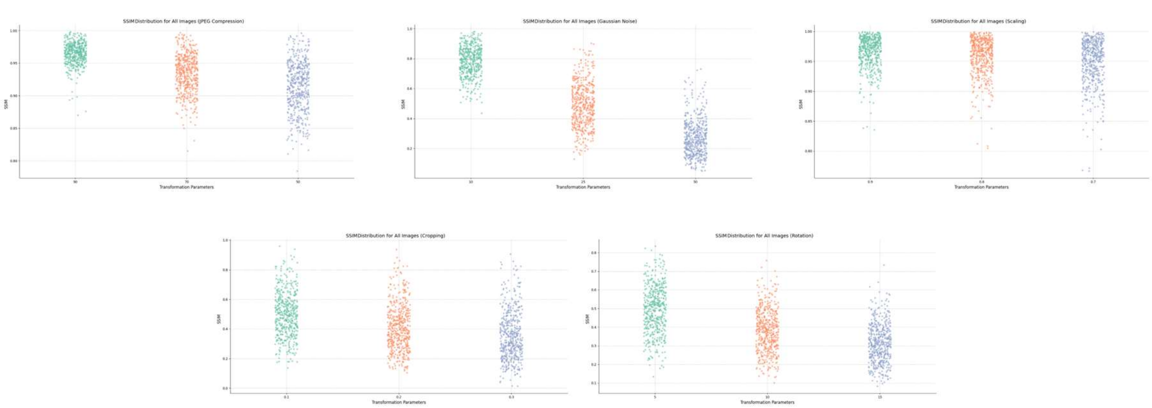


Figure 21: SSIM distribution across all transformations.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.4 Comparative Analysis

To further evaluate the robustness of the proposed SuperGAN framework, a comparative evaluation was conducted against two training-free methods DWT-DCT, and DWT-DCT-SVD [101] and a pre-trained GAN-based watermarking model, RivaGAN adapted for image [88][89]. Each method was tested on the same subset of 500 images, under the same five commonly used transformations.

Figure 22 presents the evolution of RA across severity levels for each transformation type. SuperGAN consistently achieved high recovery performance, particularly under compression and geometric distortions. Even when the image quality was significantly degraded, SuperGAN retained high RA values, especially compared to traditional methods like DWT-DCT.

Under JPEG compression and scaling, SuperGAN demonstrated remarkable stability. It maintained full watermark recovery in most cases, even at higher compression levels or scale reductions. In contrast, RivaGAN showed a more noticeable drop in performance as distortions increased, and DWT-based methods often failed entirely under the same conditions.

The effect of Gaussian noise was more pronounced. While SuperGAN handled low noise variance well, performance dropped under heavy noise (variance = 50). However, this decline was more gradual compared to the sharp collapses observed in the other methods. A similar pattern appeared under cropping and rotation: although all methods suffered under severe conditions, SuperGAN consistently outperformed the others, preserving recoverability in many cases where others did not.

Across all transformations, PSNR values were around 30 dB or higher, while SSIM values mostly exceeded 0.9. This indicates that the robustness benefits of SuperGAN were not achieved at the cost of image quality.

To consolidate these results, Table 9 provides a summary of RA statistics (mean, lower quartile, median, and maximum) for each method and attack type. These values highlight SuperGAN's overall consistency and resilience under diverse attack scenarios, reinforcing the benefits of combining semantic embedding and entropy-based region selection.

