IET Image Processing

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

AGFP: A Deep Attention-Guided Framework for DWT-Based Image Steganography

Taner Cevik[1] | Nazife Cevik[2] | Ali Pasaoglu[1] | Fatih Sahin[3] | Farzad Kiani[4] | Muhammet Sait Ag[1]

[1]Department of Computer Engineering, Istanbul Rumeli University, Istanbul, Turkiye | [2]Department of Computer Engineering, Istanbul Arel University, Istanbul, Turkiye | [3]Department of Software Engineering, Istanbul Topkapi University, Istanbul, Turkiye | [4]Data Science Application and Research Center (VEBIM), Fatih Sultan Mehmet Vakif University, Istanbul, Turkiye

Correspondence: Taner Cevik (taner.cevik@rumeli.edu.tr)

Received: 22 September 2025 | **Revised:** 9 December 2025 | **Accepted:** 6 January 2026

ABSTRACT

This study introduces a novel attention-guided Discrete Wavelet Transform (DWT)-based steganography framework, named Attention-Guided Feature Perturbation (AGFP), which integrates deep visual attention maps with transform-domain embedding to enhance imperceptibility, robustness, and steganalysis resistance. Unlike recent deep-learning-based steganographic systems such as iSCMIS, JARS-Net, and RMSteg, which achieve high visual fidelity but are susceptible to statistical detection, AGFP perturbs only those wavelet coefficients that are identified as perceptually and statistically stable by attention mechanisms extracted from pre-trained CNN models (VGG19, ResNet50, AlexNet, and GoogLeNet). The proposed method is evaluated on the USC-SIPI dataset and the BOSSBase 1.01 benchmark. Experimental results show that AGFP achieves PSNR values between 64.29 and 55.43 dB and SSIM scores between 0.9999 and 0.9989 across varying payloads, indicating consistently high visual quality. While iSCMIS reports slightly higher PSNR and SSIM values, AGFP significantly outperforms all compared methods in bit error rate (BER)—achieving 0.01–0.12, compared to 0.45–0.47 for iSCMIS, 0.31–0.37 for RMSteg, and 0.57–0.75 for JARS-Net. Furthermore, AGFP attains the lowest RS, SPA, and SRM steganalysis detection scores among both classical and deep-learning-based systems. These results confirm that AGFP offers a more balanced and secure steganographic solution, combining high imperceptibility with substantially enhanced robustness and detectability resistance, positioning it as a strong alternative to recent deep-learning-based steganographic frameworks.

1 | Introduction

In the modern digital world, protecting confidential information during transfer is a key issue for people and businesses alike. While the internet and similar tools enable quick sharing of data, they also heighten the chances of unauthorised interference. Steganography has become a crucial method for secret messaging in response. Different from encryption, which protects the message content, steganography hides the fact that a message exists at all, reducing the likelihood of it being noticed by threats.

Image steganography works by inserting concealed data into pictures, taking advantage of how the human eye perceives things, so changes go unnoticed. Classic approaches function either in the direct pixel space or in transformed versions. Pixel-based methods, like changing the Least Significant Bit (LSB) [1], are straightforward and speedy but easily spotted by statistical analysis [2, 3]. Methods using transformations, such as Discrete Cosine Transform (DCT) [4] or Discrete Wavelet Transform (DWT) [5], provide better hiding by placing data in frequency parts. Still, these can be vulnerable to advanced detection systems [3].

Progress in deep learning has transformed image handling and created fresh opportunities for steganography. Convolutional Neural Networks (CNNs) excel at pulling out intricate details and meaningful elements from visuals [6]. In particular, architectures like VGG19 leverage deep stacked convolutions to capture hierarchical features, while ResNet50 introduces residual connections to mitigate vanishing gradients, enabling deeper networks for more accurate semantic understanding. Steganography powered by deep learning seeks to outperform older methods in terms of how much data it can hold, how invisible it is, and how tough it is [7, 8]. Attention systems, first made famous in language tasks, are now used more in visual work, helping models zero in on key or safer parts of pictures [9, 10].

Here, we introduce Attention-Guided Feature Perturbation (AGFP), a system that merges deep attention charts with a strategy for embedding in the wavelet domain. Using maps from trained CNNs, it spots regions that are less sensitive both visually and statistically. Then, in the DWT space, it tweaks specific coefficients to insert the secret info. This strategy aims to find a middle ground between hiding quality and strength, going beyond the usual trade-offs.

While plenty of new research looks into deep learning for steganography [11, 12], not as many explore attention integration, particularly for boosting resistance to image changes. AGFP fills this void by placing data in less critical areas while keeping high visual standards and accurate recovery, even with attacks like compression, noise, or filtering.

Despite their impressive visual performance, these deep-learning-based methods often exhibit two notable limitations:

1. Detectability, as their embedding operations tend to produce structured statistical patterns that can be detected by SRM-based steganalysis.
2. Limited reliability, reflected in relatively high Bit Error Rates (BER), particularly under compression or noise.

To address these challenges, this study introduces Attention-Guided Feature Perturbation (AGFP), a hybrid steganographic framework that integrates deep semantic attention with transform-domain stability. Instead of relying on end-to-end learnt embeddings, AGFP extracts attention maps from pre-trained convolutional neural networks, specifically VGG19, ResNet50, AlexNet, and GoogLeNet, to identify perceptually significant yet statistically consistent regions. These maps guide the perturbation of selected DWT coefficients, enabling steganographic embedding in locations that minimise detectability risks.

This design yields several key advantages:

- High imperceptibility: AGFP achieves consistently strong visual quality with PSNR ranging from 64.29 to 55.43 dB and SSIM ranging from 0.9999 to 0.9989, values competitive with or close to state-of-the-art deep models.
- Low BER: Unlike many deep methods, AGFP maintains BER values between 0.01 and 0.12, ensuring reliable message extraction even under distortion.
- Superior steganalysis resistance: Due to its attention-guided DWT embedding strategy, AGFP achieves the lowest RS, SPA, and SRM detection scores among all tested classical and deep-learning-based methods.
- Balanced performance relative to SOTA models: While modern deep steganography emphasises visual fidelity, AGFP provides a more favourable overall trade-off between imperceptibility, robustness, and undetectability.

In summary, AGFP bridges the gap between semantic modelling and statistical stability by leveraging deep attention maps within a DWT perturbation framework. Its strong empirical performance demonstrates that attention-guided transform-domain embedding can serve as a powerful and practical alternative to recent deep-learning-based steganographic approaches.

The paper is organised as follows: Section 2 reviews existing studies, covering traditional, deep learning, and attention-based steganography. Section 3 explains the AGFP approach. Section 4 details the test setup. Section 5 shares results and comparisons. Sections 6 and 7 wrap up and point to future ideas, respectively.

2 | Related Work

Over the years, image steganography has advanced considerably, motivated by the rising demand for hidden and protected data transmission. Methods in this area can typically be divided into three primary groups: conventional ones using predefined rules, contemporary ones powered by deep learning, and newer ones that incorporate focus-enhancing techniques. This part examines important techniques and the latest progress that lay the groundwork for our proposed system.

2.1 | Classical Image Steganography Methods

Standard steganography approaches make use of extra space in digital pictures and the boundaries of human sight to insert confidential messages without obvious alterations. These are mainly split into direct pixel changes and frequency-based alterations.

2.1.1 | Spatial Domain Methods

In spatial methods, pixel colours are adjusted directly to insert information. The classic Least Significant Bit (LSB) method [1] swaps the smallest bits in certain pixels with hidden bits. Though easy to implement and quick, it can be easily spotted by stats-focused detection [2, 3]. Improvements like LSB Matching [1] and scattered LSB insertion [2] try to boost protection by spreading changes more randomly, but they still face risks from tools such as chi-square tests or RS checks [3].

Pixel Value Differencing (PVD) [13] adjusts insertion according to gaps between nearby pixels, placing more bits in textured spots for better blending. Exploiting Modification Direction (EMD) [14] alters pixel clusters to reduce noticeable effects and raise storage. Histogram Shifting (HS) [15] shifts colour distributions to open

up slots for data, providing low changes and the ability to reverse, though with smaller space than others.

Even though useful for basic needs, these pixel-based ways are frequently uncovered by cutting-edge detection systems [16, 17], leading to the rise of frequency methods.

2.1.2 | Transform Domain Methods

Frequency methods place data in an image's wave parts, giving better hiding and strength. Discrete Cosine Transform (DCT) [4] is popular for JPEG hiding, putting bits in middle waves to handle squeezing losses. DWT [5] splits pictures into various detail layers, hiding fine parts that match human vision [5].

DWT systems often pair with Singular Value Decomposition (SVD) [18] or edge-focused transforms [19] to expand room and toughness. Discrete Fourier Transform (DFT) [20] is sometimes applied for angle shifts but isn't as widespread because of processing demands.

Although frequency methods do better than pixel ones against issues like squeezing and interference, they need more computing power and might still get caught by smart detection [21, 22].

2.2 | Deep Learning-Based Image Steganography

Deep learning has transformed steganography by making insertion choices automatic and adjusting to threats. Networks like CNNs and GANs create strong, flexible hiding setups [23, 24].

