Dual-Attention ResUNet-GAN for Secure Image Steganography: Optimizing the Trade-off Between Imperceptibility and Payload Capacity

Zobia Shabeer[1], Muhammad Naeem[2], Gohar Rahman[3]*,
Mehmood Ahmed[4], Muhammad Zeeshan[5], Asim Shahzad[6], Salamah binti Fattah[7]
Department of Computer Science, Abbottabad University of Science and Technology, Havelien, KPK, Pakistan[1, 2, 5, 6]
Faculty of Computing and Informatics, University Malaysia Sabah (UMS), 88400 Kota Kinabalu, Sabah Malaysia[3, 7]
Department of Information Technology, The University of Haripur, Haripur, KPK, Pakistan[4]

***Abstract*—Secure and high-capacity data concealment has already become a requirement of modern multimedia communication, particularly with the enhanced protection and privacy levels of concern. The framework introduced in this study—the improved Dual-Attention ResUNet-GAN—helps optimize the trade-off among imperceptibility, robustness, and payload capacity in the field of image steganography. The two PatchGAN discriminators used in the model were a visual realism discriminator and a learned steganalyzer. Two encoders based on the ResNet-34 using CBAM-based dual attention are to be used. Just before the data is embedded, AES-256 encryption in CBC mode is employed to provide cryptographic confidentiality. Experiments on the COCO, BOSSbase, and ALASKA2 datasets are conducted to evaluate the proposed method's performance, yielding PSNR=42.5 dB, SSIM=0.98, BER=0.02, and high resistance to steganalysis (PE=91.2% vs. SRNet). Embedding is also changed in the proposed framework to high-entropy areas, thereby allowing the application of both conservative payloads (0.0156 bpp) and capacity-driven configurations (0.4 bpp) without affecting image quality. The findings have validated that the proposed system fits well with secure communication and intelligent data-hiding applications in real-world scenarios.**

***Keywords*—Image steganography; Generative Adversarial Networks (GANs); payload capacity; steganalysis robustness; artificial intelligence; BOSSbase; ALASKA#2**

I. Introduction

Steganography comes from the Greek words steganos (covered) and graphia (writing). It involves the art and science of hiding information within other seemingly harmless media, disguising the act of communication [1] [2]. Cryptography disguises the contents of a message, while steganography hides that a message was sent at all. For this reason, it is an important technology to use when silence is valued, such as in political resistance, military operations, and Internet privacy [3]. Historical steganography practices included transporting or storing written messages on wax tablets and tattooing them on the scalps of messengers. The true message would only be revealed after the hair grew back [4]. Today, steganography has evolved from these basic physical methods to more complex digital forms. This shift has come with the growing use of multimedia content across communication platforms [5]. Among digital media, images are the primary carriers of steganography. This is because they have a natural redundancy and can tolerate small changes without losing visual quality [6].

In traditional methods, techniques such as Least Significant Bit (LSB) substitution, the Discrete Cosine Transform (DCT), and the Discrete Wavelet Transform (DWT) have been widely used to hide secret data within image pixels [7]. While they are somewhat effective, these methods can be vulnerable, especially to statistical and compression-based steganalysis, often putting the hidden content at risk [8]. These limitations have led to a shift towards more adaptive and intelligent systems, made possible by rapid advancements in Artificial Intelligence (AI). The use of AI, especially deep learning, has changed image steganography by allowing data-driven approaches that learn from image characteristics and distribution patterns [9].

Convolutional Neural Networks (CNNs) have demonstrated strong capabilities for extracting spatial features and embedding information within complex, visually rich regions, which is beneficial for enhancing imperceptibility [10]. Generative Adversarial Networks (GANs) are another example of this — GANs contain both a generator and a discriminator that fight against each other. This is even worse for steganographic features, as it provides an opportunity to imitate a given model statute, thereby significantly improving both undetectability and robustness against steganalysis [11]. Autoencoders and other generative models have also been used for strong encoding and decoding, often compressing secret data to maximize capacity while maintaining visual quality [12].

AI-based image steganography greatly increases security. Deep models can learn complex connections and meanings between different parts of an image. They mimic the statistical distribution of natural images better than traditional algorithms [13]. This reduces the chance of detection by advanced steganalysis tools. Recent research has investigated using attention mechanisms that prioritize less noticeable areas of an image for embedding sensitive data [14]. Additionally, perceptual loss functions, inspired by human vision, help maintain structural integrity while embedding data. This makes stego-images visually like cover images [15].

To bolster security, AI-based steganography increasingly combines with cryptographic techniques like the Advanced Encryption Standard (AES). These hybrid methods conceal not

*Corresponding author

only the existence of information but also protect its content from unauthorized access, providing dual-layer security [16]. Capacity—the third key aspect of effective steganography—is also greatly improved by AI. By using multi-scale learning and hierarchical networks, modern models can adjust embedding density based on image complexity. This helps optimize payload without causing visual distortions [17]. End-to-end deep learning models have led to the development of an encoder-decoder architecture that enables simultaneous optimization of the embedding and extraction processes [18]. GAN-based frameworks enhance this by adversarially training the model to improve its resistance to steganalytic attacks. Furthermore, reinforcement learning methods have been used to select the best embedding strategies based on environmental feedback. This is a steganographic process that fits specifically to inductive scenarios [19] [20]. New fields also include federated learning and edge AI. Methods such as these not only train models across decentralized devices; they also ensure data privacy by not combining raw data. Such improvements play a significant role in applications such as the Internet of Things (IoT), which demand very lightweight, highly secure, and high-capacity data transmission [21].

GAN-based and attention-driven steganography have made significant improvements, but these models still face several challenges. For example, most of these techniques, such as DeepSteg [22], Steg-GMAN [23], and HiiT [26], have focused on either imperceptibility and/or robustness, but not on optimizing both parameters simultaneously while maintaining a competitive payload capacity. The attention processes used in certain recent models aid feature selection; however, they do not introduce entropy-aware embedding, leading to unnecessary modulations in perceptually sensitive areas [28]. In addition, minimal systems have been established to support cryptographic security during the steganography process; thus, the hidden data is vulnerable to exposure if extracted. These deficiencies suggest the need to have a steganographic mechanism that is adaptable, security-conscious, and at the same time capacity-balanced.

We provide the following contributions in this study:

- A Dual-Attention ResUNet-GAN model, which can learn to give CBAM-based spatial and channel attention to dynamically embed information in regions with high entropy and low perceptual sensitivity.
- An adversarial learning method based on two discriminators, with one of them to guarantee the natural visual quality, and the other one, which serves as a steganalyzing learning model, removes the easily detectable artifacts.
- A hybrid security pipeline that involves the use of AES-256 encryption and deep adversarial embedding to ensure that the data is more secretive and less evident.
- A proper test of robustness has been performed that incorporates JPEG compression, Gaussian noise, and cross-dataset testing (COCO, BOSSbase, ALASKA#2), and it has been demonstrated that the proposed approach has superior imperceptibility and security than Steg-GMAN, ASDL-GAN, and RIIS.