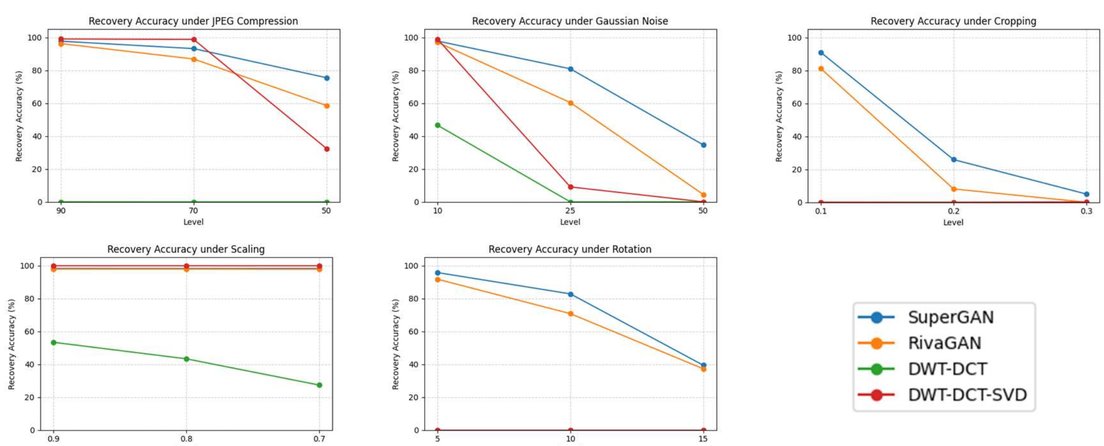


Figure 22: Comparison of Recovery Accuracy (RA) under all transformations.

Journal of Theoretical and Applied Information Technology
15th November 2025. Vol.103. No.21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 9: Summary of RA, PSNR, and SSIM across all transformations and methods.

Metric	Statistic	Method	JPEG Compression			Gaussian Noise			Scaling			Cropping			Rotation		
		Level	90	70	50	10	25	50	0.9	0.8	0.7	0.1	0.2	0.3	5	10	15
RA (%)	Mean	**SuperGAN**	**97,75**	**93,25**	**75,55**	**97,85**	**80,95**	**34,7**	**98,35**	**98,3**	**98,3**	**91**	**25,95**	**5,05**	**95,8**	**82,85**	**39,45**
		RivaGAN	96,31	86,93	58,63	97,11	60,43	4,54	97,95	97,95	97,85	81,19	8,13	0,05	91,77	70,81	37,28
		DWT-DCT	0,05	0	0	46,65	0	0,05	53,4	43,35	27,4	0	0	0,05	0	0	0
		DWT-DCT-SVD	99,1	98,85	32,5	99,05	9,1	0,05	99,95	99,95	99,95	0,05	0	0,25	0	0	0
	Min	**SuperGAN**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
		RivaGAN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		DWT-DCT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		DWT-DCT-SVD	0	0	0	0	0	0	75	75	75	0	0	0	0	0	0
	Q1	**SuperGAN**	**100**	**100**	**75**	**100**	**75**	**0**	**100**	**100**	**100**	**100**	**0**	**0**	**100**	**75**	**0**
		RivaGAN	100	75	50	100	0	0	100	100	100	75	0	0	100	50	25
		DWT-DCT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		DWT-DCT-SVD	100	100	0	100	0	0	100	100	100	0	0	0	0	0	0
	Median	**SuperGAN**	**100**	**100**	**100**	**100**	**100**	**25**	**100**	**100**	**100**	**100**	**25**	**0**	**100**	**100**	**50**
		RivaGAN	100	100	50	100	75	0	100	100	100	100	0	0	100	75	25
		DWT-DCT	0	0	0	50	0	0	75	50	0	0	0	0	0	0	0
		DWT-DCT-SVD	100	100	25	100	0	0	100	100	100	0	0	0	0	0	0
	Q3	**SuperGAN**	**100**	**100**	**100**	**100**	**100**	**75**	**100**	**100**	**100**	**100**	**50**	**0**	**100**	**100**	**75**
		RivaGAN	100	100	75	100	100	0	100	100	100	100	0	0	100	100	50
		DWT-DCT	0	0	0	100	0	0	100	100	50	0	0	0	0	0	0
		DWT-DCT-SVD	100	100	50	100	0	0	100	100	100	0	0	0	0	0	0
	Max	**SuperGAN**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**50**	**100**	**100**	**100**
		RivaGAN	100	100	100	100	100	100	100	100	100	100	100	25	100	100	100
		DWT-DCT	25	0	0	100	0	25	100	100	100	0	0	25	0	0	0
		DWT-DCT-SVD	100	100	100	100	100	25	100	100	100	25	0	25	0	0	0
PSNR (dB)	Mean	**SuperGAN**	**36,05**	**32,98**	**31,58**	**28,3**	**20,64**	**15,15**	**34,22**	**33,32**	**32,33**	**12,93**	**11,21**	**10,37**	**11,75**	**9,91**	**9,04**
		RivaGAN	35,33	32,78	31,51	28,35	20,68	15,17	33,68	32,78	31,83	12,96	11,24	10,39	11,76	9,91	9,03
		DWT-DCT	36,18	33,28	31,83	28,35	20,68	15,18	33,83	32,93	31,98	12,96	11,23	10,39	11,77	9,92	9,04
		DWT-DCT-SVD	36,25	33,11	31,76	28,25	20,6	15,12	34,09	33,16	32,17	12,62	10,83	10,06	11,46	9,71	8,86
SSIM	Mean	**SuperGAN**	**0,96**	**0,93**	**0,91**	**0,78**	**0,49**	**0,27**	**0,97**	**0,96**	**0,95**	**0,5**	**0,41**	**0,35**	**0,49**	**0,38**	**0,31**
		RivaGAN	0,96	0,93	0,91	0,78	0,49	0,27	0,96	0,96	0,95	0,5	0,41	0,35	0,49	0,38	0,31
		DWT-DCT	0,96	0,93	0,91	0,78	0,49	0,27	0,96	0,95	0,94	0,5	0,41	0,36	0,5	0,38	0,31
		DWT-DCT-SVD	0,97	0,94	0,92	0,78	0,49	0,27	0,96	0,96	0,95	0,51	0,42	0,36	0,5	0,39	0,32

