Journal of Kerbala University, Vol. 23, Issue 2, June , 2026

Available at https://www.iasj.net/iasj

Iraqi Academic Scientific Journals

Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

A Systematic Review of Deep Learning for Image Steganography

[1,]Maryam Abdulameer Oudah [2,] Baheeja Khudhair Shukur
[3,] Mohammed Abdallazez Mohammed
University of Kerbala, Iraq

Article Info
Article history:
Received 2 -5-2026
Received in revised form 1-6-2026
Accepted 22-6-2026
Available online 30 -6 -2026
Keywords: Image steganography; Deep learning; Convolutional neural networks; Generative adversarial networks; Vision Transformers; Diffusion models; Autoencoder-based steganography; Information hiding; Steganalysis resistance; Explainable artificial intelligence.

Abstract

Deep learning has brought about a change in the ways images are hidden, allowing the transition from manual embedding techniques to trainable, data-driven models. This review presents a comprehensive and structured study of deep learning-based image steganography techniques , categorizing current approaches into five common architectural categories: convolutional neural networks (CNNs), Generative Adversarial Networks (GANs) , transformer-based models, diffusion-based generative architectures, and autoencoder-based architectures. This review analyzes and compares the architectures, embedding methods, performance characteristics, and evolutionary trends of existing models , with a focus on undetectable capabilities, payload capacity, flexibility, security against steganalysis attacks , and ease of deployment. The review begins with early CNN-based encoder–decoder models and extends to adversarial, attention-based, semantic, probabilistic, and reconstruction-driven steganographic frameworks. It further examines how representation learning, self-attention mechanisms, adversarial optimization, and stochastic generative modeling have influenced the development of modern image steganography systems. In addition to synthesizing recent advances, this study identifies persistent research gaps, including limited cross-domain generalization, insufficient explainability, computational complexity, inconsistent robustness evaluation, and the absence of standardized benchmarks. By unifying developments across different architectural families and generations, this review provides a focused taxonomy and a critical understanding of the current state of deep learning-based image steganography. It also outlines future research directions toward more adaptive, explainable, robust, secure, and practically deployable steganographic systems.

Corresponding Author E-mail: maryam.abdulameer@s.uokerbala.edu.iq , baheeja.k@uokerbala.edu.iq mohammed.abdallazez@uokerbala.edu.iq
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

The increasing appeal of information security stems from the communication of multimedia data across frequently unsecured networks and the rising rate of data transmission. The data are protected by the type of encryption that is considered conventional, yet the fact that the data are present is, nevertheless, disclosed. The implementation of a supplementary kind of security has been implemented to resolve this cryptographic issue. Steganography is a type of security that conceals information by enclosing it within digital data that appears to be innocuous. Steganography takes this approach to security. This can be illustrated using text, audio, or visual media. [1]. On account of the fact that photos are the most often used digital asset throughout the internet and social media platforms, the most frequent type of steganography is steganography that is embedded into photographs. The purpose of this endeavor is to incorporate information into a cover photo in such a way that it is not visible to the naked eye that the image contains elements of information. In the past, the primary approaches for picture steganography were in the spatial domain. Earlier methods commonly used discrete cosine transform (DCT) or discrete wavelet transform (DWT) [2]. While older methods were characterized by their computational simplicity and high processing speed, they suffered from limitations. These limitations included limited data capacity, poor resistance to compression and distortion, and susceptibility to detection by newer detection techniques. Machine learning, particularly deep learning, has led to the development of techniques capable of automatically learning complex spatial relationships and embedding methods from image datasets. The major shift from manually constructed algorithms to self-learning and adaptive systems has been observed. Deep learning-based steganography systems have demonstrated significant improvements in steganography quality, attack resistance, and data storage capacity.

These systems have overcome the limitations of previous methodologies, such as limited steganography capacity and vulnerability to detection, by employing advanced techniques that enhance their performance and reliability. The integration of Generative Adversarial Networks (GANs) , pioneered by Goodfellow et al. (2014), into steganography applications [3] represents a significant step forward in this progress. GAN-based systems utilize a generator to embed hidden messages within the original images, while a discriminator distinguishes between real and generated samples [4], [5] . This results in more realistic and effective steganography images.