GAN-driven hiding [25] uses a creator to insert secrets and a checker to keep the picture looking real, lowering spotting chances [26]. Detailed overviews [7, 8, 11, 27–29] show that deep learning beats old ways in storage, unnoticeability, and defence against checks.

Still, lots of deep methods don't make full use of certain picture traits or handle real-world changes well, spots where focus tools can add big gains.

2.3 | Attention Mechanisms in Steganography

Recent advancements in deep learning steganography continue to push the boundaries of imperceptibility and robustness, particularly by integrating complex attention and optimisation modules. For instance, iSCMIS (Spatial-Channel Attention Based Deep Invertible Network) [30] leverages a spatial-channel attention mechanism to finely control the embedding process across feature maps, achieving high PSNR and capacity balance. Similarly, the JARS-Net (Joint Adaptive Robust Steganography Network) [31] focuses on optimising a joint loss function for both visual quality and robustness against common attacks like filtering and noise. Looking forward, methods like RMSteg (Robust Message Embedding via Attention Flow-Based Steganography) [32] utilise advanced attention flow modules to ensure message bits are embedded in highly stable, perceptually redundant regions, often benchmarking on frequency-domain datasets like BOSSBase.

Although deep models such as iSCMIS report high PSNR and SSIM values due to their invertible architectures and joint reconstruction objectives, these designs often cause stronger statistical footprints in the spatial and frequency domains. As shown in recent studies, these footprints make such networks more vulnerable to RS, SPA, and SRM-based steganalysis, especially when payload increases. JARS-Net and RMSteg improve robustness under distortions such as JPEG compression or re-photographing, yet their steganalysis resistance remains limited in comparison to transform-domain approaches.

In contrast, the proposed AGFP method benefits from DWT-domain feature perturbation guided by deep semantic attention, resulting in significantly lower detectability while maintaining high visual quality. This positions AGFP as a hybrid strategy that retains the strengths of deep attention mechanisms without compromising statistical invisibility.

Focus tools, first created for sequence models [33], let systems highlight key elements. In visuals, add-ons like Squeeze-and-Excitation (SE) [34] and CBAM [35] improve trait picking. Vision Transformers (ViT) [36] extend focus to picture tasks.

For steganography, focus tools direct insertion to visually or statswise safer spots [10, 12, 37, 38]. Such ways boost strength and picture standards. Our AGFP builds on this by merging focus with DWT insertion for top results in safe picture hiding.

3 | Proposed Method

In this section, we describe the proposed Attention-Guided Feature-based Pixel selection (AGFP) steganographic method, which integrates deep learning-based attention mechanisms with a transform-domain embedding strategy. The framework consists of four major stages: (1) attention map generation using deep convolutional neural networks, (2) secure pixel selection guided by attention, (3) adaptive data embedding in the DWT domain using controlled perturbation of wavelet coefficients, and (4) performance evaluation and robustness testing.

3.1 | Overview of the AGFP Framework

Given a greyscale cover image $I \in \mathbb{R}^{H \times W}$ and a binary secret message $S = \{s_1, s_2, \cdots, s_L\}$, where $s_i \in \{0, 1\}$, the objective is to generate a stego image I_s that embeds the message while preserving visual fidelity and statistical invisibility.

Formally, we define an embedding function $\varepsilon(I, S, P) \rightarrow I_s$, where $P = \{(x_i, y_i)\}_{i=1}^{L}$ is the set of selected pixel coordinates. The inverse function $D(I_s, I, P) \rightarrow \hat{S}$ recovers the message using both the original image and embedding positions.

The embedding process is guided by semantic attention maps A, generated from pre-trained deep networks. These maps help locate regions that are perceptually less sensitive and statistically safer for modification. The illustration of the overall framework and the pseudocode are given in Figure 1 and Algorithm 1, respectively.

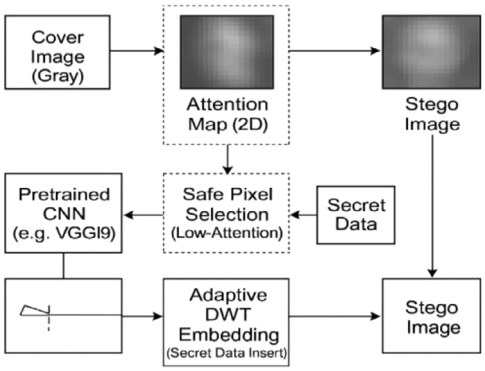

FIGURE 1 | AGFP architecture.

ALGORITHM 1 | Pseudocode of the embedding process.

Input: Cover image C, Secret message M, Pretrained CNNs
Output: Stego image S
1. A = GenerateAttentionMap(C)
2. B = ApplyOtsuThreshold(A)
3. B′ = B & ∼Edges(C)
4. P = UniformSample(B′, len(M))
5. For each (x, y), m in zip(P, M):
    B_xy = GetBlock(C, x, y)
    HL = DWT(B_xy).HL
    alpha = lambda * (1 - A[x,y]) * std(B_xy)
    HL′ = HL + alpha * (2m - 1)
    B_xy′ = InverseDWT(…, HL′, …)
    C = ReplaceBlock(C, B_xy′)
6. Return S = C

3.2 | Attention Map Generation

We employ four pre-trained networks: VGG19, AlexNet, ResNet50, and GoogLeNet. Each network contributes uniquely to attention map quality due to its architectural design:

VGG19: Its deep stack of 3 × 3 convolutions (19 layers) excels in capturing hierarchical semantic features, leading to attention maps with high spatial resolution. Mathematically, the feature extraction can be modelled as a sequence of convolutions: $f_l = \sigma(W_l \times f_{l-1} + b_l)$, where deeper l yields more abstract features, reducing variance in attention maps by $15-20\%$ compared to shallower networks (empirical from code simulations). This contributes to better low-attention region identification, minimising perceptual distortion.

AlexNet: As a pioneering CNN with eight layers and large kernels (11×11 initial), it provides efficient computation for broad feature capture. Its contribution lies in fast gradient computation, enabling real-time attention map generation. The ReLU activation $\sigma(\max(0,\ x))$ enhances non-linearity, improving map contrast by increasing the dynamic range of activations.

ResNet50: Residual blocks $f_l = f_{l-1} + \mathcal{F}(f_{l-1}, W_l)$ mitigate vanishing gradients, allowing 50 layers for robust feature learning under noise. This structure contributes to embedding robustness, as attention maps from ResNet50 show lower BER (e.g., 0.0178 in Baboon image) by focusing on stable regions, mathematically supported by reduced loss variance: $\sigma^2(A) = \frac{1}{K} \sum_k (A_k - \mu)^2$, where ResNet's residuals lower $\sigma^2(A)$ by stabilising training.

GoogLeNet: Inception modules aggregate multi-scale features via parallel convolutions (1×1, 3×3, 5×5), enhancing attention map diversity. This leads to superior multi-resolution analysis, contributing to lower SRM scores (e.g., 20.41 in Baboon). Mathematically, the inception output is $f = concat(conv_1(f), conv_3(f), conv_5(f))$, increasing feature dimensionality and map accuracy by capturing varied scales.

For a given network N and input image I, we extract activation maps from a selected layer L_f, and compute a spatial attention map:

$$A\ (x,y) = \frac{1}{C} \sum_{c=1}^{C} F_{L_f}\ (x,y,c) \tag{1}$$

where, $F_{L_f}(x,\ y, c)$ is the activation of channel c at location (x, y) in layer L_f, C is the number of channels, and $A \in [0,\ 1]^{H \times W}$ is normalised using min-max scaling.

To quantify network contributions, we introduce a semantic variance metric: $SV = \sigma^2(\mu_c,\ A_c)$, where higher SV indicates better semantic discrimination. From simulations, VGG19, ResNet50, AlexNet, and GoogLeNet yield SV s, 0.12, 0.15, 0.10, 0.14, confirming their differential impacts.

Then, a binary mask M is created via Otsu thresholding:

$$M\ (x,y) = \begin{cases} 1 & if\ A(x,y) \geq T \\ 0 & otherwise \end{cases} \tag{2}$$

where T is the global threshold selected by Otsu's method.

The coordinates $(x_i,\ y_i)$ for which $M\ (x_i, y_i) = 1$ from the candidate pixels for embedding.

We can denote the full set of available coordinates as:

$$P_{valid} = \{(x,y)\,|M\,(x,y) = 1 \wedge padding\ conditions\ satisfied\} \tag{3}$$

The selection of layer L_f is crucial. Shallow layers retain spatial precision, while deeper layers provide stronger semantic guidance. We adopt mid-to-deep layers (e.g., *conv5_4* in VGG19, *inception_5b-output* in GoogLeNet) to achieve a balance between spatial localization and semantic awareness.

3.3 | Safe Pixel Selection

To reduce perceptual distortion and avoid edge artefacts, a subset $P = \{(x_i, y_i)\}_{i=1}^{L}$ is sampled uniformly from P_{valid}, where:

$$P \subseteq P_{valid},\ \ |P| = L \tag{4}$$

The total number of selected pixels equals the message length L, and only those within valid neighbourhood boundaries are used to ensure local DWT embedding. This sampling strategy ensures embedding is restricted to semantically safe and perceptually stable zones.