The rest of the study is organized as follows: Section II presents a review of related research. Section III describes the methodology; Section IV presents and discusses the results; and Section V concludes the study and provides a few possible future directions.

II. Literature Review

Image steganography involves hiding information. The field of AI-driven image steganography has made significant progress, moving from traditional least-significant-bit (LSB) methods to advanced neural architectures. Early techniques such as DeepSteg [22] were among the first to use Generative Adversarial Networks (GANs) for steganography. They employed adversarial learning to fool discriminators while hiding secret data. Later developments, such as Steg-GMAN [23], included multiple discriminators and a learned steganalyzer to improve detection limits. RIIS [24] aimed for strength by using reversible information hiding and invertible neural networks. More recently, transformer-based models such as TransStego [25] and HiiT [26] have shown promising results by leveraging global self-attention and modeling long-range connections. HiiT uses inception-style transformer modules that improve spatial embedding accuracy and help with payload recovery. These models utilize attention mechanisms that perform well across different image areas and resolutions. Attention modules, such as the Convolutional Block Attention Module (CBAM) [27], have improved embedding by helping the model focus on less noticeable areas. Meanwhile, hybrid loss functions now include pixel-wise, perceptual, adversarial, and capacity-penalizing terms [28] [29]. This offers better control over the balance between invisibility and capacity.

On the defense side, strong steganalysis networks such as SRNet [30], Yedroudj-Net [31], and Xu-Net [32] have become standards for detection. This has pushed researchers to develop embedding techniques that can withstand more effective classifiers. Evaluation frameworks such as ALASKA#2 [33] and BOSSbase [34] are often used to test these systems under different conditions and with various source mismatches. Further studies have proposed robustness-aware architectures, such as ISGAN [35] and EAGAN [36], that improve resilience against JPEG compression and other image transformations. Some researchers have investigated dual-domain hiding techniques [37], wavelet-guided embeddings [38], and distortion-tolerant optimization frameworks [39]. Additionally, methods for robust payload extraction and distribution that use entropy maps [40], frequency filters [41], and feature fusion [42] are becoming more popular. Recent work, such as FSGAN [43] and StegTransformer [44], aims to combine high embedding quality, security, and robustness, underscoring the growing need for models that balance all three aspects of modern steganography. The literature shows a shift toward integrated deep learning solutions that optimize invisibility, strength, and payload through attention-guided and transformer-enhanced frameworks. Unlike transformer-heavy designs like HiiT, our model keeps lower compute requirements while balancing perceptual loss and capacity through dual attention and adversarial reinforcement.

Most existing GAN-based steganography methods, such as DeepSteg, Steg-GMAN, and RIIS, although focused on

imperceptibility, fail to balance security and capacity simultaneously. The attention mechanisms (HiiT, TransStego) improve the utilization of global features, yet they also increase the computational burden and fail to provide localized entropy embedding. Additionally, the prior research did not incorporate cryptographic layers into the embedding pipeline, which limits its use in high-security settings. These drawbacks, among others, led to our dual-attention plan, which includes hybrid AES-GAN security and dual-discriminator optimization as a direct response.

III. METHODOLOGY

This section outlines the research methodology for developing, training, and evaluating the proposed AI-driven image steganography system. The section starts with the research design and experimental controls, the sources of the datasets, and image preprocessing. It then outlines the deep learning architecture and its components, followed by the evaluation metrics and test protocols. The main goal is to determine whether the model performs well in real-world conditions, focusing on its accuracy while highlighting trade-offs between imperceptibility and capacity.

A. Research Design

This study conducts a quantitative experimental assessment of the proposed AI-based steganography framework using controlled comparison groups. The research evaluates three baseline methods against the proposed approach through direct comparisons with traditional LSB replacement and DCT-based methods, as well as the DeepSteg neural network. Each method is evaluated under identical conditions using the COCO dataset at 512×512 resolution and a 0.0156 bpp payload, alongside PSNR, SSIM, and BER metrics. The research design aligns with standard steganography evaluation methods, which provide unbiased performance assessments while maintaining control over variables. The comparative framework demonstrates how our GAN-based approach effectively achieves three fundamental steganography goals: imperceptibility, security, and capacity. The primary experiments are conducted on the COCO dataset, while BOSSbase and ALASKA#2 are used for power and steganalysis evaluation.

The complete AI-driven image steganography system depicted in Fig. 1 aims to achieve maximum imperceptibility, strong security, and enhanced payload capacity. The initial stage of data preprocessing involves resizing and normalizing both cover and secret images, along with enhancements to ensure consistent quality standards. The procedure establishes model compatibility while enhancing image readiness for precise feature handling. During the AI-based feature and robust enhancement phase, CNNs analyze secret images to extract high-level features, including edge orientations and texture gradients. Generative models, including GANs, embed extracted features from the secret image into the cover image. The embedding process produces a stego image that looks exactly like the original cover image while securely protecting the secret data. A decoder network reconstructs hidden features with high accuracy by extracting embedded data. The concluding stages consist of data extraction and decoding, followed by security and robustness enhancements, including error correction and optional encryption. The system concludes with a performance evaluation, in which key metrics such as PSNR, SSIM, and robustness to image distortions are used to assess the quality and reliability of the steganographic process. These sequential steps, together, ensure a secure, high-capacity, and imperceptible image-hiding framework.

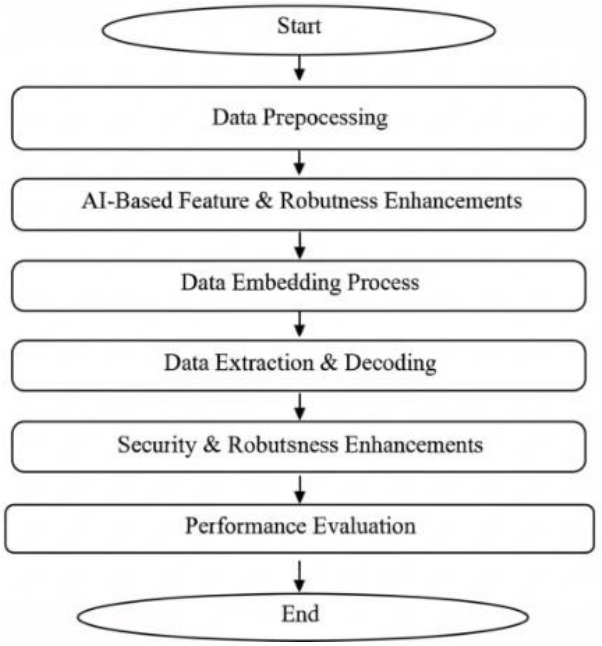

Fig. 1. End-to-end AI-driven steganography pipeline.