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION

This work introduces SuperGAN, a hybrid image watermarking framework that combines superpixel-based region selection with GAN-driven embedding. Embedding is carried out by first segmenting the host image into superpixels, then selecting the most suitable regions according to entropy and size. Within these regions, an attention-guided GAN encoder inserts the watermark in a content-aware manner, which helps preserve imperceptibility while improving resilience to distortions. On the extraction side, the decoder leverages the same attention mechanism to locate and retrieve the watermark, while spatial redundancy across multiple superpixels ensures that recovery remains reliable even if some regions are altered or lost. Evaluations demonstrate that SuperGAN outperforms baseline techniques in invisibility and robustness, as measured by PSNR, SSIM, and retrieval accuracy. While promising for copyright enforcement and secure image transmission, future work could explore dynamic content control and semantic image analysis to further optimize watermark placement and improve adaptability.

REFERENCES:

[1] L. Tanwar et J. Panda, « Review of different transforms used in digital image watermarking », in *2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, IEEE, 2018, p. 1165-1171.

[2] S. Fazli et G. Khodaverdi, « Trade-off between imperceptibility and robustness of LSB watermarking using SSIM quality metrics », in *2009 Second International Conference on Machine Vision*, IEEE, 2009, p. 101-104.

[3] T. K. Araghi, A. Manaf, M. Zamani, et S. K. Araghi, « A survey on digital image watermarking techniques in spatial and transform domains », *Int J Adv Image Process Tech*, vol. 3, nº 1, p. 6-10, 2016.

[4] N. Jimson et K. Hemachandran, « DFT based coefficient exchange digital image watermarking », in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2018, p. 567-571.

[5] D. K. Mahto et A. K. Singh, « A survey of color image watermarking: State-of-the-art and research directions », *Comput. Electr. Eng.*, vol. 93, p. 107255, 2021.

[6] T. K. Tsui, X.-P. Zhang, et D. Androutsos, « Color image watermarking using multidimensional Fourier transforms », *IEEE Trans. Inf. Forensics Secur.*, vol. 3, nº 1, p. 16-28, 2008.

[7] I. J. Goodfellow *et al.*, « Generative adversarial nets », *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[8] Z. Pang, M. Wang, L. Cao, X. Chai, et Z. Gan, « Pairwise open-sourced dataSet protection based on adaptive blind watermarking », *Appl. Intell.*, vol. 53, nº 14, p. 17391-17410, 2023.

[9] J.-Y. Zhu, T. Park, P. Isola, et A. A. Efros, « Unpaired image-to-image translation using cycle-consistent adversarial networks », in *Proceedings of the IEEE international conference on computer vision*, 2017, p. 2223-2232.