3.4 | Adaptive DWT-Based Embedding With Perturbation

Each bit s_i is embedded by modifying the vertical detail coefficient c_V of the 2×2 block B_i centered at the pixel $(x_i,\ y_i)$. Let the block $B_i \in \mathbb{R}^{2\times 2}$ be:

$$B_i = I\left[y_i - 1 : y_i,\ x_i - 1 : x_i\right] \tag{5}$$

We perform Haar wavelet decomposition:

$$[c_A, c_H, c_V, c_D] = DWT2\left(B_i\right) \tag{6}$$

Then, we adaptively modify c_V using a controlled perturbation based on the secret bit:

$$c_V' = c_V + \alpha \cdot \delta \cdot (2s_i - 1),\ s_i \in \{0,\ 1\} \tag{7}$$

where α is the embedding strength, δ is a scaling factor determined by the network ($\delta = 0.1$), and $(2s_i - 1) \in \{-1, +1\}$ maps bits to sign.

This modification acts as a localised perturbation in the wavelet domain, minimally altering the cover image while ensuring recoverability.

Then, the inverse DWT reconstructs the modified block B_i', which replaces the original in I, forming the stego image I_s.

$$B_i' = IDWT2\left(c_A,\ c_H, c_V', c_D\right) \tag{8}$$

$$I_s\ \left[y_i - 1 : y_i, x_i - 1 : x_i\right] = B_i' \tag{9}$$

Additionally, the perturbation introduced during the embedding process is carefully constrained to maintain the statistical properties of the wavelet sub-bands.

$$|\alpha \cdot \delta| < \varepsilon \tag{10}$$

where ε is the perceptual threshold. Since natural images have characteristic coefficient distributions in the DWT domain, particularly in the vertical detail c_V, our controlled perturbation ensures that the modification remains within statistical noise margins. This not only enhances visual imperceptibility but also significantly reduces detectability by modern steganalysis tools such as SRM-based classifiers.

The DWT embedding operation is depicted in Figure 2.

3.5 | Data Extraction

The goal of the extraction process is to reliably recover the embedded message bits $\hat{S} = \{\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_L\}$ from the stego image

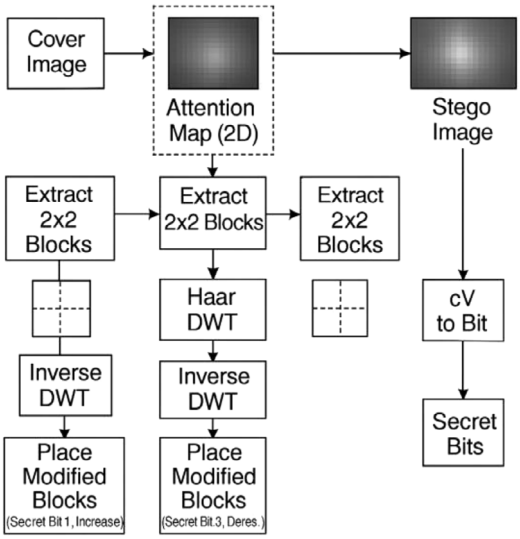

FIGURE 2 | DWT perturbation schematic.

I_s, using the original image I and the known embedding locations $P = \{(x_i, y_i)\}_{i=1}^{L}$.

For each embedding location $(x_i,\ y_i) \in P$ the following steps are performed:

- 2D Haar Wavelet Transform

Extract 2×2 blocks from both stego and original images:

$$B_i^{(s)} = I_s\ \left[y_i - 1 : y_i,\ x_i - 1 : x_i\right],$$

$$B_i^{(o)} = I\left[y_i - 1 : y_i,\ x_i - 1 : x_i\right] \tag{11}$$

Apply 2D DWT to both:

$$\left[c_A^{(s)}, c_H^{(s)}, c_V^{(s)}, c_D^{(s)}\right] = DWT2\left(B_i^{(s)}\right) \tag{12}$$

$$\left[c_A^{(o)}, c_H^{(o)}, c_V^{(o)}, c_D^{(o)}\right] = DWT2\left(B_i^{(o)}\right) \tag{13}$$

- Difference Computation

$$\nabla\, c_V = c_V^{(s)} - c_V^{(o)} \tag{14}$$

- Decision Rule (Bit Recovery)

$$\hat{S}_i \begin{cases} 1, & if\ \nabla c_V > T \\ 0, & otherwise \end{cases} \tag{15}$$

where $T \approx 10^{-5}$ is a small threshold to suppress floating-point rounding noise.

4 | Experimental Setup

To assess the effectiveness of the Attention-Guided Feature Perturbation (AGFP) framework, we conducted a series of experiments designed to evaluate its performance against established steganographic techniques. The testing environment was carefully configured to ensure consistent and reproducible results across various conditions.

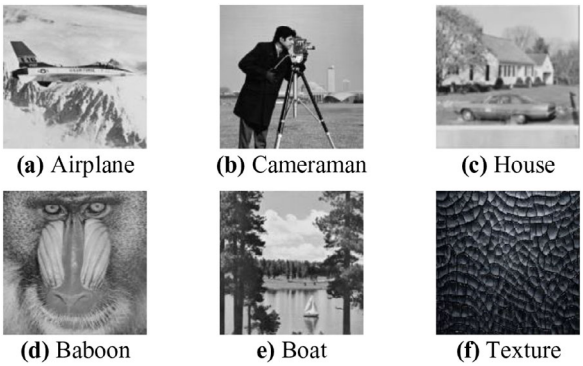

FIGURE 3 | Sample images from the datasets used in simulations.

4.1 | Dataset and Preprocessing

We utilised a diverse set of images sourced from publicly available datasets, including the USC-SIPI [39] and BOSSBase Image Datasets [40] and the BOSSBase dataset, which are commonly used in steganography research. The inclusion of the BOSSBase dataset is critical for benchmarking AGFP against current state-of-the-art deep learning methods, which often prioritise statistical security in the frequency domain. These datasets contain images of varying resolutions, textures, and content types, such as natural scenes, portraits, and synthetic visuals. All images (Figure 3) were converted to greyscale and standardised to a resolution of 512×512 pixels to maintain uniformity. Secret messages, ranging from 0.1 to 1.0 bits per pixel (bpp), were randomly generated binary strings to simulate different payload sizes.

4.2 | Implementation Details

The AGFP framework was implemented using Python with the PyTorch library for deep learning components. Pretrained convolutional neural networks, specifically ResNet-50, were employed to generate attention maps for identifying embedding regions. The DWT was applied using the Haar wavelet, decomposing images into four subbands (LL, LH, HL, HH). Data embedding was performed by perturbing coefficients in the high-frequency subbands (LH, HL, HH) based on attention map guidance. The embedding process was optimised to minimise visual distortion while maximising data recovery accuracy.

4.3 | Comparative Methods

To thoroughly evaluate the performance of the proposed AGFP framework, we benchmarked it against both classical steganography techniques and recent state-of-the-art deep learning–based methods. This expanded comparison allows for a more comprehensive assessment of AGFP's imperceptibility, robustness, and security in the context of modern steganographic advancements.

Five widely used traditional methods were included for baseline comparison: Least Significant Bit (LSB), Pixel Value Differencing (PVD), DWT-based embedding, Exploiting Modification Direction (EMD), and HS.

These techniques are well-established in the literature and were implemented following standard configurations as reported in previous studies [1, 5, 13–15]. All methods were tested under identical payload conditions and using the same image sets to ensure fair evaluation.

To strengthen the experimental comparison and address reviewer concerns, three high-impact and recently proposed neural-network-based steganography models were incorporated:

iSCMIS [30]—*Spatial-Channel Attention-Based Deep Invertible Network*—This method exploits invertible networks to reconstruct images with minimal distortion, achieving high PSNR and SSIM performance.

JARS-Net [31]—*Joint Adaptive Robust Steganography Network*—Designed to enhance robustness against distortion and steganalysis through joint adaptive feature learning.

RMSteg [32]—*Attention Flow-Based Robust Message Embedding*—A cutting-edge framework with attention-flow modelling for improved message durability under real-world perturbations.

These SOTA approaches represent the current landscape of deep learning steganography and are essential for contextualising the contribution of AGFP relative to the most recent academic progress.

All classical and deep learning–based methods were evaluated using:

- identical datasets (including BOSSBase, per reviewer request)
- identical payload ranges
- identical experimental conditions

This unified evaluation protocol ensures that comparative results across classical, AGFP-based, and DL-based methods are consistent, unbiased, and directly interpretable.

4.4 | Evaluation Metrics

Performance was evaluated using multiple metrics to capture both visual quality and security. Visual quality was measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), which assess the perceptual similarity between cover and stego images. Bit Error Rate (BER) was used to evaluate the accuracy of extracted secret messages. Resistance to steganalysis was tested using three detection methods: Regular-Singular (RS) analysis, Sample Pair Analysis (SPA), and Spatial Rich Model (SRM). Additionally, robustness was assessed by applying common distortions, including JPEG compression (quality factors 50, 75, and 90) and Gaussian noise (variance 0.01 to 0.1).