It has been shown that generative models can directly encode data while creating images through steganography without embedding. Recent developments in deep learning-based steganography systems depend on the creation of suggested transformer-based topologies, the integration of attentional processes, and the incorporation of spatial frequency feature learning. The maximum signal-to-noise ratio (PSNR), bit error rate (BER), and structural similarity index (SSIM) all significantly improved as a result, showing that these techniques outperformed earlier steganography techniques. Additionally, by using Explainable Artificial Intelligence (XAI) techniques, the researchers improved the transparency and dependability of deep learning-based information steganography systems and offered new insights into the embedding mechanism. This helped users better understand the decision-making processes in these systems and increased confidence in their efficacy.

This systematic review focuses on image Steganography using deep learning. It can be summarized as follows:

1. **Comprehensive Mapping of Existing Research:** A systematic review aggregates and organizes all exemplary studies on deep learning-based image steganography. This means putting tasks into groups like spatial-domain, frequency-domain, hybrid, adversarial, and resilient steganography; finding the most common DL architectures (CNNs, GANs, Autoencoders, Transformers, etc.); and giving a brief overview of the most common embedding and extraction strategies.
2. **Finding open issues and research gaps:** The systematic review can identify neural architectures that have not received enough attention, a lack of studies that demonstrate how they function in the real world, a need for improved resistance to contemporary steganalysis tools, a lack of large and standardized datasets, inadequate benchmarking, and reproducibility issues by comparing methods and results across studies. This aids in directing subsequent studies.
3. **Developing Conceptual or Taxonomy Models:** Structured taxonomies, such as DL architecture classification, embedding domain classification, robustness against attack taxonomy, and optimization target classification, may result from the evaluation. The field is kept organized by these frameworks.
4. **Trends and Future Directions:** More transformer-based architectures, diffusion models for stego synthesis, improved adversarial robustness, privacy-preserving and federated steganography, and multi-modal steganography (image-in-audio, text-in-image, and video-in-image) could all arise from a thorough review.
5. **Standardizing evaluation protocols and criteria:** Due to the different criteria used in different research, it may be difficult to evaluate the results and draw reliable conclusions.

2. The Importance of Deep Learning with Image Steganography

In the field of information security, image steganography is essential because it protects sensitive information from public disclosure by hiding it inside digital photographs. Early advances in this area were on transform-domain algorithms using techniques like Least Significant Bit (LSB) replacement, such as the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT). However, there are a number of issues with these older methods that must be resolved, such as their reduced capacity, susceptibility to noise and compression, and increased susceptibility to steganalysis detection.

Image steganography has changed dramatically since deep learning emerged, moving from data-driven models to ones with human-like intelligence. Deep neural architectures such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Transformer networks [6], [7], [8] have made notable advancements in enhancing imperceptibility, robustness, and adaptability. Compared to conventional methods, deep learning presents numerous advantages and is increasingly supplanting standard practices in image steganography. These advanced models possess the ability to autonomously learn features.

CNNs utilize hierarchical spatial representations—such as edges, textures, and semantic structures—to identify optimal spatial locations for embedding data, resulting in less noticeable alterations than those produced by older techniques relying on LSB or DCT [9]. Encoder-decoder networks can be trained within a unified deep framework through parallel training. This approach allows for comprehensive optimization

throughout the entire process while improving the balance between distortion and recovery accuracy.

An example of this innovation is HiNet, which incorporates Invertible Neural Networks (INNs) for image concealment, enabling high-quality image retransformation with minimal loss. Subsequent research has facilitated multi-image embedding capabilities and frequency-domain learning through various extensions [6, 10, 32]. Furthermore, deep learning serves as a more secure and robust mechanism for maintaining data integrity and protecting hidden information against transformations, compression, or noise introduction. The adversarial training methodologies inherent in GAN-based

models render them particularly challenging to assess [6], [11]. Additionally, adaptive hybrid architectures like U-Net combined with DWT optimize the payload-imperceptibility trade-off by achieving enhanced capacity alongside improved image quality [7, 28, 8]. Lastly, recent research has used Explainable AI (XAI) tools like Grad-CAM and LRP to show embedding behavior in order to make models easier to understand and more open [9]. Overall, deep learning has enabled image steganography systems to learn adaptive embedding and recovery strategies, improving imperceptibility, robustness, and capacity compared with traditional handcrafted methods. Recent work on diffusion models, Transformers, and autoencoders further indicates a shift toward more secure, interpretable, and scalable steganographic frameworks. [28, 8].