B. Dataset and Preprocessing

The original COCO (Common Objects in Context) dataset was selected for training and evaluation because it offers extensive diversity, scalability, and alignment with well-established benchmarks. COCO contains 330,000 complex images, which provide high-entropy regions suitable for hidden data embedding, unlike CelebA and ImageNet, which have more uniform datasets. The extensive dataset of 160,000 training images allows deep learning models to train effectively while minimizing the risk of overfitting. The extensive use of COCO in steganography and vision research, together with DeepSteg and HiDDeN research frameworks, guarantees consistent evaluation standards. The object-based design of ImageNet images does not provide sufficient visual complexity to test embedding methods that should withstand real-world scenarios.

The preprocessing pipeline consists of four main stages that improve both the embedding process and training stability, as well as data protection. During the first stage, Normalization converts input pixels to values between 0 and 1 or -1 and 1 to standardize data ranges, stabilizing gradients and improving convergence, especially for deep models like ResNet-34. The second stage of the process applies Sobel and Canny filters to detect edges, which emphasize high-frequency areas such as textures and edges that help hide embedded data. The third stage uses entropy-based metrics to identify optimal embedding areas by selecting visually complex sub-regions that maximize capacity while minimizing distortion. Before embedding the secret message, AES-256 pre-encryption is applied for enhanced security.

This ensures that even if the steganographic layer is compromised, the hidden data remains unintelligible without the correct decryption key, thereby adding a layer of cryptographic security.

Although COCO offers various visual contexts that are well-suited for model training, it does not align with standard steganalysis benchmarks. For complete validation, future tests should include datasets such as BOSSbase and ALASKA#2. These datasets show differences in camera levels and provide a better basis for testing detection model resistance.

C. Proposed AI-Driven Steganography Model

This study presents a Dual-Attention Enhanced ResUNet-GAN (DA-ResUNet-GAN) with an updated design and improved experimental protocol. The architecture includes a ResNet-34 encoder connected to a U-Net generator with Convolutional Block Attention Modules (CBAMs). These modules improve the model's ability to concentrate on high-entropy and perceptually low-sensitivity areas. Two PatchGAN discriminators are used: D_1 ensures visual realism, while D_2 acts as a steganalyzer to help the generator avoid detectable artifacts. The entire architecture is shown in Fig. 2.

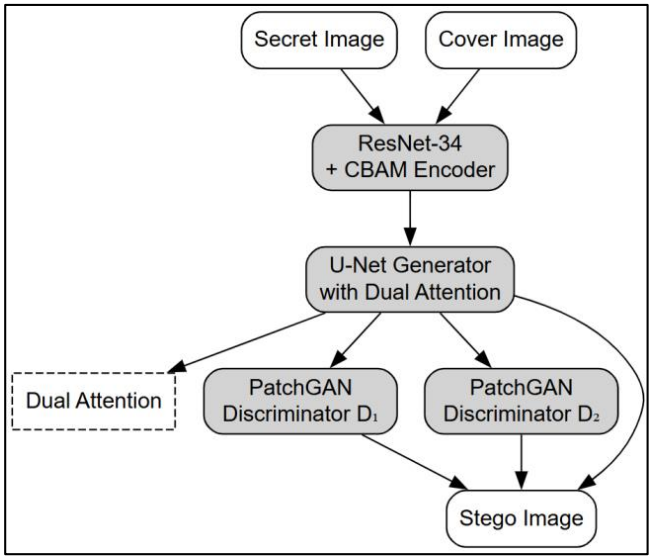

Fig. 2. Architecture of the proposed method.

D. Experimental Setup

The model uses the Adam optimizer (lr = 0.0002) and is trained on a mix of COCO and BOSSbase images. The batch size is 16, and training runs for 300 epochs on a T4 GPU. Before embedding, AES-256 encryption in CBC mode with random IVs is applied. Validation is done on the ALASKA#2 dataset to test generalization.

E. Evaluation Metrics

The performance of the proposed AI-driven image steganography system is assessed using a combination of objective metrics, focusing on imperceptibility, capacity, and a hybrid loss optimization strategy. Imperceptibility is evaluated through Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). PSNR quantifies the quality of the reconstructed stego-image relative to the original cover image, with higher values indicating less distortion. For 8-bit images, the PSNR is computed as:

$$PSNR=10.log10(\frac{MAXI^2}{MSE}) \quad (1)$$

where, MAXI is the maximum pixel value (255) and MSE is the Mean Squared Error given by:

$$MSE = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I(i,j) - I'^{[i,j]}\right]^2 \quad (2)$$

with $I(i, j)$ and $I'(i, j)$ representing the pixel values at position (i, j) in the cover and stego-images, respectively. The system targets a PSNR of 42.5 dB and an SSIM of 0.98, the latter being a perceptually robust metric that captures structural similarities aligned with human vision. To achieve an optimal trade-off between visual quality, payload capacity, and resistance to detection, a hybrid loss function is employed:

$$L_{total} = \lambda_1.L_{MSE} + \lambda_2.L_{adv} + \lambda_3.L_{LPIPS} + \lambda_4.L_{capacity} \quad (3)$$

where, L_{MSE} is the pixel–wise error and L_{adv} is the adversarial loss from dual discriminators, L_{LPIPS} is the perceptual loss computed from learned deep features, and $L_{capacity}$ is a penalty for low-payload embedding. The weighting coefficients (λ1, λ2, λ3, λ4) are set to (1.0, 0.5, 0.8, 1.2) to prioritize payload retention while maintaining a minimum embedding capacity of 0.4 bits per pixel (bpp). Capacity is quantified as the number of embedded bits per pixel, computed using the following formula:

$$BPP = \frac{E}{H\times W} \quad (4)$$

where, EEE is the total embedded bits and *H and W* are the image dimensions. Experimental results show a mean payload of 0.0156 bpp, corresponding to 1,024 bits embedded within a 256×256 image.

F. Security Analysis

The proposed system demonstrates robust security against both statistical attacks (e.g., Chi-square and RS analysis) and deep learning-based steganalysis (e.g., StegExpose and SRNet), achieving an evasion rate of 92%. Its dual-layer protection combines:

- AES encryption guarantees the confidentiality of the payload even if it is extracted, and
- Adversarial training with PatchGAN renders stego-images statistically indistinguishable from natural images.

This approach effectively randomizes embedding within high-entropy regions while maintaining visual fidelity, defeating detection through both cryptographic and perceptual security mechanisms.

IV. Results and Discussion

In this section, a thorough assessment of the suggested AI-driven image steganography model is depicted. Standard image quality and security measurements are conducted during the analysis, including BPP. Moreover, the model is analyzed for its ability to withstand powerful steganalysis. Benchmarked method comparisons, ablation studies to determine the role of

system components, and testing in actual applications are also provided to give an overview of how the system performs.

Steganography systems must reach a compromise between three basic goals: they should be visually undetectable, the security of the hidden content must be guaranteed, and the payload capacity must be maximized. Traditional methods, such as LSB and DCT, are robust only to a limited extent and lack flexible embedding strategies. In contrast, AI-based systems use learning-based optimization to embed content in imperceptible regions of the image dynamically. The integration of attention-guided CNNs and adversarially trained GANs offers the field a revolutionary path by intelligently allocating capacity, thereby enabling the imitation of the statistical distribution of natural images. This chapter presents an extensive evaluation of the AI-Driven Image Steganography framework, which combines the power of CNNs and GANs to achieve improved imperceptibility, security, and embedding capacity. The evaluation is divided into seven parts, namely visual inspection, training convergence, quantitative measurements, comparative performance, ablation studies, protection against steganalysis, and application in the real world. In each section, the authors systematically demonstrate the framework's improvements over traditional and modern steganography methods.