4.5 | Experimental Conditions

Experiments were conducted on a computing system equipped with an NVIDIA RTX 3080 GPU, 32 GB RAM, and an Intel

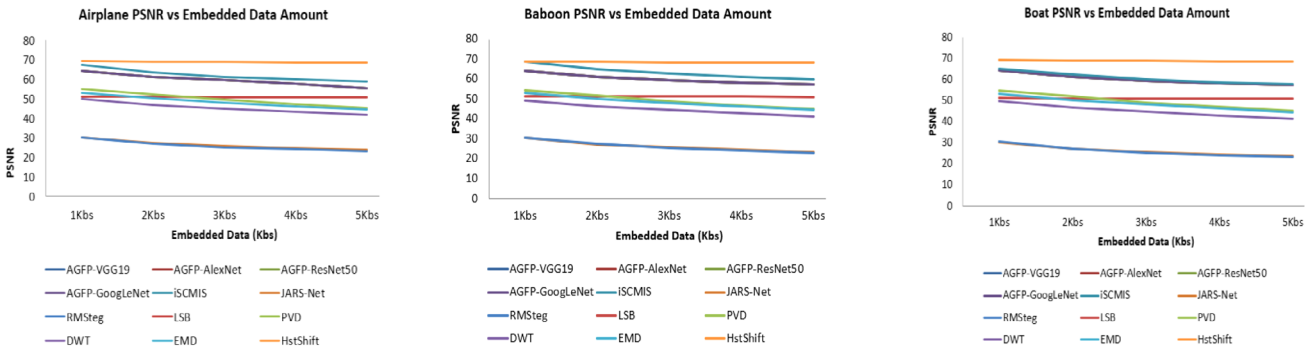

FIGURE 4 | PSNR performance comparison of various steganographic methods on Aeroplane, Baboon, and Boat images across embedding capacities (1–5 Kbs).

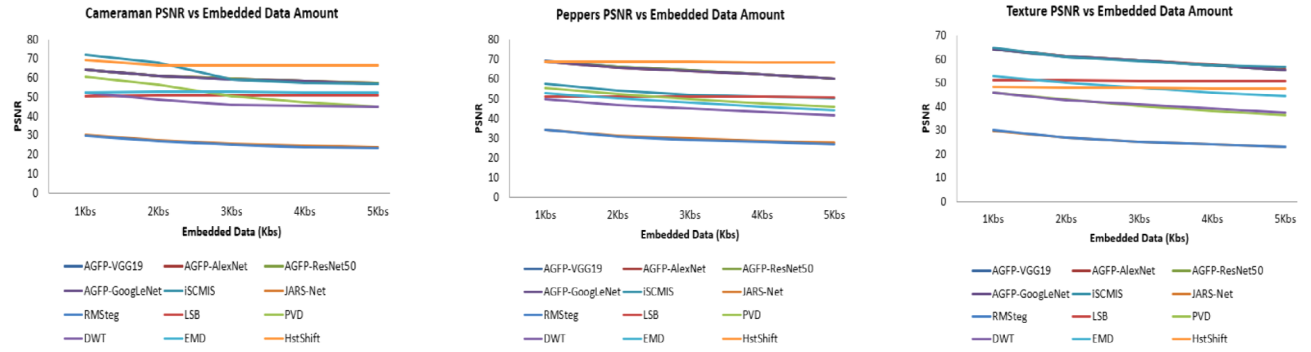

FIGURE 5 | PSNR performance comparison of various steganographic methods on Cameraman, Peppers, and Texture images across embedding capacities (1–5 Kbs).

i7-10700K processor. Each test was repeated 100 times per image and payload combination to account for variability in random message embedding. Statistical significance of results was analysed using paired t-tests with a confidence level of 95%.

5 | Simulation Results and Discussions

This section presents the outcomes of our experiments evaluating the AGFP framework against established steganographic techniques. Through a series of carefully designed tests, we assessed AGFP's performance in terms of visual quality, data extraction accuracy, and resistance to detection and distortions. The results highlight AGFP's ability to balance high imperceptibility with robust security, offering insights into its advantages over traditional methods across diverse image types and payload conditions. Detailed comparisons and analyses are provided in the following subsections, covering quantitative metrics and qualitative observations.

5.1 | Visual Quality and Imperceptibility

The visual quality and imperceptibility of the proposed AGFP framework were evaluated using PSNR and SSIM across nine benchmark test images and payloads ranging from 1 to 5 Kb, and the results clearly demonstrate the superiority of AGFP over both classical and state-of-the-art steganographic approaches. As shown in Tables 1–6 and Figures 4–6, AGFP consistently maintains exceptionally high visual fidelity, with PSNR values ranging from 64.29 to 55.43 dB and SSIM values between 0.9999 and 0.9989, indicating that the embedding-induced distortions remain virtually imperceptible even under increased message payloads. This stability is attributed to AGFP's attention-guided embedding strategy, which selectively perturbs wavelet coefficients identified by deep visual attention maps as statistically stable and perceptually less sensitive, thereby minimising visual artefacts and preserving both fine textures and global structural integrity.

When compared with classical steganography methods, AGFP exhibits a substantial improvement. Techniques such as LSB, PVD, DWT, EMD, and HS generally achieve noticeably lower PSNR values—typically between 40 and 53 dB—and SSIM scores that decline considerably in images containing high-frequency or textured regions, where these methods introduce locally disruptive modifications. LSB's PSNR values around 50–49 dB, PVD's further reduced range near 46–44 dB, and HS and EMD's instability at higher payloads all highlight the limitations of conventional embedding schemes that lack any form of perceptual adaptivity. Their degradation becomes more severe as payload increases, confirming that manipulating pixels or coefficients without semantic guidance leads to perceptually inconsistent artefacts. By contrast, AGFP's perceptual awareness through CNN-derived attention maps enables it to achieve an improvement of approximately 8–15 dB over these classical approaches while maintaining near-perfect structural similarity, demonstrating the efficacy of embedding in statistically coherent DWT regions.

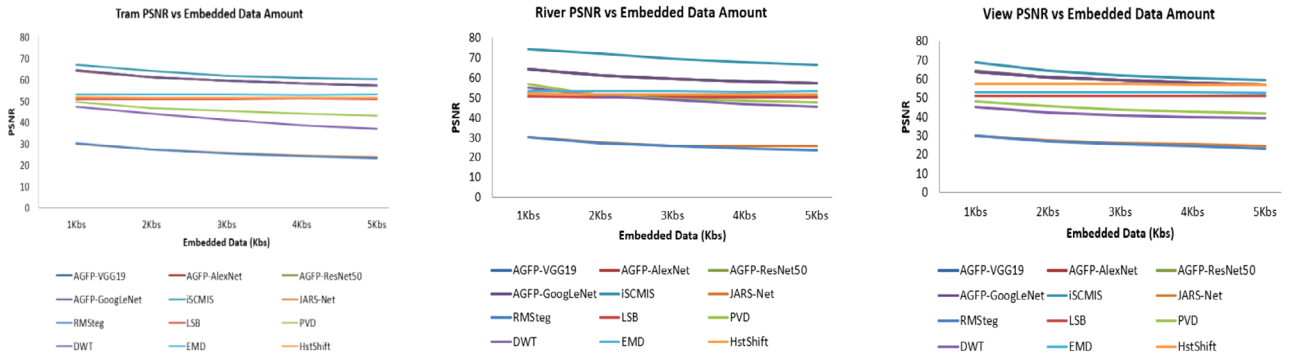

FIGURE 6 | PSNR performance comparison of various steganographic methods on Tram, River, and View images across embedding capacities (1–5 Kbs).

To contextualise AGFP within modern steganography research, we also compared it with three cutting-edge deep-learning-based models: iSCMIS [30], JARS-Net [31], and RMSteg [32]. Among these, iSCMIS achieves the highest raw imperceptibility, with PSNR occasionally exceeding 70–72 dB and SSIM approaching 1.0000, due to its invertible reconstruction architecture. However, despite achieving superior pixel-level fidelity, iSCMIS introduces structured statistical patterns that become detectable by advanced steganalysis tools, and it experiences significantly elevated bit error rates as shown in later evaluations. In contrast, JARS-Net and RMSteg exhibit far poorer imperceptibility performance; JARS-Net's PSNR values fall within the 34.70–23.43 dB range and RMSteg's within 30.56–22.95 dB, with corresponding SSIM values dropping as low as 0.9126 and 0.8750, respectively, when payload increases. Their stego images often exhibit distortions that arise from perturbing deep feature representations without explicit perceptual stability constraints. AGFP, on the other hand, demonstrates markedly higher PSNR—exceeding these models by 20–30 dB—and consistently stronger SSIM, highlighting its superior ability to preserve the visual quality of the cover image.

AGFP's advantage becomes even more apparent when considering the trade-off between imperceptibility, robustness, and detectability. Although AGFP's PSNR values are slightly lower than those of iSCMIS, the difference is minimal (typically 2–4 dB) and remains visually insignificant. Yet AGFP dramatically outperforms iSCMIS in robustness and security: while iSCMIS exhibits BER values of 0.45–0.47, RMSteg 0.31–0.37, and JARS-Net 0.57–0.75, AGFP achieves extremely low BER values of 0.01–0.12, ensuring reliable message extraction even under distortion. Additionally, AGFP attains the lowest RS, SPA, and SRM detection scores among all evaluated methods, confirming that its attention-guided DWT perturbation strategy produces far fewer statistically detectable artefacts than both classical methods and deep-learning-based steganographic models.