Table 1: Systematic Review Methodology and Selection Framework

Steps	Description
Data Sources / Digital Libraries	ScienceDirect (Elsevier), IEEE Xplore, SpringerLink, MDPI
Additional Academic Resources	Google Scholar, arXiv, Publisher Platforms (e.g., NeurIPS, CVPR Proceedings)
Search Type	Systematic Literature Search
Search Method	Boolean Query-Based Retrieval
Coverage Period	Recent Publications (2021– March /2026)
Inclusion Criteria	• Publications written in English • Studies directly related to deep learning-–based image steganography • Journal articles and conference papers • Relevant deep learning architectures (CNN, GAN, Transformer, Autoencoder, Diffusion Models)
Exclusion Criteria	• Non-English publications • Studies unrelated to deep learning or image steganography • Papers with inaccessible full text • Outdated publications (before 2020) • Non-academic or low-credibility sources
Study Selection Stages	Identification → Duplicate Removal → Title/Abstract Screening → Full-Text Eligibility Assessment → Final Inclusion
Distribution of Selected Publications	ScienceDirect (40%), IEEE Xplore (32%), SpringerLink (15%), MDPI (5%), Other Sources (8%)

To improve the systematic review methodology, the process of study selection was clarified. The literature search was performed by using Boolean search strings combining terms related to deep learning

and image steganography such as "deep learning" AND "image steganography," "CNN" AND "image hiding," "GAN" AND "steganography," "Transformer" AND "image steganography," "diffusion model" AND "steganography," and "auto-encoder" AND "image hiding." The retrieved studies were initially screened by titles and abstracts, and then assessed for full-text eligibility against the inclusion/exclusion criteria as described in Table 1 The included studies directly investigated image steganography based on deep learning, reported relevant architectures or evaluation metrics, and provided sufficient methodological or experimental details. Studies were excluded if they were not related to image steganography, full text was not accessible, there was no sufficient technical information or they could not be verified by reliable bibliographic information. In order to reduce the risk of selection bias, the screening process was checked with the pre-defined criteria and conflicts were solved by discussion between the authors. A basic quality assessment was also carried out looking at publication status, source credibility, methodological clarity, dataset description, reported evaluation metrics, robustness/security evaluation and relevance to the five architectural families discussed in this study.

A PRISMA-style workflow was also added to visually summarize the identification, screening, eligibility assessment, and final inclusion stages of the systematic review process.