[10] M. Arjovsky, S. Chintala, et L. Bottou, « Wasserstein generative adversarial networks », in *International conference on machine learning*, PMLR, 2017, p. 214-223.

[11] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, et P. Abbeel, « Infogan: Interpretable representation learning by information maximizing generative adversarial nets », *Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.

[12] K. A. Zhang, L. Xu, A. Cuesta-Infante, et K. Veeramachaneni, « Robust invisible video watermarking with attention », *ArXiv Prepr. ArXiv190901285*, 2019.

[13] A. H. Allaf et M. A. Kbir, « Adapting RivaGAN for Robust Image Watermarking with Attention Mechanisms », *Int. J. Eng. Trends Technol.*, vol. 73, nº 4, p. 81-91, avr. 2025, doi: 10.14445/22315381/IJETT-V73I4P108.

[14] H.-C. Huang et W.-C. Fang, « Metadata-based image watermarking for copyright protection », *Simul. Model. Pract. Theory*, vol. 18, nº 4, p. 436-445, 2010.

[15] S. P. Vaidya et P. C. Mouli, « Adaptive digital watermarking for copyright protection of digital images in wavelet domain », *Procedia Comput. Sci.*, vol. 58, p. 233-240, 2015.

[16] A. H. Allaf et M. A. Kbir, « A review of digital watermarking applications for medical image exchange security », in *The proceedings of the third international conference on smart city applications*, Springer, 2018, p. 472-480.

[17] A. H. Allaf et M. A. Kbir, « Usage of watermarking techniques in medical imaging », in *Proceedings of the 3rd International Conference on Smart City Applications*, 2018, p. 1-6.

[18] G. Hua, J. Huang, Y. Q. Shi, J. Goh, et V. L. Thing, « Twenty years of digital audio

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

watermarking—a comprehensive review », *Signal Process.*, vol. 128, p. 222-242, 2016.

[19] M. Asikuzzaman et M. R. Pickering, « An overview of digital video watermarking », *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, nº 9, p. 2131-2153, 2017.

[20] K. Tanaka, Y. Nakamura, et K. Matsui, « Embedding secret information into a dithered multi-level image », in *IEEE Military Communications Conference*, 1990, p. 216-220.

[21] I.-K. Yeo et H. J. Kim, « Generalized patchwork algorithm for image watermarking », *Multimed. Syst.*, vol. 9, nº 3, p. 261-265, 2003.

[22] G. Manjula et A. Danti, « A novel hash based least significant bit (2-3-3) image steganography in spatial domain », *ArXiv Prepr. ArXiv150303674*, 2015.

[23] G. N. Mohammed, A. Yasin, et A. M. Zeki, « Robust image watermarking based on dual intermediate significant bit (DISB) », in *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, IEEE, 2014, p. 18-22.

[24] M. Saqib et S. Naaz, « Spatial and frequency domain digital image watermarking techniques for copyright protection », *Int J Eng Sci TechnolIJEST*, vol. 9, nº 6, p. 691-699, 2017.

[25] K. Rinki, P. Verma, et R. K. Singh, « A novel matrix multiplication based LSB substitution mechanism for data security and authentication », *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, nº 8, p. 5510-5524, 2022.

[26] M. Mustaqim Abrar, A. Pal, et T. Shahriar Sazzad, « Bit Plane Slicing and Quantization-Based Color Image Watermarking in Spatial Domain », in *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020*, Springer, 2021, p. 371-383.

[27] W. Belferdi, A. Behloul, et L. Noui, « A Bayer pattern-based fragile watermarking scheme for color image tamper detection and restoration », *Multidimens. Syst. Signal Process.*, vol. 30, nº 3, p. 1093-1112, 2019.

[28] S. Bhalerao, I. A. Ansari, et A. Kumar, « A secure image watermarking for tamper detection and localization », *J. Ambient Intell. Humaniz. Comput.*, vol. 12, nº 1, p. 1057-1068, 2021.

[29] K. Thongkor, T. Amornraksa, et E. J. Delp, « Digital watermarking for camera-captured images based on just noticeable distortion and Wiener filtering », *J. Vis. Commun. Image Represent.*, vol. 53, p. 146-160, 2018.