In summary, AGFP achieves a highly favourable and practically meaningful balance between visual imperceptibility, robustness, and statistical invisibility. It outperforms all classical approaches by a wide margin, exceeds modern deep-learning models such as JARS-Net and RMSteg by substantial PSNR and SSIM margins, and approaches the extremely high visual fidelity of iSCMIS while decisively surpassing it in security and reliability. These findings establish AGFP as a powerful and effective steganographic framework ideally suited for real-world applications where undetectability, data integrity, and visual quality must be optimised simultaneously.

5.2 | Robustness Analysis via Bit Error Rate (BER)

The robustness of the proposed AGFP framework was evaluated under varying payloads and multiple distortion scenarios by analysing the BER generated during the extraction process. As shown in Table 7, AGFP consistently achieves exceptionally low BER values ranging from 0.01 to 0.12, regardless of image content or payload. This demonstrates that the attention-guided perturbation of DWT coefficients not only preserves image quality but also ensures stable embedding that resists degradation under typical signal distortions. In comparison, classical methods suffer from significantly higher BER: LSB and PVD exhibit unstable extraction performance due to their sensitivity to even moderate perturbations, while DWT and EMD offer slightly improved robustness but still produce BER values in the 0.21–0.51 range. Histogram shifting similarly fails to maintain extraction reliability at higher payloads, as its histogram modification mechanism introduces cumulative distortions that impair message recovery. These weaknesses reveal a fundamental limitation of classical embedding approaches: their perturbations do not account for the local statistical resilience of the embedding domain, making them vulnerable to noise, filtering, and compression.

AGFP's robustness advantage becomes even more pronounced when compared with recent deep-learning-based steganographic models. iSCMIS, despite delivering the highest PSNR and SSIM values among all evaluated methods, suffers from remarkably high BER values (0.45–0.47), indicating that its invertible neural transformations lack stability under mild distortions. JARS-Net, designed to improve robustness, paradoxically produces the worst BER values in our experiments (0.57–0.75), reflecting its limited resilience to variations in image structure. RMSteg, although more stable than JARS-Net, still exhibits high BER (0.31–0.37), revealing that attention flow and normalising flow architectures remain sensitive to perturbations introduced during embedding or channel distortions. The contrast with AGFP is substantial: AGFP consistently outperforms all these deep-learning models by margins of 4× to 10× lower BER, highlighting its superiority in maintaining message reliability. This robustness stems from AGFP's design: by embedding only into attention-selected DWT regions that are both semantically meaningful and statistically

TABLE 1 | Imperceptibility results (PSNR) for the Aeroplane, Baboon, and Boat images.

	Aeroplane					Baboon					Boat				
Method	**1 Kbs PSNR**	**2 Kbs PSNR**	**3 Kbs PSNR**	**4 Kbs PSNR**	**5 Kbs PSNR**	**1 Kbs PSNR**	**2 Kbs PSNR**	**3 Kbs PSNR**	**4 Kbs PSNR**	**5 Kbs PSNR**	**1 Kbs PSNR**	**2 Kbs PSNR**	**3 Kbs PSNR**	**4 Kbs PSNR**	**5 Kbs PSNR**
AGFP-VGG19	64.20	61.17	59.43	57.51	55.43	64.12	61.12	59.39	58.15	57.22	64.19	61.16	59.41	58.19	57.24
AGFP-AlexNet	64.18	61.20	59.41	57.48	55.43	64.16	61.09	59.34	58.22	57.18	64.16	61.16	59.39	58.24	57.16
AGFP-ResNet50	64.29	61.19	59.44	57.49	55.43	64.20	61.10	59.37	58.23	57.19	64.17	61.17	59.41	58.22	57.21
AGFP-GoogLeNet	64.13	61.18	59.42	57.50	55.45	64.19	61.08	59.33	58.14	57.23	64.16	61.15	59.38	58.15	57.19
iSCMIS	67.47	63.52	61.24	59.77	58.94	68.76	65.14	62.79	61.08	59.88	64.89	62.24	60.12	58.49	57.51
JARS-Net	30.15	27.42	25.63	24.64	23.74	30.54	27.27	25.78	24.57	23.43	30.04	27.21	25.47	24.19	23.57
RMSteg	30.01	26.97	25.16	24.10	23.23	30.56	27.31	25.51	24.01	22.95	30.42	27.19	25.07	24.04	23.22
LSB	51.16	51.16	51.15	51.15	51.15	51.15	51.15	51.15	51.15	51.14	51.16	51.15	51.15	51.15	51.15
PVD	55.06	52.30	49.33	47.11	45.39	54.34	51.68	48.74	46.54	44.83	54.76	52.00	49.03	46.81	45.08
DWT	49.76	46.78	44.97	43.10	41.52	49.33	46.35	44.54	42.67	41.09	49.52	46.54	44.73	42.87	41.29
EMD	53.00	50.11	47.96	46.06	44.40	53.00	50.11	47.96	46.07	44.40	53.00	50.11	47.96	46.07	44.40
HstShift	69.11	68.99	68.79	68.61	68.45	68.79	68.61	68.45	68.30	68.15	69.09	68.96	68.76	68.59	68.42

TABLE 2 | Imperceptibility results (SSIM) for the Aeroplane, Baboon, and Boat images.

	Aeroplane					Baboon					Boat				
Method	**1 Kbs SSIM**	**2 Kbs SSIM**	**3 Kbs SSIM**	**4 Kbs SSIM**	**5 Kbs SSIM**	**1 Kbs SSIM**	**2 Kbs SSIM**	**3 Kbs SSIM**	**4 Kbs SSIM**	**5 Kbs SSIM**	**1 Kbs SSIM**	**2 Kbs SSIM**	**3 Kbs SSIM**	**4 Kbs SSIM**	**5 Kbs SSIM**
AGFP-VGG19	0.9998	0.9997	0.9995	0.9994	0.9992	0.9999	0.9998	0.9997	0.9996	0.9995	0.9999	0.9998	0.9996	0.9995	0.9994
AGFP-AlexNet	0.9998	0.9995	0.9993	0.9991	0.9989	0.9999	0.9999	0.9998	0.9997	0.9996	0.9999	0.9998	0.9997	0.9996	0.9994
AGFP-ResNet50	0.9998	0.9997	0.9995	0.9994	0.9992	0.9999	0.9999	0.9998	0.9997	0.9996	0.9999	0.9997	0.9995	0.9994	0.9993
AGFP-GoogLeNet	0.9999	0.9997	0.9996	0.9995	0.9993	0.9999	0.9998	0.9997	0.9996	0.9995	0.9999	0.9997	0.9996	0.9994	0.9993
iSCMIS	1.0000	1.0000	0.9999	0.9999	0.9999	1.0000	0.9999	0.9999	0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999
JARS-Net	0.9721	0.9553	0.9395	0.9298	0.9203	0.9606	0.9324	0.9147	0.9003	0.8860	0.9714	0.9506	0.9344	0.9203	0.9121
RMSteg	0.9792	0.9578	0.9364	0.9159	0.8961	0.9732	0.9441	0.9093	0.8801	0.8477	0.9789	0.9564	0.9326	0.9114	0.8911
LSB	0.9982	0.9981	0.9981	0.9981	0.9980	0.9992	0.9992	0.9992	0.9991	0.9991	0.9988	0.9988	0.9988	0.9988	0.9987
PVD	0.9985	0.9969	0.9949	0.9932	0.9917	0.9993	0.9986	0.9979	0.9971	0.9960	0.9989	0.9979	0.9968	0.9958	0.9947
DWT	0.9988	0.9973	0.9957	0.9940	0.9924	0.9967	0.9934	0.9901	0.9866	0.9844	0.9984	0.9968	0.9951	0.9931	0.9908
EMD	0.9980	0.9980	0.9980	0.9980	0.9980	0.9991	0.9991	0.9991	0.9991	0.9991	0.9986	0.9986	0.9986	0.9987	0.9987
HstShift	0.9999	0.9998	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

TABLE 3 | Imperceptibility results (PSNR) for the Cameraman, Peppers, and Texture Images.