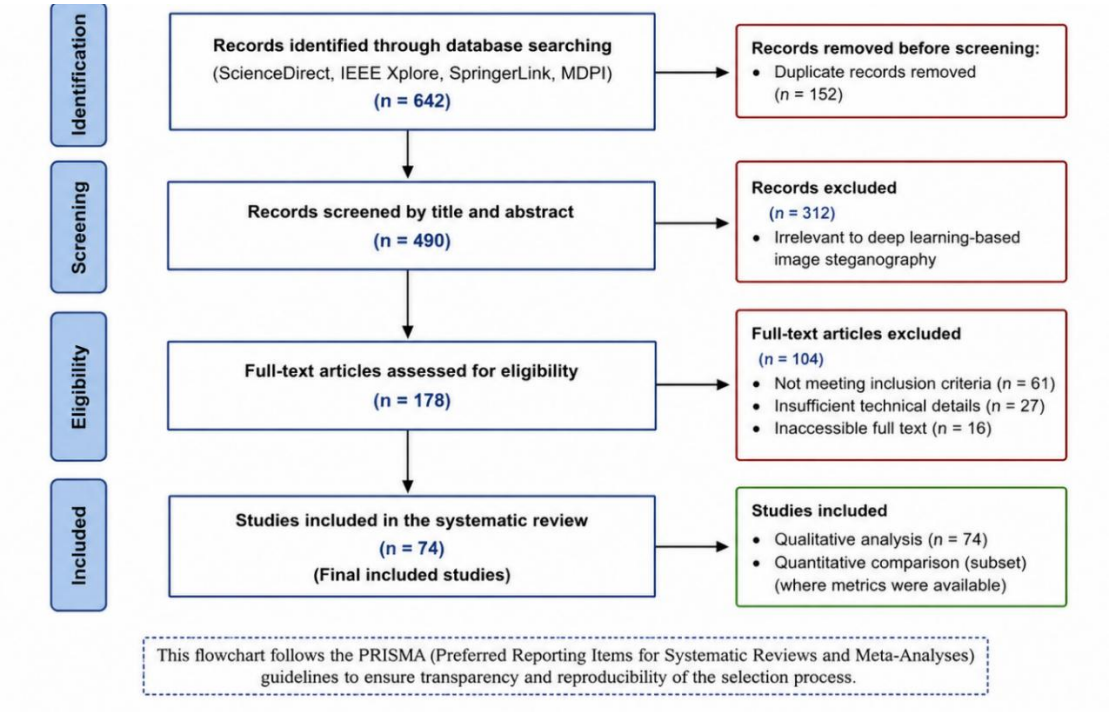

Fig 1: PRISMA flowchart of the study selection process.

This figure illustrates the systematic review workflow, including the number of records identified, screened, excluded, assessed for eligibility, and finally included in the review.

3. Evaluation Metrics for Image Steganography

The study found that deep learning for image steganographyhas been applied to four different multimedia elements; image, video, audio and text. The organization of the systematic review is shown in Figure 2. There is a lack of standardized evaluation metrics to compare the image steganography methods between different studies. In deep learning-based image steganography, evaluation usually concentrates on four main aspects: imperceptibility of the cover image, quality of the secret image recovery, extraction accuracy, and embedding capacity [5, 4]. The four major metrics used in this review are defined as follows:

1.Peak Signal-to-Noise Ratio (PSNR)

The pixel-level similarity between two images, such as the original cover image and the generated stego image or the original secret image and the recovered secret image, is measured by PSNR. It measures the ratio of the distortion induced during embedding or recovery to the highest pixel intensity that can be achieved [19, 7]. Higher PSNR levels often indicate less distortion and improved visual imperceptibility. PSNR is measured in decibels (dB). It is frequently employed as the main stego image quality metric in CNN-based [19, 22, 34], GAN-based [41, 48, 66], Transformer-based [77, 80], diffusion-based [83], and autoencoder-based systems [6, 103]. The definition of PSNR is:

$$PSNR = 10 \cdot \log_{10} (MAX^2_I / MSE)$$

where MAX_I is the maximum possible pixel value of the image (255 for 8-bit images), and MSE is the mean squared error between the reference image and the compared image, defined as:

$$MSE = (1 / MN) \cdot \Sigma_i \Sigma_j [I(i,j) - K(i,j)]^2$$

where I and K represent two compared images of size M × N. Despite its widespread use, PSNR has been criticized for measuring only pixel-level amplitude distortion without capturing statistical detectability by steganalysis tools [5, 19], and thus should not be interpreted as a measure of security.

2.Structural Similarity Index Measure (SSIM)

By concurrently evaluating brightness, contrast, and structural data, SSIM assesses the perceived similarity between two images. SSIM captures local structural degradation more correctly and is more compatible with the human visual system than PSNR, which only assesses pixel-level error [7, 8]. Higher structural similarity and superior perceptual quality are indicated by SSIM values, which range from 0 to 1. In CNN-based [7, 20, 34], GAN-based [48, 51, 66], Transformer-based [77, 80, 87], diffusion-based [91, 92], and autoencoder-based systems [6, 103], it is reported alongside PSNR in the reviewed literature as a complementary imperceptibility metric. The definition of SSIM is:

$$SSIM(x, y) = [(2\mu_x\mu_\gamma + C_1)(2\sigma_{x\gamma} + C_2)] / [(\mu_x^2 + \mu_\gamma^2 + C_1)(\sigma_x^2 + \sigma_\gamma^2 + C_2)]$$

where μ_x and μ_γ are the mean intensities of images x and y, σ_x^2 and σ_γ^2 are their variances, $\sigma_{x\gamma}$ is their covariance, and C_1 and C_2 are stabilizing constants.