[30] N. Pan-Pan, W. Xiang-Yang, L. Yu-Nan, et Y. Hong-Ying, « A robust color image watermarking using local invariant significant bitplane histogram », *Multimed. Tools Appl.*, vol. 76, nº 3, p. 3403-3433, 2017.

[31] P. Bas, J. Chassery, et B. Macq, « Toward a content-based watermarking scheme », *Trait. Signal*, vol. 19, nº 1, p. 11-18, 2002.

[32] A. H. Allaf et M. A. Kbir, « Watermarking Image Scheme Based on Image Content and Corners Decomposition Method », in *The Proceedings of the Third International Conference on Smart City Applications*, Springer, 2019, p. 693-703.

[33] A. H. Allaf et M. A. Kbir, « Corners based spatial decomposition method for image watermarking », in *Proceedings of the 4th International Conference on Smart City Applications*, 2019, p. 1-7.

[34] Q. Su et B. Chen, « Robust color image watermarking technique in the spatial domain », *Soft Comput.*, vol. 22, nº 1, p. 91-106, 2018.

[35] Z. Nana, « Watermarking algorithm of spatial domain image based on SVD », in *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, IEEE, 2016, p. 361-365.

[36] J. Abraham et V. Paul, « An imperceptible spatial domain color image watermarking scheme », *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 31, nº 1, p. 125-133, 2019.

[37] H. Kaur et V. Kaur, « Invisible video multiple watermarking using optimized techniques », in *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, IEEE, 2016, p. 1-9.

[38] M. Masoumi, M. Rezaei, et A. B. Hamza, « A blind spatio-temporal data hiding for video ownership verification in frequency domain », *AEU-Int. J. Electron. Commun.*, vol. 69, nº 12, p. 1868-1879, 2015.

[39] Z. Bahrami et F. Akhlaghian Tab, « A new robust video watermarking algorithm based on SURF features and block classification », *Multimed. Tools Appl.*, vol. 77, nº 1, p. 327-345, 2018.

[40] X. Li, X. Wang, W. Yang, et X. Wang, « A robust video watermarking scheme to scalable recompression and transcoding », in *2016 6th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, IEEE, 2016, p. 257-260.

[41] P. Venugopala, H. Sarojadevi, N. N. Chiplunkar, et V. Bhat, « Video watermarking by adjusting the pixel values and using scene

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

change detection », in *2014 Fifth International Conference on Signal and Image Processing*, IEEE, 2014, p. 259-264.

[42] J. Wang, W. B. Wan, X. X. Li, J. De Sun, et H. X. Zhang, « Color image watermarking based on orientation diversity and color complexity », *Expert Syst. Appl.*, vol. 140, p. 112868, 2020.

[43] N. A. Loan, N. N. Hurrah, S. A. Parah, J. W. Lee, J. A. Sheikh, et G. M. Bhat, « Secure and robust digital image watermarking using coefficient differencing and chaotic encryption », *IEEE Access*, vol. 6, p. 19876-19897, 2018.

[44] M. Moosazadeh et G. Ekbatanifard, « A new DCT-based robust image watermarking method using teaching-learning-based optimization », *J. Inf. Secur. Appl.*, vol. 47, p. 28-38, 2019.

[45] Y. Zhang, Z. Wang, Y. Zhan, L. Meng, J. Sun, et W. Wan, « JND-aware robust image watermarking with tri-directional inter-block correlation », *Int. J. Intell. Syst.*, vol. 36, nº 12, p. 7053-7079, 2021.

[46] V. P. Vishwakarma et V. Sisaudia, « Gray-scale image watermarking based on DE-KELM in DCT domain », *Procedia Comput. Sci.*, vol. 132, p. 1012-1020, 2018.

[47] K. Rangel-Espinoza, E. Fragoso-Navarro, C. Cruz-Ramos, R. Reyes-Reyes, M. Nakano-Miyatake, et H. M. Pérez-Meana, « Adaptive removable visible watermarking technique using dual watermarking for digital color images », *Multimed. Tools Appl.*, vol. 77, nº 11, p. 13047-13074, 2018.