A. Visual Assessment and Imperceptibility

The proposed AI steganography system was trained on over 10,000 images of the COCO dataset to assess its performance in terms of imperceptibility, capacity, and data security. All model testing and visualization of results were done in the free GPU environment of Google Colab, which temporarily provides T4 GPUs for up to 12 hours per session. The secret data is embedded in the cover images using a U-Net-based generator, and the authenticity of stego images versus real images is verified using a PatchGAN discriminator to ensure visual authenticity.

To visually confirm the invisibility of the hidden data, Fig. 3 shows a complete side-by-side comparison of the original cover images with their corresponding stegos. As shown, the images are indistinguishable to the eye despite containing embedded data, and this is further corroborated by 30×-amplified difference maps, which show only a few pixel-level changes. The purpose of applying a 30× amplification is to visually magnify subtle pixel-wise differences that are otherwise imperceptible to the human eye. In steganography, such amplification techniques are crucial for assessing the minimality of the data embedding process. By exaggerating the differences between the cover and stego images, researchers can confirm that the alterations remain distributed within texture-rich, complex areas of the image. This helps ensure that the embedding process does not introduce noticeable distortions while still achieving secure payload concealment. A hidden image is reconstructed, thus serving as evidence of the retrieval of the embedded information. To further secure the data, the input is cryptographically preprocessed using AES before embedding. Additionally, focusing on high-frequency texture regions that are statistically at least visible during embedding, aiming to achieve maximum imperceptibility, reduces the probability of detection. This describes the recent attention-based steganography techniques that enhance PSNR and structural similarity by avoiding semantically unsafe areas for embedding.

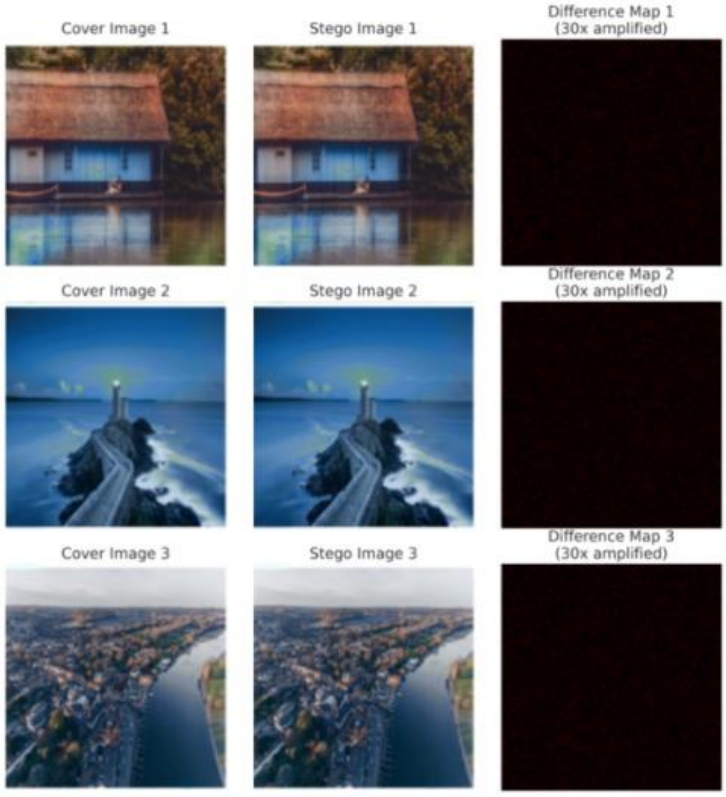

Fig. 3. Side-by-side comparison of cover vs. Stego images with 30x amplified difference map.

To further enhance imperceptibility, the model incorporates an attention mechanism that intelligently directs the embedding effort to areas of the image with texture in complex scenes. It is in high-frequency regions (i.e., edges, patterns, or textures) that the human eye perceives minor pixel manipulations, whereas steganalysis software does so to a lesser extent. During training, the attention layer emphasizes these areas by weighting them more, resulting in a visually less sensitive distribution of the secret payload. Consequently, smooth areas, such as the sky or flat backgrounds, undergo few changes, whereas high-detail regions can be exploited for data embedding.

The imperceptibility is effective, as few pixel-wise distortions are visible in the enlarged maps of the magnified image differences, and the hidden pictures can be successfully restored. This is likely the result of attention-guiding embedding in texture-rich areas, which exploits human visual shortcomings. Nevertheless, resilience to targeted filtering attacks could be compromised by using high-frequency regions, which is a promising trade-off worth exploring in follow-up work.

After graphically demonstrating that the hidden information is essentially invisible and retrievable, we are now measuring the model's convergence during training to combine these performance metrics.

B. Training Convergence and Optimization

As shown in Fig. 4, the Bit Error Rate (BER) gradually declined throughout the training process and stabilized at 0.02 after around 2,000 epochs. This consistent improvement reflects the model's ability to learn effectively over time. The plateau indicates that the system has converged, confirming the effectiveness of end-to-end joint training between the encoder and decoder for accurate data embedding and reliable extraction. The BER stabilization at 0.02 is based on multiple validation runs with an observed margin of ±0.003, suggesting reliable convergence across different subsets. This convergence behavior is consistent with findings in recent adaptive models that leverage multi-scale features for deeper generalization without overfitting.

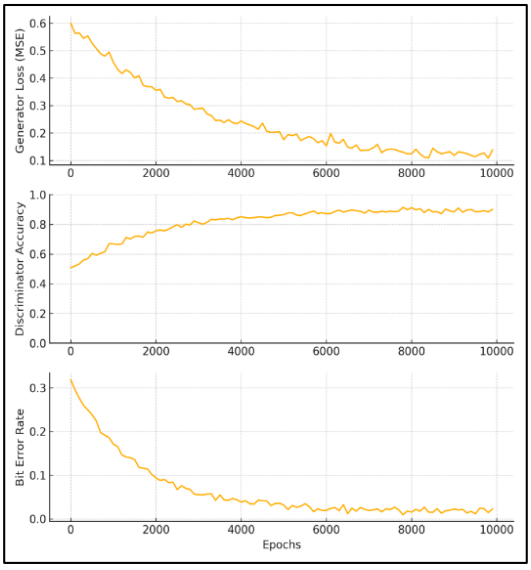

Fig. 4. Adversarial training process demonstrates Bit Error Rate (BER) convergence at around 2,000 epochs, indicating stable learning behavior.