3.Bit Error Rate (BER)

The percentage of improperly recovered secret bits in relation to the total number of embedded bits is measured by BER. It is mostly used to assess the precision of the extraction or recovery procedure in steganographic systems, and it is especially important for systems assessed under distortion conditions like Gaussian noise and JPEG compression [77, 34, 88]. Higher BER values indicate more extraction error,

whilst a BER of 0 denotes full recovery. BER is reported in robustness-focused studies where message recovery accuracy under distortion is a primary goal, such as the CNN adaptive loss system [34], TRPSteg [77], and the reversible Transformer latent embedding system [88]. The definition of BER is:

$$BER = N\_error / N\_total$$

where N_error is the number of incorrectly recovered bits and N_total is the total number of embedded bits.

4.Payload Capacity

Payload capacity, which is typically defined as the ratio of the secret picture size to the cover image size or measured in bits per pixel (bpp), is the amount of secret information that may be placed within a cover image [19, 6, 80]. More secret information can be concealed with a larger payload capacity, but as the reviewed literature has shown, this increases the risk of discovery and decreases imperceptibility by introducing more visual or statistically identifiable artifacts [5, 7, 28]. The definition of payload capacity is:

$$Payload\ Capacity = N\_embedded\ bits / N\_cover\ pixels$$

where N_embedded bits is the total number of embedded secret bits and N_cover pixels is the total number of pixels in the cover image. Payload capacity ranges widely across the reviewed systems, from approximately 0.4–1.0 bpp in security-focused CNN and GAN methods [19, 20, 51] to full-image hiding at approximately 24 bpp in invertible autoencoder systems such as HiNet [6] and SMILENet . This range reflects the fundamental trade-off at the core of image steganography: payload capacity must always be balanced against imperceptibility, robustness under distortions, and resistance to steganalysis attacks [5, 7, 8].

4. Deep Learning-Based Approaches for Image Steganography

This section contains the included research papers that passed all the exclusion criteria and it is divided into five parts according to the deep learining techniques used for image steganography .

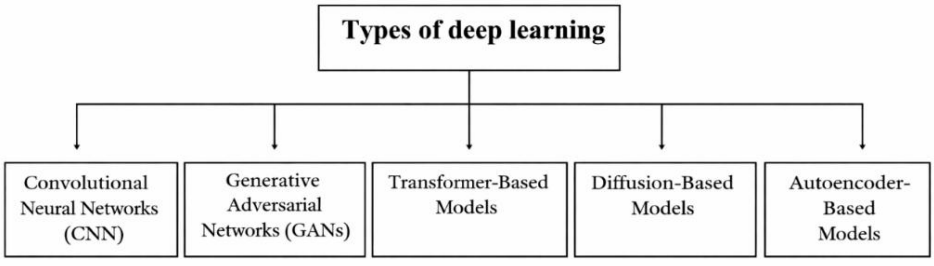

Fig 2: Diagram of the structure of the systematic review paper.