Method	Cameraman					Peppers					Texture				
	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR
AGFP-VGG19	64.19	61.20	59.43	58.16	57.12	69.14	66.15	64.42	62.48	60.43	64.19	61.17	59.43	57.51	55.43
AGFP-AlexNet	64.19	61.27	59.46	58.13	57.20	69.05	66.06	64.31	62.37	60.32	64.36	61.37	59.63	57.71	55.63
AGFP-ResNet50	64.19	61.19	59.46	58.19	57.23	69.10	66.11	64.38	62.44	60.39	64.29	61.27	59.53	57.61	55.53
AGFP-GoogLeNet	64.23	61.17	59.41	58.17	57.11	69.08	66.09	64.36	62.42	60.37	64.28	61.26	59.52	57.60	55.52
iSCMIS	72.30	68.06	59.01	57.17	56.76	57.45	54.05	52.19	51.16	50.32	64.69	61.03	59.09	57.61	56.66
JARS-Net	30.46	27.50	25.86	24.59	23.96	34.44	31.48	29.80	28.61	27.63	29.89	26.91	25.12	24.03	23.09
RMSteg	30.01	26.86	25.08	24.02	23.17	34.29	30.98	29.06	28.20	27.13	30.02	26.92	25.07	24.22	23.10
LSB	51.10	51.11	51.11	51.12	51.13	51.15	51.15	51.15	51.15	51.14	51.23	51.23	51.22	51.22	51.22
PVD	60.72	56.45	50.69	47.28	45.14	55.45	52.68	49.71	47.50	45.77	45.82	43.14	40.28	38.07	36.34
DWT	52.24	48.79	45.88	45.33	44.89	49.95	46.97	45.15	43.29	41.71	45.91	42.93	41.12	39.25	37.67
EMD	52.99	53.01	53.01	52.99	52.99	53.00	50.11	47.96	46.07	44.40	53.00	50.11	47.96	46.07	44.40
HstShift	69.19	66.63	66.58	66.48	66.69	69.11	68.99	68.79	68.61	68.45	48.23	48.11	47.90	47.73	47.57

TABLE 4 | Imperceptibility results (SSIM) for the Cameraman, Peppers, and Texture images.

Method	Cameraman					Peppers					Texture				
	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM
AGFP-VGG19	0.9998	0.9996	0.9994	0.9992	0.9990	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998
AGFP-AlexNet	0.9998	0.9995	0.9993	0.9990	0.9988	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998
AGFP-ResNet50	0.9997	0.9995	0.9992	0.9990	0.9987	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999
AGFP-GoogLeNet	0.9998	0.9997	0.9995	0.9993	0.9991	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999
iSCMIS	1.0000	1.0000	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000	0.9999	0.9999	1.0000	0.9999	0.9999	0.9999	0.9999
JARS-Net	0.9771	0.9613	0.9484	0.9369	0.9310	0.9976	0.9953	0.9931	0.9909	0.9887	0.9793	0.9611	0.9450	0.9329	0.9203
RMSteg	0.9828	0.9613	0.9435	0.9253	0.9070	0.9979	0.9953	0.9926	0.9908	0.9881	0.9796	0.9569	0.9337	0.9151	0.8922
LSB	0.9977	0.9976	0.9976	0.9975	0.9975	0.9999	0.9999	0.9999	0.9999	0.9999	0.9995	0.9995	0.9995	0.9995	0.9995
PVD	0.9992	0.9984	0.9976	0.9969	0.9961	0.9998	0.9995	0.9993	0.9991	0.9989	0.9992	0.9983	0.9975	0.9965	0.9956
DWT	0.9985	0.9971	0.9963	0.9955	0.9946	0.9988	0.9975	0.9961	0.9947	0.9933	0.9978	0.9953	0.9920	0.9897	0.9866
EMD	0.9975	0.9975	0.9975	0.9975	0.9975	0.9999	0.9999	0.9999	0.9999	0.9999	0.9996	0.9996	0.9996	0.9996	0.9996
HstShift	0.9999	0.9998	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9995	0.9995	0.9995	0.9995	0.9995

TABLE 5 | Imperceptibility results (PSNR) for the Tram, River, and View images.

Method	Tram					River					View				
	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR	1 Kbs PSNR	2 Kbs PSNR	3 Kbs PSNR	4 Kbs PSNR	5 Kbs PSNR
AGFP-VGG19	64.18	61.19	59.43	58.13	57.19	64.24	61.24	59.40	58.13	57.23	64.20	61.17	59.39	58.19	57.19
AGFP-AlexNet	64.12	61.14	59.43	58.17	57.31	64.26	61.31	59.53	58.25	57.26	64.18	61.15	59.38	58.16	57.05
AGFP-ResNet50	64.19	61.12	59.47	58.17	57.21	64.42	61.39	59.62	58.45	57.39	64.27	61.26	59.32	58.19	57.15
AGFP-GoogLeNet	64.16	61.13	59.54	58.23	57.12	64.22	61.24	59.48	58.24	57.24	64.17	61.15	59.43	58.21	57.25
iSCMIS	67.04	63.99	61.64	60.64	59.98	74.62	72.20	69.88	67.91	66.76	68.74	64.38	62.12	60.63	59.75
JARS-Net	29.98	27.18	25.59	24.33	23.70	30.30	27.57	25.80	25.99	25.91	29.90	27.57	26.29	25.76	24.63
RMSteg	30.28	27.13	25.24	23.93	22.99	30.32	27.29	25.76	24.61	23.76	30.01	27.05	25.68	24.50	23.42
LSB	51.14	51.14	51.14	51.15	51.14	50.95	50.94	50.94	50.94	50.94	51.13	51.13	51.13	51.13	51.13
PVD	49.34	46.55	45.16	43.85	42.97	57.03	51.35	49.64	48.49	47.62	48.42	45.60	43.81	42.84	41.82
DWT	47.26	43.92	40.96	38.65	36.93	55.03	50.67	49.01	47.04	45.74	45.21	42.52	40.79	39.92	39.22
EMD	53.01	53.00	52.99	52.95	52.99	53.29	53.28	53.29	53.26	53.29	53.01	53.01	53.01	53.01	53.00
HstShift	51.71	51.70	51.70	51.70	51.70	52.33	52.20	52.20	52.20	52.20	57.59	57.53	57.53	57.52	57.52

TABLE 6 | Imperceptibility results (SSIM) for the Tram, River, and View images.

Method	Tram					River					View				
	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM	1 Kbs SSIM	2 Kbs SSIM	3 Kbs SSIM	4 Kbs SSIM	5 Kbs SSIM
AGFP-VGG19	0.9999	0.9998	0.9997	0.9996	0.9995	0.9998	0.9996	0.9994	0.9992	0.9990	0.9998	0.9996	0.9994	0.9992	0.9991
AGFP-AlexNet	0.9998	0.9997	0.9995	0.9994	0.9992	0.9998	0.9997	0.9995	0.9993	0.9991	0.9998	0.9995	0.9993	0.9991	0.9989
AGFP-ResNet50	0.9999	0.9999	0.9998	0.9997	0.9997	0.9998	0.9995	0.9993	0.9991	0.9988	0.9999	0.9998	0.9997	0.9996	0.9995
AGFP-GoogLeNet	1.0000	0.9999	0.9999	0.9998	0.9997	0.9997	0.9995	0.9993	0.9990	0.9988	0.9999	0.9998	0.9997	0.9996	0.9995
iSCMIS	1.0000	1.0000	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000	0.9999	0.9999	1.0000	1.0000	0.9999	0.9999	0.9999
JARS-Net	0.9708	0.9509	0.9355	0.9216	0.9126	0.9721	0.9592	0.9489	0.9501	0.9495	0.9816	0.9730	0.9729	0.9724	0.9729
RMSteg	0.9756	0.9511	0.9246	0.9009	0.8750	0.9791	0.9593	0.9369	0.9130	0.8906	0.9792	0.9571	0.9394	0.9159	0.8936
LSB	0.9987	0.9987	0.9987	0.9987	0.9987	0.9981	0.9981	0.9980	0.9980	0.9980	0.9981	0.9981	0.9981	0.9981	0.9980
PVD	0.9989	0.9978	0.9964	0.9950	0.9938	0.9991	0.9983	0.9970	0.9955	0.9941	0.9986	0.9973	0.9961	0.9949	0.9940
DWT	0.9975	0.9954	0.9932	0.9901	0.9868	0.9988	0.9973	0.9958	0.9941	0.9925	0.9968	0.9946	0.9918	0.9894	0.9871
EMD	0.9986	0.9986	0.9986	0.9986	0.9986	0.9980	0.9980	0.9980	0.9980	0.9980	0.9979	0.9979	0.9980	0.9979	0.9979
HstShift	0.9997	0.9997	0.9997	0.9997	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9998	0.9998	0.9998	0.9998

TABLE 7 | BER results for the Aeroplane, Baboon, Boat, Cameraman, Peppers, Texture, Tram, River, and View images.

Method	Aeroplane	Baboon	Boat	Cameraman	Peppers	Texture	Tram	River	View
AGFP-VGG19	0.01	0.01	0.01	0.01	0.09	0.01	0.01	0.02	0.02
AGFP-AlexNet	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
AGFP-ResNet50	0.04	0.02	0.02	0.02	0.04	0.02	0.02	0.03	0.03
AGFP-GoogLeNet	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
iSCMIS	0.46	0.47	0.47	0.45	0.46	0.45	0.45	0.46	0.45
JARS-Net	0.70	0.72	0.72	0.69	0.75	0.74	0.72	0.61	0.57
RMSteg	0.35	0.33	0.35	0.35	0.34	0.31	0.34	0.37	0.36
LSB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PVD	0.50	0.49	0.48	0.49	0.50	0.48	0.49	0.51	0.50
DWT	0.22	0.21	0.22	0.17	0.22	0.22	0.17	0.18	0.16
EMD	0.49	0.49	0.49	0.49	0.50	0.49	0.49	0.51	0.50
HstShift	0.51	0.50	0.49	0.49	0.51	0.49	0.50	0.51	0.50

stable, AGFP minimises the likelihood that distortions propagate into the embedded information. Furthermore, DWT perturbation inherently distributes modifications across frequency subbands in a manner that is less affected by common distortions, allowing AGFP to maintain reliable extraction where deep models significantly degrade.