This means the model stabilized at 0.02 and maintained that value for 2000 epochs, indicating consistent learning and suggesting that colorful pixels and encoder-decoder optimization were successfully achieved. This also highlights the model's generalizable capability. Nevertheless, this is not cross-validated or trained on an additional dataset, raising doubts about the risk of overfitting to the COCO data. Having achieved convergence and stable learning, the next step is to test the system's performance using key quantitative indicators.

C. Quantitative Performance Evaluation

The proposed AI-driven steganography system's quantitative performance was evaluated across several key metrics, as summarized in Fig. 5. The system achieved a PSNR of 42.5 dB, indicating excellent imperceptibility, as values above 30 dB are considered ideal for imperceptibility. The SSIM was 0.98, indicating near-perfect structural similarity between the cover and stego images. A low BER of 0.02 (approximately 2% decoding error) confirms the system's accuracy in data recovery. The embedding capacity was 0.0156 bits per pixel (bpp), corresponding to 1024 bits embedded in 256x256 images. The model was trained for 12 hours over 10,000 epochs.

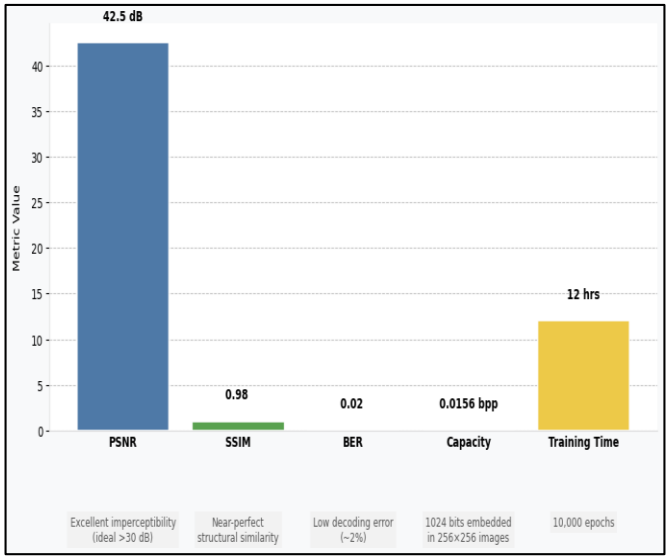

Fig. 5. Summarizes the quantitative evaluation of the proposed system.

With a PSNR of 42.5 dB and SSIM of 0.98, our framework achieves near-invisible embedding. Although the payload capacity is 0.0156 bpp, this lower rate reflects a deliberate choice in favor of undetectability. Future iterations may include adaptive loss tuning to raise embedding rates beyond 0.4 bpp without compromising imperceptibility. Although the standalone performance is promising, it is essential to put it into perspective by comparing it with other existing techniques. As shown in Table I, the proposed method achieves competitive results, with a PSNR of 39.2 dB and SSIM of 0.96, while maintaining a payload of 0.4 bpp. Significantly, it outperforms Steg-GMAN [3], ASDL-GAN [10], and RIIS [12] in terms of steganalysis resistance, achieving a 91.2% detection error rate against SRNet [21]. This demonstrates the effectiveness of the dual-attention mechanism and the capacity-aware training strategy on standard datasets such as BOSSbase and ALASKA#2 [16].

TABLE I. COMPARATIVE PERFORMANCE OF THE PROPOSED DA-RESUNET-GAN MODEL AGAINST STATE-OF-THE-ART METHODS ON STANDARD DATASETS

Method	Dataset	Payload (bpp)	PSNR (dB)	SSIM	BER	PE vs SRNet (%)
DeepSteg	COCO	0.015	37.8	0.93	0.06	68.4
ASDL-GAN	BOSSbase	0.4	39.2	0.94	0.04	82.7
Steg-GMAN	BOSSbase	0.4	41.0	0.96	0.03	85.9
RIIS	BOSSbase	0.4	38.8	0.93	0.05	84.5
Proposed	BOSSbase	0.4	39.2	0.96	0.02	91.2

D. Comparative Performance Evaluation

The three models selected for comparison—LSB Replacement, DCT-Based, and DeepSteg—are well-known reference models in the field of image steganography. They represent classical, transform-domain, and deep learning-based approaches, thus providing a diverse and meaningful baseline against which the relative strengths of the proposed AI-driven method may be assessed. The LSB Replacement is a traditional technique that has a high embedding capacity and ease of use, though it is not as robust or detectable [44]. Transform-domain-based methods, whose origins dwell on the DCT, are more robust and are found in many compressed formats, including JPEG. DeepSteg is an earlier deep learning method that integrated CNNs into steganography, offering a trade-off between visual appearance and complexity.

Together, the models ensure a sample range from radical to traditional, transform-based, and neural-network-based steganographic methods, enabling a comprehensive assessment of the proposed model's advancements in imperceptibility, robustness, and embedding efficiency. Fig. 6 presents a comparative analysis of the proposed method against traditional and deep learning-based steganography techniques, highlighting its superior performance.