4.1 CNN based Image Steganography

LeCun et al. [12] proposed the LeNet-5 architecture that was a major advance of CNNs. Such information was distributed among fully connected layers, pooling layers, and convolutional layers. It is relatively cost-effective in parameter configuration; however, stability was attained with a few deviations when using a parameter that encodes handwritten digits. It also brought in important notions like shared weights, subsampling, and local receptive fields. AlexNet [13] further raised the interest in CNN research while showing that deep networks, powered by GPU systems, could better do things like ImageNet [14], where it was shown that the networks performed faster in scaling data such as ImageNet when using ReLU activations, dropout techniques, and new data augmentation techniques. In each study following the previous class of CNNs, the following design was a derivative of each other: VGGNet used the constant 3×3 kernels for focusing on depth; ResNet [15] integrated the residual connections for ultra-deep training; Inception [16] employed the multi-scale modular processing approach; and EfficientNet [17] suggested compound scaling for better efficiency. In recent years, there was a variety of attention mechanisms, transfer learning schemes, hybrid CNNs with transformers hybridized designs and different compression methods have also been explored for efficient deployment in CNN [18]. These progresses, all of which are represented by the recent developments in CNN architectures, from LeNet-5, to efficiently scalable CNN architectures. Convolutional neural network (CNN) architectures are now essential as part of deep learning-based image steganography, providing an alternative to end-to-end learning and relying on predefined embedding rules. Early encoder–decoder architectures outperformed standard least significant bit (LSB) methods on terms of imperceptibility and payload capacity [19]. Loss functions and architectural adjustments have enhanced CNNs' capability to hide data while it is concluded from the reviewed literature projects that CNNs were considered as the front-runner backbone for various applications [5, 9]. Attention-enhanced CNNs increased the embedding accuracy , and hybrid approaches based on a spatial frequency of discrete wavelet transform (DWT) or discrete cosine transform (DCT) decreased the sensitivity to compression or distortion effects [20]. GANs for adversarial training significantly improved CNN frames' defenses to steganalysis attacks [21]. Nevertheless, adaptive optimization, robustness assessment, and explainability still pose challenges regarding their performance. Recent works have been turning increasingly towards scaling U-Net variants; multi-scale CNNs and adaptation to the specific application [22]. Although there has been a great deal of recent work toward CNN-based image steganography, there are some important limitations that still exist. While a lot of techniques utilize visual fidelity metrics like PSNR and SSIM extensively, they do not yet adequately consider security or robustness to real-world distortions—such as compression artifacts or noise—or current learning-based steganalysis ; [9]. In addition, the majority of the loss function models are static and have fixed-weight multi-objective models. This stiffness restricts the approach to trade-off imperceptibility with payload capacity and robustness in different transmission environments or image properties [19], [21]. Very few studies have been conducted on cross-domain generalization as a majority of CNN based generalization models have been investigated with small datasets and with only a few image formats and distribution. Their effectiveness against heterogeneous domains, such as JPEG or HDR imagery has thus far been subjected to unstructuredly untried rigorously on them [5],

[22]. And while such hybrid spatial-frequency methods that integrate the DWT or DCT with hybrid SpF techniques are also proposed for hybrid methods, they often fail to integrate those approaches together in coherent manner and do not establish an overall optimization framework that maximizes the optimal harmony between the spatial and frequency settings [20], . Cited in, there are no interpretable AI (XAI) approaches present in CNN-based steganographic models, reducing transparency of learned embedding behaviors in CNN-based steganographic models, thereby emphasizing the dearth for XAI methods, illustrating there remains a significant explainability gap. The bulk of the research focuses exclusively on specific architectures in CNN with no systematic approach of comparing such architectures with other encoder-decoder architectures or feature fusion algorithms, severely constraining both architectural exploration and comparative avenues [19], [20]. The difficulties associated with scale and deployment have made practical applications even more difficult. In the meantime, these are often neglected to favor computational efficiency and memory usage; but solving real-world applicability is important as well [22]. Finally, the lack of standard practices prevents unbiased comparisons across studies and thus attempts at reproducibility; so methods are inconsistent across this area of research [5] . In conclusion, these shortcomings highlighted by the present report show an importance for the development of CNN-based steganography systems: not only do they need to adopt adaptive optimization strategies in conjunction with thorough robustness testing, but also cross-domain generalization capabilities and architectural flexibility together with robust measure of explainability must be stressed upon. Although the advantages brought to the process of spatial-frequency fusion using wavelet transformations are evident in reference work cited as [35], nevertheless this scheme still presents challenges to further refine as it neglects parts of the adaptive optimization and also provides no comprehensive robustness evaluations together with interpretability measures.