[48] S. Prajwalasimha, S. Chethan Suputhra, et C. Mohan, « Performance analysis of combined discrete fourier transformation (DFT) and successive division based image watermarking scheme », *Int. J. Recent Technol. Eng.*, vol. 8, nº 3, p. 34-39, 2019.

[49] X. Liao, K. Li, et J. Yin, « Separable data hiding in encrypted image based on compressive sensing and discrete fourier transform », *Multimed. Tools Appl.*, vol. 76, nº 20, p. 20739-20753, 2017.

[50] R. Riad, F. Ros, K. Gourrame, M. E. Hajji, H. Douzi, et R. Harba, « A preventive and curative watermarking scheme for an industrial solution », *Multimed. Tools Appl.*, vol. 82, nº 1, p. 651-679, 2023.

[51] W. Chen, C. Zhu, N. Ren, T. Seppänen, et A. Keskinarkaus, « Screen-cam robust and blind watermarking for tile satellite images », *IEEE Access*, vol. 8, p. 125274-125294, 2020.

[52] L.-Y. Hsu et H.-T. Hu, « Blind watermarking for color images using EMMQ based on QDFT », *Expert Syst. Appl.*, vol. 149, p. 113225, 2020.

[53] M. Cedillo-Hernandez, A. Cedillo-Hernandez, et F. J. Garcia-Ugalde, « Improving DFT-based image watermarking using particle swarm optimization algorithm », *Mathematics*, vol. 9, nº 15, p. 1795, 2021.

[54] S. Yuan *et al.*, « A blind watermarking scheme based on computational ghost imaging in wavelet domain », *Opt. Commun.*, vol. 482, p. 126568, 2021.

[55] S. Kumar et B. K. Singh, « DWT based color image watermarking using maximum entropy », *Multimed. Tools Appl.*, vol. 80, nº 10, p. 15487-15510, 2021.

[56] A. Anand et A. K. Singh, « Joint watermarking-encryption-ECC for patient record security in wavelet domain », *IEEE Multimed.*, vol. 27, nº 3, p. 66-75, 2020.

[57] H. M. Al-Otum, « Secure and robust host-adapted color image watermarking using inter-layered wavelet-packets », *J. Vis. Commun. Image Represent.*, vol. 66, p. 102726, 2020.

[58] Y. Tan, J. Qin, X. Xiang, W. Ma, W. Pan, et N. N. Xiong, « A robust watermarking scheme in YCbCr color space based on channel coding », *IEEE Access*, vol. 7, p. 25026-25036, 2019.

[59] L. Zhang et D. Wei, « Image watermarking based on matrix decomposition and gyrator transform in invariant integer wavelet domain », *Signal Process.*, vol. 169, p. 107421, 2020.

[60] S. M. Darwish et L. D. S. Al-Khafaji, « Dual watermarking for color images: a new image copyright protection model based on the fusion of successive and segmented watermarking », *Multimed. Tools Appl.*, vol. 79, nº 9, p. 6503-6530, 2020.

[61] A. Alzahrani et N. A. Memon, « Blind and robust watermarking scheme in hybrid domain for copyright protection of medical images », *IEEE Access*, vol. 9, p. 113714-113734, 2021.

[62] K. Swaraja, K. Meenakshi, et others, « An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine », *Biomed. Signal Process. Control*, vol. 55, p. 101665, 2020.

[63] X. Kang, Y. Chen, F. Zhao, et G. Lin, « Multi-dimensional particle swarm optimization for robust blind image watermarking using intertwining logistic map and hybrid domain », *Soft Comput.*, vol. 24, nº 14, p. 10561-10584, 2020.