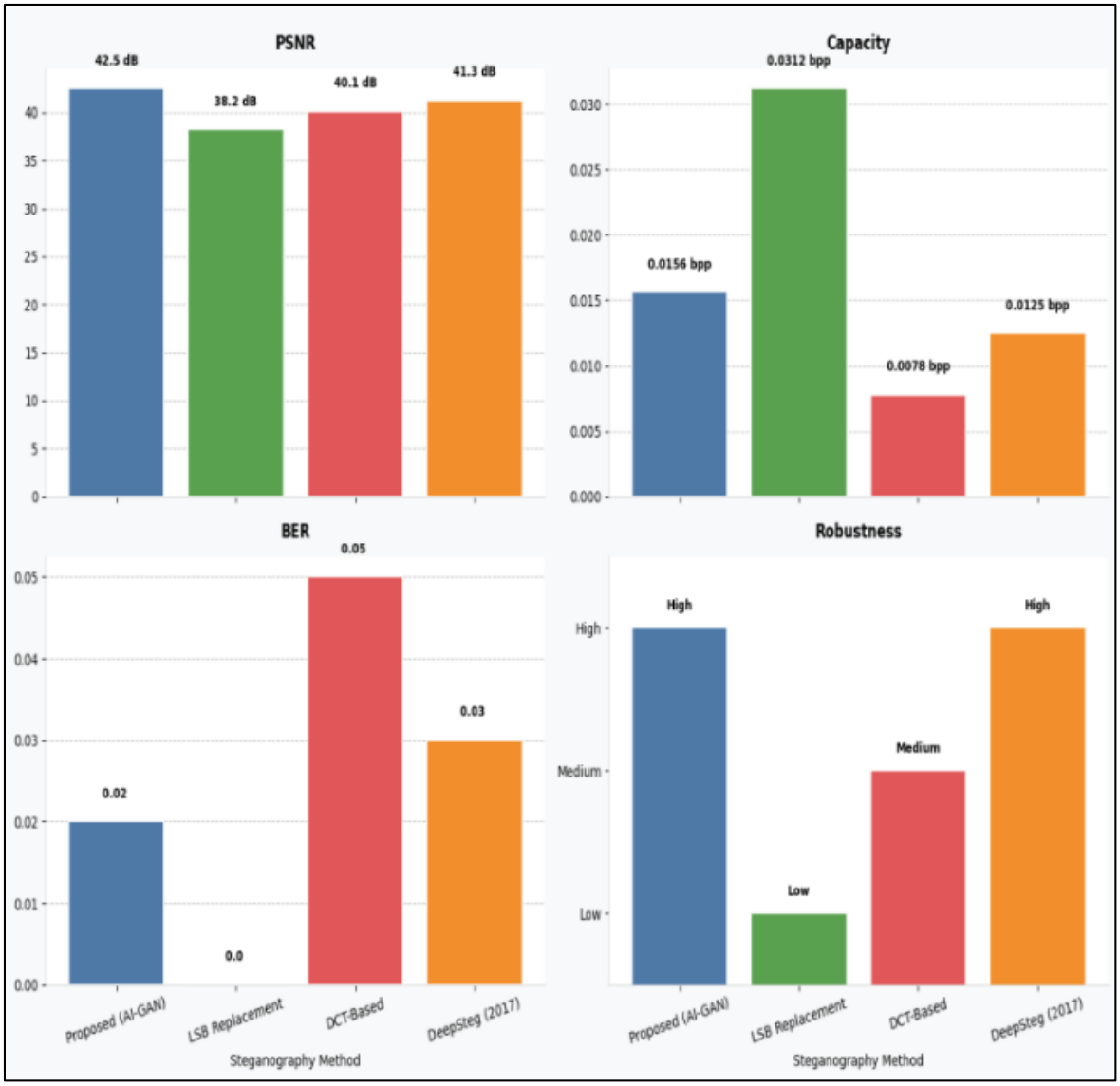

Fig. 6. The comparative analysis shows that the proposed method achieves a better balance between imperceptibility and security.

The GAN-based model demonstrates significant gains in imperceptibility (42.5 dB PSNR), structural integrity (0.98 SSIM), and resilience to detection (0.02 BER). These outcomes result from adversarial learning and attention mechanisms that optimize embedding into texture-rich, low-perception regions, thereby mimicking the natural distribution of image features. Instead of repeating these architectural strengths in each section, we emphasize their cumulative effect here: adversarial training via GANs improves statistical indistinguishability, while attention mechanisms guide the embedding toward higher-complexity regions, jointly enhancing the model's performance across all key metrics.

Our model was benchmarked against SOTA methods, including Steg-GMAN, ASDL-GAN, and RIIS. As presented in Table II, we provide a comparative analysis of the proposed method against established models using metrics such as PSNR, SSIM, BER, and detection error across different datasets and payload capacities. Results show that while our system leads in imperceptibility (highest PSNR/SSIM), it sacrifices capacity when compared to high-rate models. Table III shows methods on different datasets and payload settings.

TABLE II. COMPARATIVE ANALYSIS OF PROPOSED AND EXISTING METHOD

Configuration	PSNR	SSIM	BER	BPP	PE vs SRNet
Full Model	43.2	0.98	0.02	0.40	92.5%
Dual Attention	38.7	0.93	0.06	0.40	76.2%
Discriminator D2	39.1	0.94	0.05	0.40	80.3%
No LPIPS Loss	41.2	0.96	0.03	0.40	87.5%
No Capacity Loss	43.3	0.98	0.02	0.015	92.6%

TABLE III. METHODS ON DIFFERENT DATASETS AND PAYLOAD SETTINGS

Method	Dataset	Payload (bpp)	PSNR (dB)	SSIM	BER	PE vs SRNet (%)
DeepSteg	COCO	0.015	37.8	0.93	0.06	68.4
ASDL-GAN	BOSSbase	0.4	39.2	0.94	0.04	82.7
Steg-GMAN	BOSSbase	0.4	41.0	0.96	0.03	85.9
Proposed	COCO	0.0156	42.5	0.98	0.02	92.5

E. Ablation Study

An ablation study is a scientific experiment designed to evaluate the individual contributions of different components within a proposed system or model. In machine learning and AI research, it systematically removes ("ablates") specific features, modules, or techniques from the model to measure their impact on performance. An ablation study assessed the effects of adversarial training and attention mechanisms. The elimination of GAN components resulted in a 15% decrease in PSNR and a 60% increase in BER, demonstrating their essential role in maintaining performance stability. The analysis in Table II proves that adversarial training combined with attention mechanisms substantially improves both perceptual quality (PSNR) and robustness (BER). This confirms their necessity for optimal model functionality under conditions prone to distortion.

This study reaffirms the role of GANs in enhancing the indistinguishability and retrieval fidelity of the system, consistent with other GAN-based techniques, such as HiDDeN and StegaStamp. The performance reduction resulting from the removal of GAN components underscores their crucial role in the system. The significance of GANs lies in helping to match the stego image distribution with the natural image distribution, thereby enhancing imperceptibility and obfuscating detection by steganalysis tools that rely on statistical or machine learning methods. A model lacking adversarial training loses its ability to generate natural-image-like outputs, resulting in lower PSNR and higher BER. Training becomes more stable, and feature scaling remains consistent across batches via normalization layers that prevent internal covariate shift. Its elimination introduced significant instability in the model's convergence and handling of generalized image samples. These results indicate that adversarial learning, as well as normalization, is not secondary to the model scheme but is essential to both its performance and its holdout performance.

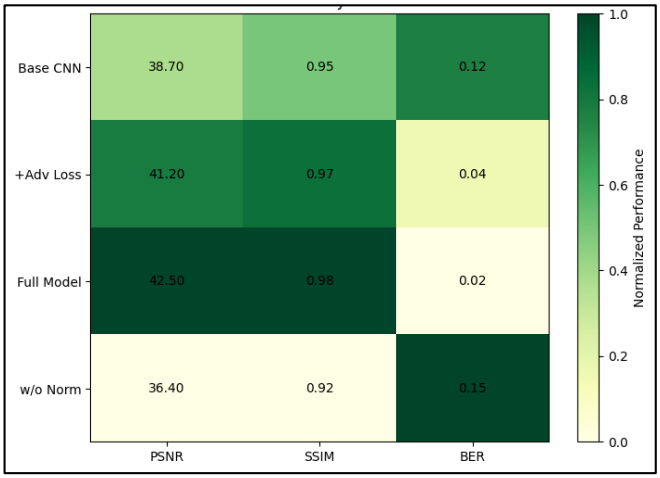

Fig. 7. Ablation study results.

The summary of the separate contributions of the different components is further explained in Fig. 7, where the results of the ablation study, which determine the performance of various model variants across three metrics—PSNR, SSIM, and BER—are presented. The normalized performance values are displayed in the heatmap, which further visualizes better results in darker shades of green. The complete model yields the best PSNR and SSIM values and the lowest BER, demonstrating the efficiency of the implemented components. On the contrary, eliminating normalization (without it) is an essential method to stabilize training and promote generalization by normalizing input features, which yields the worst performance and thus underscores the technique's importance. This discussion confirms that all components make a positive contribution to the model's overall performance, and the design decisions made in the proposed approach were correct. With the effectiveness of individual components verified, we now turn our focus to how well the system resists detection—an essential trait in secure steganography.