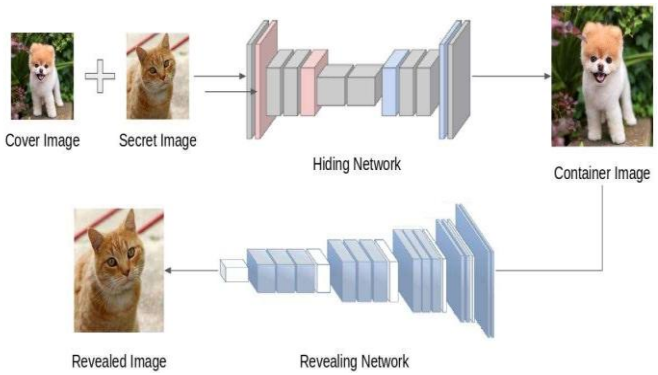

Fig 3: Conceptual architecture of CNN-based image steganography.

Table 2: Comparative Summary of CNN-Oriented Steganography Studies

Ref.	Year	Method / Framework	Dataset	Key Focus	Robustness Evaluated	Steganalysis / Security Tested	Adaptive Loss	Spatial–Frequency Fusion	Using XAI	Scalability / Efficiency
[22]	2020	SteganoCNN (CNN Encoder–Decoder)	ImageNet	Generalization and multi-image hiding	Partially addressed	Not clearly evaluated	Not addressed	Not addressed	Not reported	Heavy model
[23]	2021	Siamese CNN Steganalysis	BOSSBase / ALASKA#2	Feature similarity-based steganalysis detection	Explicitly addressed	Explicitly addressed	Not applicable	Not addressed	Not reported	Normal
[25]	2022	Loss Function Robustness Study	BOSSBase / DIV2K	Stability analysis under different loss functions	Partially addressed	Not the core focus	Partially addressed	Not addressed	Not reported	Normal
[26]	2022	NAFF + Attention CNN	VOC / RESISC45-type datasets	Imperceptibility improvement with low complexity	Limited	Not evaluated	Not addressed	Not addressed	Not reported	Explicitly addressed
[27]	2023	Extractor Matching CNN	BOSSBase / COCO	Decoder-aware embedding and extraction matching	Limited	Partially addressed	Not addressed	Not addressed	Not reported	Normal
[28]	2024	Hybrid CNN High-Capacity Framework	Natural RGB dataset	Payload maximization	Not clearly evaluated	Not evaluated	Not addressed	Not addressed	Not reported	Not clear
[20]	2021	Frequency-Domain CNN	BOSSBase-JPEG	Compression-robust embedding	Explicitly addressed	Limited	Not addressed	Explicitly addressed	Not reported	Normal
[29]	2023	Lightweight StegoNet	Natural image dataset	Edge/mobile deployment	Limited	Not evaluated	Not addressed	Not addressed	Not reported	Explicitly addressed
[30]	2023	Dual Attention U-Net	COCO / DIV2K-type datasets	Feature preservation and attention-guided hiding	Limited	Not evaluated	Not addressed	Not addressed	Not reported	Normal
[31]	2021	Hybrid Autoencoder + DWT	ImageNet / BOSSBase	Spatial–frequency synergy	Partially addressed	Limited	Not addressed	Explicitly addressed	Not reported	Normal
[32]	2026	End-to-End CNN + Encryption	STL-10	Joint encryption and image hiding	Limited	Not evaluated	Not addressed	Not addressed	Not reported	Not focused
[33]	2026	WSERNet Steganalysis	BOSSBase / BOWS2	Weak steganographic signal extraction	Explicitly addressed	Explicitly addressed	Not applicable	Not addressed	Not reported	Efficient
[34]	2025	CNN + Adaptive Loss (LF3)	Tiny-ImageNet and related datasets	Payload–robustness balance	Explicitly addressed	Explicitly addressed	Explicitly addressed	Not addressed	Not reported	Normal
[35]	2021	Data Hiding Scheme Based on U-Net and Wavelet Transform	VOC 2012 / ImageNet	Image-in-image hiding using grayscale secret and color cover images	Partially addressed	Limited	Not addressed	Explicitly addressed	Not reported	Normal