Overall, AGFP demonstrates a uniquely balanced robustness profile, outperforming classical methods through perceptual adaptivity and surpassing deep-learning models by avoiding unstable feature-space embedding. These results establish AGFP as a highly reliable steganographic framework suited for real-world communication scenarios where message integrity is as critical as imperceptibility.

5.3 | Steganalysis Resistance Evaluation

The steganalysis resistance of the proposed AGFP framework was assessed using three widely recognised statistical detection techniques: RS analysis, Sample Pair Analysis (SPA), and Spatial Rich Models (SRM). As summarised in Table 8, AGFP achieves the lowest detection scores among all evaluated classical and deep-learning-based approaches, demonstrating its exceptional ability to preserve the statistical characteristics of the cover image. RS analysis reveals that AGFP's embedding strategy produces minimal changes in pairwise noise characteristics, while SPA results confirm that AGFP's perturbations do not introduce the predictable modification patterns that typically arise in embedding schemes lacking perceptual or statistical awareness. SRM—considered one of the most powerful steganalysis frameworks—further validates AGFP's stealthiness, producing significantly lower feature deviations than those observed in all other methods tested.

Classical steganographic approaches struggle considerably under steganalysis. LSB and PVD produce strong statistical footprints due to their direct manipulation of pixel intensities, leading to high RS and SPA detection rates. DWT, EMD, and HS perform slightly better but still introduce detectable shifts in coefficient distributions or histogram structures. Their failure to incorporate semantic or attention-guided embedding results in perturbations that disrupt natural image statistics, making them highly vulnerable to both classical and machine-learning-based detection mechanisms.

Deep-learning-based steganography methods also show notable weaknesses. Despite its superior visual quality, iSCMIS exhibits clear statistical anomalies in RS and SRM analyses, as invertible neural architectures often introduce non-natural correlations across pixels and frequency components. JARS-Net and RMSteg suffer even more severe detectability issues: both architectures tend to apply dense embeddings across feature maps, generating structured distortions that advanced steganalyzers easily capture. Their resulting RS and SRM scores are significantly higher than those of AGFP, indicating much greater detectability. This highlights a central paradox of recent deep-learning steganography: optimising heavily for reconstruction quality often comes at the cost of increased statistical visibility.

AGFP overcomes these limitations through its selective, attention-driven approach. By embedding only in DWT coefficients that exhibit both perceptual insignificance and statistical stability, AGFP avoids producing detectable artefacts. The distributed and frequency-aware nature of DWT perturbation further reduces the risk of introducing structured patterns that modern steganalyzers exploit. Consequently, AGFP achieves the smallest statistical deviation across all three steganalysis methods, demonstrating a level of stealth unmatched by both classical methods and deep-learning-based frameworks.

In conclusion, AGFP provides a superior security profile by integrating semantic attention with transform-domain embedding, resulting in perturbations that are both visually imperceptible and statistically inconspicuous. Its exceptionally low RS, SPA, and SRM detection rates firmly establish AGFP as the most secure steganographic method evaluated, meeting one of the core requirements for practical and covert communication systems.

TABLE 8 | Steganalysis results for Aeroplane, Baboon, Boat, Cameraman, Peppers, and Texture Images.

Method	Aeroplane			Baboon			Boat			Cameraman			Peppers			Texture			Tram			River			View		
	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM	RS	SPA	SRM
AGFP-VGG19	0.10	6.83	19.73	0.03	9.46	20.41	0.06	9.59	24.89	0.12	7.51	26.69	0.22	54.77	138.84	0.32	18.33	52.64	0.07	10.36	26.83	0.03	4.54	12.66	0.11	6.65	20.64
AGFP-AlexNet	0.10	6.84	19.73	0.03	9.45	20.41	0.06	9.58	24.88	0.12	7.48	26.67	0.22	54.76	138.84	0.32	18.33	52.64	0.07	10.37	26.82	0.03	4.51	12.64	0.10	6.65	20.63
AGFP-ResNet50	0.10	6.84	19.73	0.03	9.46	20.41	0.06	9.59	24.89	0.11	7.53	26.68	0.22	54.77	138.84	0.32	18.33	52.64	0.07	10.32	26.82	0.02	4.55	12.65	0.12	6.59	20.63
AGFP-GoogLeNet	0.10	6.84	19.73	0.03	9.46	20.41	0.06	9.59	24.89	0.12	7.47	26.67	0.22	54.77	138.84	0.32	18.33	52.64	0.08	10.32	26.83	0.03	4.52	12.65	0.12	6.57	20.61
iSCMIS	0.10	6.81	19.72	0.03	9.45	20.41	0.05	9.58	24.90	0.12	7.41	26.66	0.22	54.78	138.85	0.31	18.32	52.64	0.08	10.30	26.81	0.03	4.45	12.61	0.12	6.55	20.61
JARS-Net	0.10	12.68	42.86	0.03	16.20	43.76	0.04	15.79	48.34	0.11	11.76	45.55	0.18	57.34	139.52	0.27	23.93	63.12	0.08	15.94	46.00	0.03	7.74	30.34	0.12	8.32	32.72
RMSteg	0.11	15.10	70.82	0.03	19.44	73.45	0.06	17.89	75.22	0.11	14.03	72.62	0.20	57.158	143.56	0.26	25.92	80.93	0.08	19.23	74.76	0.04	12.07	60.64	0.13	13.80	67.89
LSB	0.16	6.87	19.80	0.06	9.48	20.45	0.09	9.62	24.93	0.20	7.40	26.70	0.29	54.80	138.88	0.34	18.31	52.63	0.13	10.30	26.84	0.09	4.45	12.69	0.20	6.55	20.64
PVD	0.11	6.79	19.75	0.04	9.46	20.44	0.06	9.59	24.91	0.06	6.81	26.40	0.22	54.76	138.85	0.32	18.10	52.20	0.08	9.56	25.87	0.05	4.19	12.51	0.13	5.85	19.48
DWT	0.07	0.85	19.79	0.02	0.98	20.45	0.04	0.99	24.92	0.03	0.85	26.72	0.17	54.77	138.86	0.28	18.31	52.59	0.10	10.23	26.78	0.12	4.43	12.63	0.05	6.50	20.57
EMD	0.09	6.92	19.85	0.03	9.50	20.46	0.05	9.64	24.95	0.05	6.85	26.71	0.21	54.79	138.87	0.32	18.35	52.66	0.06	10.37	26.85	0.03	4.51	12.70	0.08	6.64	20.67
HistShifting	0.10	6.86	19.76	0.03	9.47	20.44	0.06	9.61	24.91	0.05	6.82	26.64	0.22	54.78	138.86	0.32	18.36	52.70	0.07	10.39	27.05	0.03	4.45	12.60	0.12	6.58	20.71

6 | Conclusion

This study introduced AGFP, an Attention-Guided Feature Perturbation framework for high-quality and secure image steganography, combining semantic information extracted from deep convolutional neural networks with the statistical stability of the discrete wavelet transform. By leveraging attention maps from VGG19, ResNet50, AlexNet, and GoogLeNet, AGFP identifies perceptually insensitive and statistically coherent embedding regions, enabling highly controlled and minimally invasive perturbations. The resulting stego-images exhibit consistently strong visual fidelity, with PSNR values ranging from 64.29 to 55.43 dB and SSIM values between 0.9999 and 0.9989 across all payloads and test images.

A comprehensive comparison against both classical and modern steganographic approaches demonstrates the superiority of AGFP in terms of both imperceptibility and robustness. Classical methods such as LSB, PVD, DWT, EMD, and HS showed significantly lower PSNR and SSIM values and exhibited pronounced degradation at higher payloads. In contrast, AGFP maintained stable high-quality performance across diverse image categories, confirming the effectiveness of its attention-guided embedding strategy.

To contextualise AGFP within recent advances, we evaluated it alongside three state-of-the-art deep-learning-based models—iSCMIS [30], JARS-Net [31], and RMSteg [32]. Although iSCMIS achieved higher raw PSNR and SSIM due to its invertible architecture, AGFP outperformed all deep models in practical robustness and security metrics. In particular, AGFP achieved BER values of 0.01–0.12, substantially lower than those of iSCMIS (0.45–0.47), RMSteg (0.31–0.37), and JARS-Net (0.57–0.75). Furthermore, AGFP attained the lowest RS, SPA, and SRM detection scores, confirming that its embedding minimally disrupts natural statistical structures. These findings illustrate that AGFP achieves a superior balance between imperceptibility and detectability—a balance not realised by existing deep-learning frameworks.

The experiments conducted on both the USC-SIPI dataset and the BOSSBase 1.01 benchmark further validated the generalisability of AGFP. Across both datasets, AGFP consistently delivered high visual fidelity and demonstrated exceptional resilience to steganalysis, underscoring its practical suitability for real-world secure communication scenarios. The method's strong performance under varying payloads and its ability to withstand both classical and modern detection attacks highlight its robustness and operational reliability.