[64] P. Zheng et Y. Zhang, « A robust image watermarking scheme in hybrid transform

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

domains resisting to rotation attacks », *Multimed. Tools Appl.*, vol. 79, nº 25, p. 18343-18365, 2020.
[65] Z. Yuan, Q. Su, D. Liu, X. Zhang, et T. Yao, « Fast and robust image watermarking method in the spatial domain », *IET Image Process.*, vol. 14, nº 15, p. 3829-3838, 2020.
[66] F. Ernawan, D. Ariatmanto, et A. Firdaus, « An improved image watermarking by modifying selected DWT-DCT coefficients », *IEEE Access*, vol. 9, p. 45474-45485, 2021.
[67] S. Ge, Z. Xia, J. Fei, Y. Tong, J. Weng, et M. Li, « A robust document image watermarking scheme using deep neural network », *Multimed. Tools Appl.*, vol. 82, nº 25, p. 38589-38612, 2023.
[68] S. Jaiswal et M. K. Pandey, « Color watermark extraction using deep neural network in IWT domain with PCA-based statistical feature reduction », *SN Comput. Sci.*, vol. 4, nº 5, p. 669, 2023.
[69] R. Sinhal et I. A. Ansari, « Machine learning based multipurpose medical image watermarking », *Neural Comput. Appl.*, vol. 35, nº 31, p. 23041-23062, 2023.
[70] L. R. Battarusetty, G. R. N. Kumari, R. Tamilkodi, et B. S. Kumar, « Modeling automated image watermarking using meta-heuristic-based deep learning with wavelet approach », *Sens. Imaging*, vol. 24, nº 1, p. 24, 2023.
[71] L. Wang, Y. Song, et D. Xia, « Deep neural network watermarking based on a reversible image hiding network », *Pattern Anal. Appl.*, vol. 26, nº 3, p. 861-874, 2023.
[72] G. Liu, R. Xiang, J. Liu, R. Pan, et Z. Zhang, « An invisible and robust watermarking scheme using convolutional neural networks », *Expert Syst. Appl.*, vol. 210, p. 118529, 2022.
[73] A. Chacko et S. Chacko, « Deep learning-based robust medical image watermarking exploiting DCT and Harris hawks optimization », *Int. J. Intell. Syst.*, vol. 37, nº 8, p. 4810-4844, 2022.
[74] H.-C. Wu, W.-L. Fan, C.-S. Tsai, et J. J.-C. Ying, « An image authentication and recovery system based on discrete wavelet transform and convolutional neural networks », *Multimed. Tools Appl.*, vol. 81, nº 14, p. 19351-19375, 2022.
[75] H. K. Singh et A. K. Singh, « Digital image watermarking using deep learning », *Multimed. Tools Appl.*, vol. 83, nº 1, p. 2979-2994, 2024.
[76] M. M. Darwish, A. A. Farhat, et T. M. El-Gindy, « Convolutional neural network and 2D logistic-adjusted-Chebyshev-based zero-watermarking of color images », *Multimed. Tools Appl.*, vol. 83, nº 10, p. 29969-29995, 2024.
[77] L.-Y. Hsu, H.-T. Hu, et H.-H. Chou, « A high-capacity QRD-based blind color image watermarking algorithm incorporated with AI technologies », *Expert Syst. Appl.*, vol. 199, p. 117134, 2022.
[78] D. K. Mahto, A. Anand, et A. K. Singh, « Hybrid optimisation-based robust watermarking using denoising convolutional neural network », *Soft Comput.*, vol. 26, nº 16, p. 8105-8116, 2022.
[79] Z. Zhang, H. Wang, et H. Fu, « A convolutional neural network-based blind robust image watermarking approach exploiting the frequency domain », *Vis. Comput.*, vol. 39, nº 8, p. 3533-3544, 2023.
[80] D. Li *et al.*, « Hybrid encrypted watermarking algorithm for medical images based on DCT and improved DarkNet53 », *Electronics*, vol. 12, nº 7, p. 1554, 2023.
[81] S. A. Nawaz, J. Li, M. U. Shoukat, U. A. Bhatti, et M. A. Raza, « Hybrid medical image zero watermarking via discrete wavelet transform-ResNet101 and discrete cosine transform », *Comput. Electr. Eng.*, vol. 112, p. 108985, 2023.
[82] R. Xiang, G. Liu, K. Li, J. Liu, Z. Zhang, et M. Dang, « Zero-watermark scheme for medical image protection based on style feature and ResNet », *Biomed. Signal Process. Control*, vol. 86, p. 105127, 2023.
[83] C. Gong, J. Liu, M. Gong, J. Li, U. A. Bhatti, et J. Ma, « Robust medical zero-watermarking algorithm based on Residual-DenseNet », *IET Biom.*, vol. 11, nº 6, p. 547-556, 2022.
[84] F. Dong, J. Li, U. A. Bhatti, J. Liu, Y.-W. Chen, et D. Li, « Robust zero watermarking algorithm for medical images based on improved NasNet-mobile and DCT », *Electronics*, vol. 12, nº 16, p. 3444, 2023.
[85] S. A. Nawaz, J. Li, U. A. Bhatti, M. U. Shoukat, D. Li, et M. A. Raza, « Hybrid watermarking algorithm for medical images based on digital transformation and MobileNetV2 », *Inf. Sci.*, vol. 653, p. 119810, 2024.
[86] H. K. Singh, N. Baranwal, K. N. Singh, A. K. Singh, et H. Zhou, « GAN-based watermarking for encrypted images in healthcare scenarios », *Neurocomputing*, vol. 560, p. 126853, 2023.
[87] C. Annadurai, I. Nelson, K. N. Devi, R. Manikandan, et A. H. Gandomi, « Image watermarking based data hiding by discrete wavelet transform quantization model with convolutional generative adversarial