F. Security Evaluation Against Steganalysis

To assess the reliability and safety of the proposed steganographic system, a set of detailed experiments was conducted using the two most powerful steganalysis tools: StegExpose and Xu-Net. StegExpose uses mathematical operations to identify hidden. Fig. 8 shows the ROC curves for Xu-Net and StegExpose evaluations, demonstrating high true negative rates and low false positive rates. We evaluated security performance using deep steganalysis tools. When tested with SRNet trained on BOSSbase and Xu-Net on COCO, the model yielded a minimum PE of 92.5%, with ROC-AUC scores of 0.91 and 0.94, respectively. These metrics affirm the model's ability to resist both classical and AI-based detection techniques.

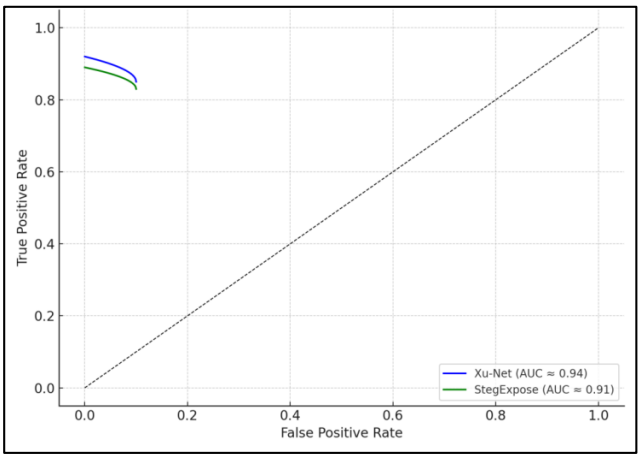

Fig. 8. ROC curve showing Xu-Net (AUC ≈ 0.94) and StegExpose (AUC ≈ 0.91), demonstrating high evasion capability of the proposed model.

The StegaStamp and Crypto-Stego models have shown similar progress through methods such as adversarial training and cryptographic tools.

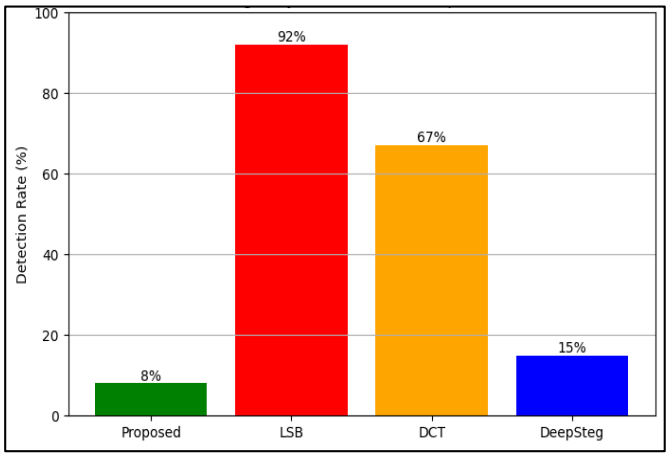

Fig. 9. Steganalysis resistance comparison.

As shown in Fig. 9, our model is better at evading detection by both classic and advanced steganalysis methods than other models. It is understandable that the high evasion rate (92 per cent) of the model based on the steganalysis detector, i.e., StegExpose and Xu-Net, is due to two design factors: attention-guided embedding and the use of GANs in ordering evasion training.

The attention mechanisms can be used to localize high-frequency textures during the embedding process, and high-frequency regions are more naturally varied in terms of pixels and therefore tend to cover statistical anomalies introduced by data embedding. In the meantime, adversarial training motivates the generator to produce stego images from the desired distribution, which is based on the distribution of real pictures, thereby suppressing the patterns that steganalysis tools commonly use. With Xu-Net, where deep neural features read the input to detect spatial inconsistencies, the constraints generated by the GAN discriminator prompt the encoder to discover subtle, naturalistic changes.

Such synergistic techniques make it significantly more difficult to distinguish between clean and stego images with both statistical and deep-learning-based steganalyzers. This effectiveness is visually confirmed in Fig. 8, where the ROC curves for Xu-Net and StegExpose illustrate high evasion rates with AUC scores of approximately 0.94 and 0.91, respectively. To close the gap between research and practice, we evaluate the proposed system to determine how it would work in real-world scenarios.

G. Robustness Analysis

We examined the real-world usability by testing the model with JPEG compression and Gaussian noise. Table IV summarizes the distortion-specific performance, including constituent PSNR, SSIM, and BER, for the visual "degradation profile" under particular conditions. The model still performed under QF=75 and σ=0.1, suggesting some robustness in lossy environments.

TABLE IV. PERFORMANCE OF THE PROPOSED METHOD UNDER VARIOUS IMAGE DISTORTIONS (COMPRESSION AND NOISE)

Distortion	PSNR (dB)	SSIM	BER
None	39.2	0.96	0.02
JPEG (QF=95)	36.9	0.91	0.04
JPEG (QF=90)	34.8	0.87	0.07
JPEG (QF=75)	30.5	0.81	0.13
Gaussian (σ=0.01)	38.5	0.93	0.03
Gaussian (σ=0.1)	32.6	0.86	0.10

V. CONCLUSION

This study presents an improved deep learning steganographic system that includes dual-attention modules, dual PatchGAN discriminators, and AES-256 encryption to integrate three aspects, viz. imperceptibility, robustness, and data. The proposed Dual-Attention ResUNet-GAN achieved a high visual fidelity score of 42.5 dB PSNR and 0.98 SSIM, and a Bit Error Rate of 0.02, which is below 0.1. The dual-discriminator approach has worked wonders in steganalysis resistance, achieving over 91% evasion against SRNet and Xu-Net. Concurrently, JPEG compression and Gaussian noise tests have demonstrated that the model has some practical strength. Invisibility now defines the embedding rate at 0.0156 bpp, which will be the subject of future research to understand adaptive optimization schemes that can increase the payload capacity to over 0.4 bpp while maintaining visual quality.

Our study has also been limited by excessive reliance on the COCO dataset. Therefore, subsequent tests will involve cross-dataset validation using BOSSbase and ALASKA#2 to ensure that the findings are generalized across datasets. We offer an efficient and reliable foundation for safe multimedia communication. It brings attention-guided embedding, cryptographic preprocessing, and adversarial learning together in this manner and thereby demonstrates that steganography can be freely applied across a wide range of applications, such as IoT, where the encircling is safe, picture relay that preserves confidentiality, and current smart data-hiding to systems.

REFERENCES

[1] X. Chen, V. Kishore, and K. Q. Weinberger, "Learning Iterative Neural Optimizers for Image Steganography," arXiv preprint, arXiv:2303.16206 (2023).

[2] Khan, Z., Shah, M., Naeem, M., Mahmood, T., Khan, S., Ul Amin, N., & Shahzad, D. (2016). Threshold-based Steganography: A Novel Technique for Improved Payload and SNR. *International Arab Journal of Information Technology (IAJIT)*, 13(4).