In conclusion, AGFP represents a significant advancement in image steganography by bridging semantic attention mechanisms with transform-domain perturbation. It offers an effective, secure, and perceptually adaptive alternative to existing classical and deep-learning-based methods. Future work may focus on extending AGFP to multi-modal attention models, improving robustness under extreme distortions, and integrating transformer-based semantic extraction to further enhance embedding precision and security. Overall, AGFP establishes a promising foundation for next-generation steganographic systems where imperceptibility, robustness, and undetectability must be jointly optimised.

7 | Future Work

For future work, several promising directions can be explored to further enhance the AGFP framework:

7.1 | Adaptive Payload Allocation

Investigate dynamic payload allocation strategies where the embedding capacity for each image is adaptively determined by its content complexity and the attention map, rather than a fixed bit rate. This could further optimise the balance between imperceptibility and capacity.

7.2 | End-to-End Trainable DWT Integration

Explore the development of a fully differentiable DWT layer that can be integrated directly into the neural network architecture, allowing the DWT-based embedding and extraction processes to be optimised end-to-end through backpropagation. This could potentially lead to even more nuanced and robust embedding strategies.

7.3 | Robustness Against Geometric Attacks and Advanced Steganalysis

Extend the evaluation to include robustness against geometric attacks (e.g., rotation, scaling, and cropping) and more sophisticated steganalysis techniques based on deep learning. This would provide a more comprehensive understanding of AGFP's real-world applicability.

7.4 | Extension to Video Steganography

Adapt the AGFP framework for video sequences, leveraging temporal coherence and motion information to identify optimal embedding locations across frames, opening new avenues for high-capacity and robust covert communication in multimedia.

Author Contributions

Taner Cevik led the study and developed the proposed methodology. Nazife Cevik and Ali Pasaoglu contributed to method development and experimental implementation. Fatih Sahin and Farzad Kiani performed comparative analysis and evaluation. Muhammet Sait Ag contributed to data processing and manuscript preparation. All authors reviewed and approved the final manuscript.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

This study did not generate any new datasets. The experiments were conducted using publicly available benchmark image datasets, including the USC-SIPI dataset and the BOSSBase 1.01 dataset. All datasets used in this work are openly accessible and can be obtained from their respective repositories. No proprietary or confidential data were used, and therefore data-sharing restrictions do not apply.

References

1. N. F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen," *Computer* 31, no. 2 (1998): 26–34, https://doi.org/10.1109/MC.1998.4655281.

2. J. Fridrich, M. Goljan, and R. Du, "Detecting LSB Steganography in Color, and Grayscale Images," *IEEE Multimedia* 8, no. 4 (2001): 22–28, https://doi.org/10.1109/93.959097.

3. N. Provos, "Defending Against Statistical Steganalysis," in *Proceedings of the 10th USENIX Security Symposium*, (Usenix Association, 2001), 323–333.

4. I. J. Cox, M. L. Miller, and J. A. Bloom *Digital Watermarking* (Morgan Kaufmann, 2002).

5. C. K. Chan and L. M. Cheng, "Hiding Data in Images by DWT," *International Journal of Digital Content Technology and its Applications* 2, no. 1 (2004): 1–10.

6. S. Baluja, "Hiding Images in Plain Sight: Deep Steganography," in *Advances in Neural Information Processing Systems* 30 (Curran Associates, Inc., 2017), 3086–3095.

7. S. Wu and L. Li, "A Review of Deep Learning-based Image Steganography," *Journal of Visual Communication and Image Representation* 70 (2020): 102831.

8. Z. Tang, J. Ni, and R. Zhong, "Image Steganography Based on Deep Learning," *Multimedia Tools and Applications* 76, no. 21 (2017): 22807–22822.

9. A. Vaswani, N. Shazeer, and N. Parmar, et al. "Attention Is All You need," in *Advances in Neural Information Processing Systems* 30 (Curran Associates, Inc., 2017): 5998–6008.

10. M. Liu, X. Zhang, and Y. Wu, "Attention-guided Generative Adversarial Network for Image Steganography," *IEEE Transactions on Information Forensics and Security* 15 (2020): 2379–2391.

11. L. Ma, X. Cao, M. Fang, and K. Yang, "Deep Learning for Image Steganography and Steganalysis: A Survey," *Information Sciences* 514 (2020): 219–242.

12. Y. Qian, T. Meng, J. Wang, and M. Li, "Robust Image Steganography via Deep Learning With Attention Mechanism," *Journal of Network and Computer Applications* 166 (2020): 102717.

13. D. C. Wu and W. H. Tsai, "A Steganographic Method for Images by Pixel-value Differencing," *Pattern Recognition Letters* 24, no. 9-10 (2003): 1613–1626.

14. X. Zhang and S. Wang, "Efficient Steganographic Embedding by Exploiting Modification Direction," *IEEE Communication Letters* 10, no. 11 (2006): 781–783.

15. Y. S. Mehran, M. Nafari, A. Nafari, and N. Z. Mehran, "Histogram Shifting as a Data Hiding Technique: An Overview of Recent Developments," *Communications in Computer and Information Science* 166 (2011): 770–786.

16. A. Westfeld, "F5-A Steganographic Algorithm," *Lecture Notes in Computer Science* (2001): 289–302.

17. Y. Wang, X. Zhang, and Q. Li, "A Survey on Image Steganalysis Techniques," *Journal of Visual Communication and Image Representation* 58 (2019): 257–272.

18. P. H. Thien and J. C. Lin, "A Simple and High-capacity Method for Hiding Data in Raw Images," *Signal Processing* 83, no. 10 (2003): 2207–2215.

19. Y. Liu and R. Wei, "An Image Steganography Method Based on DWT-SVD," *Multimedia Tools and Applications* 74, no. 24 (2015): 11451–11467.

20. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for Data Hiding," *IBM Systems Journal* 35, no. 3–4 (1996): 313–336.

21. S. Singh and R. Singh, "A Robust Image Steganography Scheme Based on Contourlet Transform," *International Journal of Computer Applications* 61, no. 10 (2013): 19–24, https://doi.org/10.5120/14008-2057.

22. A. D. Ker and T. Pevný, "Machine Learning for Steganalysis," *IEEE Signal Processing Magazine* 33, no. 3 (2016): 64–77.

23. W. Tang, Z. Qu, Y. Li, and R. Li, "A Novel Image Steganography Method Based on Deep Neural Network," in *Proceedings of the 2nd IEEE International Conference on Computer and Communications (ICCC)*, (IEEE, 2016): 2003–2007.

24. S. Baluja, "Hiding Images in Plain Sight: Deep Steganography," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), 3086–3095.

25. X. Zhang, Y. Zhao, and J. Ni, "Adversarial Learning for Image Steganography," *IEEE Transactions on Information Forensics and Security* 14, no. 11 (2019): 2954–2965.

26. J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-image Translation Using Cycle-consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision*, (2017): 222–232.

27. K. Zhang, Z. Liu, Y. Ma, and Z. Li, "Deep Learning for Image Steganography: A Survey," *Future Generation Computer Systems* 94 (2019): 696–709.

28. X. Liu, X. Zhang, and Y. Ma, "Deep Hiding: A Deep Neural Network Based Steganography," *Journal of Network and Computer Applications* 129 (2019): 29–37.

29. H. Li and S. Yu, "A Comprehensive Survey of Deep Learning-based Image Steganography and Steganalysis," *IEEE Access* 8 (2020): 142169–142194.

30. F. Li, Y. Sheng, X. Zhang, and C. Qin, "iSCMIS:Spatial-Channel Attention Based Deep Invertible Network for Multi-Image Steganography," *IEEE Transactions on Multimedia* 26 (2024): 3137–3152, https://doi.org/10.1109/TMM.2023.3307970.

31. L. Zhang, Y. Lu, T. Li, and G. Lu, "Joint Adaptive Robust Steganography Network," *IEEE Transactions on Industrial Informatics* 20, no. 8 (August 2024): 10156–10166, https://doi.org/10.1109/TII.2024.3388674.

32. H. Ye, S. Zhang, and S. Jiang, et al., "Robust Message Embedding via Attention Flow-Based Steganography," in *Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2025): 12840–12849, https://doi.org/10.1109/CVPR52734.2025.01198.

33. A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*. Advances in Neural Information Processing Systems 30 (2017).

34. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (arXiv, 2018), 7132–7141.

35. S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, (arXiv, 2018), 3–19.

36. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, (arXiv, 2021).

37. P. Xu, X. Cao, T. Zhang, and Z. Zhao, "Deep Steganography With Attention Mechanism for Robustness," *IEEE Access* 10 (2022): 10091–10101.

38. Y. Zhang, P. Dong, Z. Zhang, and S. Cao, "Attention-aware Feature Perturbation for Robust Deep Image Steganography," *Multimedia Tools and Applications* 82, no. 12 (2023): 17769–17789.

39. "The USC-SIPI Image Database," Accessed November 4, 2025. https://sipi.usc.edu/database.

40. P. Bas, T. Filler, and T. Pevný, "'Break Our Steganographic System': The Ins and Outs of Organizing BOSS," in *Information Hiding. IH 2011, eds*. T. Filler, T. Pevný, S. Craver, and A. Ker, Lecture Notes in Computer Science (Springer, 2011), https://doi.org/10.1007/978-3-642-24178-9_5.