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

architectures », *Appl. Sci.*, vol. 13, nº 2, p. 804, 2023.

[88] J. Huang, T. Luo, L. Li, G. Yang, H. Xu, et C.-C. Chang, « ARWGAN: Attention-guided robust image watermarking model based on GAN », *IEEE Trans. Instrum. Meas.*, vol. 72, p. 1-17, 2023.

[89] Ren et Malik, « Learning a classification model for segmentation », in *Proceedings ninth IEEE international conference on computer vision*, IEEE, 2003, p. 10-17.

[90] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, et K. Siddiqi, « Turbopixels: Fast superpixels using geometric flows », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, nº 12, p. 2290-2297, 2009.

[91] A. P. Moore, S. J. Prince, et J. Warrell, « "Lattice Cut"-Constructing superpixels using layer constraints », in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2010, p. 2117-2124.

[92] M.-Y. Liu, O. Tuzel, S. Ramalingam, et R. Chellappa, « Entropy rate superpixel segmentation », in *CVPR 2011*, IEEE, 2011, p. 2097-2104.

[93] X. Qiao, R. Ni, et Y. Zhao, « Superpixel-based watermarking scheme for image authentication and recovery », in *International Workshop on Digital Watermarking*, Springer, 2014, p. 160-173.

[94] P. Niu, L. Wang, X. Shen, S. Zhang, et X. Wang, « A novel robust image watermarking in quaternion wavelet domain based on superpixel segmentation », *Multidimens. Syst. Signal Process.*, vol. 31, nº 4, p. 1509-1530, 2020.

[95] P. K. Singh, B. Jana, et K. Datta, « Robust watermarking scheme for compressed image through dct exploiting superpixel and arnold transform », in *Proceedings of the Sixth International Conference on Mathematics and Computing: ICMC 2020*, Springer, 2020, p. 43-54.

[96] P. K. Singh, B. Jana, et K. Datta, « Superpixel based robust reversible data hiding scheme exploiting Arnold transform with DCT and CA », *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, nº 7, p. 4402-4420, 2022.

[97] N. Jani et A. Jani, « SUPERPIXEL SEGMENTATION BASED DIGITAL IMAGE WATERMARKING FOR COLOR IMAGE AUTHENTICATION. », *Int. J. Adv. Res. Comput. Sci.*, vol. 8, nº 9, 2017.

[98] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, et S. Süsstrunk, « SLIC superpixels compared to state-of-the-art superpixel methods », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, nº 11, p. 2274-2282, 2012.

[99] C. E. Shannon, « A mathematical theory of communication », *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, nº 1, p. 3-55, 2001.

[100] A. Hore et D. Ziou, « Image quality metrics: PSNR vs. SSIM », in *2010 20th international conference on pattern recognition*, IEEE, 2010, p. 2366-2369.

[101] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, et T. Kalker, « Digital watermarking », *Morgan Kaufmann Publ.*, vol. 54, p. 56-59, 2008.