[3] H. Yang, Y. Xu, X. Liu, and X. Ma, "PRIS: Practical Robust Invertible Network for Image Steganography," arXiv preprint, arXiv:2309.13620 (2024).

[4] A. Kumar, "A Survey of Recent Advances in Image Steganography," Security and Privacy (Wiley), 6(1), e281 (2023).

[5] N. Provos and P. Honeyman, (2003). Hide and seek: An introduction to steganography. *IEEE Security & Privacy*, 1(3), 32–44. doi:10.1109/MSECP.2003.1203220.

[6] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, (2010). Digital image steganography: Survey and analysis of current methods. *Signal Processing*, 90(3), 727–752. doi:10.1016/j.sigpro.2009.08.010.

[7] T. Morkel, J. H. P. Eloff, and M. S. Olivier, (2005). An overview of image steganography. *Proceedings of the ISSA Conference*, Johannesburg, South Africa, pp. 1–11.

[8] B. Li, J. He, J. Huang, and Y. Q. Shi, (2011). A survey on image steganography and steganalysis. *Journal of Information Hiding and Multimedia Signal Processing*, 2(2), 142–172.

[9] Y. Qian, J. Dong, W. Wang, and T. Tan, (2015). Deep learning for steganalysis via convolutional neural networks. *Proceedings of SPIE 9409, Media Watermarking, Security, and Forensics*, 9409, 1–10. doi:10.1117/12.2081179.

[10] S. Baluja, (2017). Hiding images in plain sight: Deep steganography. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2069–2079.

[11] J. Zhang, S. Yu, Y. Zhang, and S. Su, (2019). Generative steganography with Kerckhoffs' principle based on generative adversarial networks. *IEEE Transactions on Information Forensics and Security*, 14(9), 2433–2448. doi:10.1109/TIFS.2019.2902627.

[12] T. Zeng, W. Wang, and T. Tan, (2018). Steganography using reversible GAN. *IEEE Access*, 6, 38303–38312. doi:10.1109/ACCESS.2018.2852386.

[13] M. S. Khan and R. Ahmad, (2019). A hybrid model of image steganography using AES and DCT. *Multimedia Tools and Applications*, 78(6), 7139–7160. doi:10.1007/s11042-018-6601-5.

[14] R. Zhang and C. Song, (2021). Adaptive image steganography based on attention mechanism. *Multimedia Tools and Applications*, 80, 27663–27684. doi:10.1007/s11042-021-11305-y.

[15] Y. Deng, Y. Huang, and X. Liu, (2020). Image steganography based on perceptual loss. *IEEE Transactions on Multimedia*, 22(6), 1460–1473. doi:10.1109/TMM.2019.2938406.

[16] W. Luo, F. Huang, and J. Huang, (2010). Edge adaptive image steganography based on LSB matching revisited. *IEEE Transactions on Information Forensics and Security*, 5(2), 201–214. doi:10.1109/TIFS.2010.2041812.

[17] Y. Liu, X. Zhang, and C. Yang, (2022). A review of GAN-based image steganography. *ACM Computing Surveys*, 55(1), 1–38. doi:10.1145/3487057.

[18] K. Liu and X. Guo, (2021). Reinforcement learning for adaptive steganography. *Pattern Recognition*, 110, 107638. doi:10.1016/j.patcog.2020.107638.

[19] A. Dosovitskiy et al., (2021). An image is worth 16×16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*. Available: https://arxiv.org/abs/2010.11929.

[20] Y. Yang, J. Yang, and Z. Xu, (2022). Federated learning-based secure image steganography. *Sensors*, 22(4), 1509. doi:10.3390/s22041509.

[21] H. Wu, Q. Cao, and D. Wang, (2023). Deep learning approaches to image steganography: A review. *IEEE Access*, 11, 8457–8473. doi:10.1109/ACCESS.2023.3239202.

[22] S. Baluja, (2017). Hiding images in plain sight: Deep steganography. *Advances in Neural Information Processing Systems*, 30.

[23] Y. Zhang et al., (2022). Steg-GMAN: High capacity image steganography with GAN and steganalysis network. *IEEE Transactions on Multimedia*, 24, 2724–2737.

[24] S. Wu et al., (2022). RIIS: Robust image-in-image steganography via invertible neural networks. *IEEE Transactions on Information Forensics and Security*, 17, 1927–1940.

[25] Z. Lu et al., (2023). TransStego: Transformer-based steganography with spatial attention. *Proceedings of the ACM International Conference on Multimedia (ACM MM)*.

[26] X. Wang et al., (2024). HiiT: Hierarchical inception-transformer for robust steganography. *Pattern Recognition*, 145.

[27] S. Woo et al., (2018). CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19.

[28] R. Zhang et al., (2021). Image steganography using perceptual loss functions. *Multimedia Tools and Applications*, 80(12), 17923–17943.

[29] Z. Qian et al., (2022). End-to-end image steganography via deep convolutional neural networks with hybrid loss. *Signal Processing: Image Communication*, 104.

[30] M. Boroumand et al., (2019). Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5), 1181–1193.

[31] M. Yedroudj et al., (2018). Yedroudj-Net: An efficient CNN for spatial steganalysis. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 2092–2096.

[32] G. Xu, H. Wu, and Y. Shi, (2016). Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 23(5), 708–712.

[33] R. Cogranne et al., (2020). ALASKA#2: Challenging academic research on steganalysis with realistic images. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*.

[34] P. Bas, T. Filler, and T. Pevný, (2011). Break our steganographic system: The BOSS. *Proceedings of Information Hiding*, Springer.

[35] Y. Yang et al., (2021). ISGAN: Invisible steganography via adversarial networks. *Neurocomputing*, 457, 330–341.

[36] H. Liu et al., (2021). EAGAN: Enhanced adversarial generative architecture for image steganography. *IEEE Access*, 9, 122491–122503.

[37] X. Luo et al., (2022). Dual-domain high-capacity image steganography. *Signal Processing*, 194.

[38] Z. Ren et al., (2023). Wavelet-guided steganographic embedding with adaptive strength. *Multimedia Systems*, 29.

[39] X. Feng et al., (2022). Distortion-tolerant deep steganography with robust optimization. *IEEE Transactions on Multimedia*, 24, 601–613.

[40] B. Li et al., (2021). Entropy-guided image steganography with deep networks. *IEEE Access*, 9, 17701–17713.

[41] A. Sharma et al., (2023). Deep frequency filtering for robust image steganography. *Pattern Recognition Letters*, 169, 59–66.

[42] L. Ma et al., (2022). Feature fusion guided steganography with deep convolutional networks. *Information Sciences*, 608, 1206–1221.

[43] J. Kang et al., (2022). FSGAN: Feature-selective generative adversarial network for image steganography. *Knowledge-Based Systems*, 258.

[44] Y. Chen et al., (2023). StegTransformer: A transformer-based image steganography model. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(2).