4.2 GAN-Based Image Steganography

Goodfellow et al. [3] developed Generative Adversarial Networks (GANs), which represent a fundamental framework in contemporary image steganography and a notable advancement in generative modeling. The GAN framework originally facilitated adversarial learning to achieve data-driven approximations of complex image distributions by establishing a minimax game between a discriminator and a generator [3]. This theory was the basis for steganographic systems that use GANs. The main goals of the first changes to the architecture were to make training more stable and synthesis more accurate. With the addition of batch normalization and the switch from fully linked layers to convolutional structures in DCGAN , GANs became good for modeling pictures at the pixel level. Theoretical advancements, including Wasserstein GAN (WGAN) [36] and WGAN-GP [37], enhanced adversarial training by addressing gradient instability and mode collapse through novel loss formulations and gradient penalties. Conditional and translation-based variations, like Conditional GANs , InfoGAN [38], Pix2Pix [39], and CycleGAN [40], made it possible to extend adversarial learning even further to controlled and paired picture production. The most recent versions had cover-secret mapping and translation-based concealing schemes.Between 2018 and 2021, spatial-domain GAN-based steganography models like HCISNet [41] and SteganoGAN made the payload capacity and invisibility much better. These methods were still very much limited to pixel-space representations, which made them easy to change and compress. After 2021, steganography using GANs moved toward hybrid and multi-stage architectures. Architectures that used residual feedback, rate-distortion optimization, attention mechanisms, and hierarchical generation [40]–[41] made reconstruction more accurate, more stable, and more realistic in terms of meaning. Attention-driven and hybrid CNN-Transformer designs [42]–[44] made it possible to use adaptive and context-aware embedding methods, but there were still problems with training stability and computing complexity. Recent advances have been mostly about designs that are focused on efficiency, representations that take geometry into account, and managing latent space. To make things more stable, models like SPDGAN [55] use geometry-aware latent representations. To make things more coherent, models like StegaStyleGAN [53] use StyleGAN-based latent manipulation. Lightweight adversarial generative network (GAN) frameworks [56] allow for model switching and aid in deployment issues as well as scalability.

One cannot directly place the GAN Steganography in that framework, as it can be grouped into non-embedding (SWE) but they does not embedded implicitly available cover image directly with concealed information instead learns to encode between they're and hidden while producing them [4].

By matching hidden images with the distributions extracted from the images, this method enhances the statistical normality and detection resistance of hidden images. Generative Adversarial Networks (GANs) and their generators use convolutional neural networks (CNNs) to extract spatial features and model textures [1, 2].

New techniques that leverage explainable artificial intelligence (XAI) methods, for instance Grad-CAM [42] and SHAP [43], to better understand embedding functions/model purposes have also been encompassed into higher trust transparency adherence. While substantial progress has been made, there are several ongoing challenges.

Table 3: Comparative Summary of GAN-Based Steganography Studies

Ref	Year	Method / Framework	Dataset	Key Focus	Robustness Evaluated	Steganalysis / Security Tested	Adaptive Loss	Spatial–Frequency Fusion	Using XAI	Scalability / Efficiency
[6]	2021	HiNet	ImageNet / COCO	Reversible full-image hiding with high-fidelity recovery	Explicitly Addressed	Limited	Not Addressed	Explicitly Addressed (Wavelet/DWT)	Not Reported	Heavy Model
[47]	2021	GFR-Net / Residual Feedback GAN	ImageNet-type	Multi-stage refinement and robustness enhancement	Explicitly Addressed	Limited	Explicitly Addressed	Not Addressed	Not Reported	Computationally Heavy
[49]	2022	Cross-Feedback GAN	ImageNet-type	Cross-feedback guided hiding and recovery stabilization	Explicitly Addressed	Partially Addressed	Explicitly Addressed	Not Addressed	Not Reported	Training Intensive
[50]	2023	IDGAN	COCO-type	Attention-guided adaptive adversarial embedding	Partially Addressed	Explicitly Addressed	Explicitly Addressed	Not Addressed	Not Reported	Normal
[51]	2023	Adaptive GAN Steganography	ImageNet-type	Dynamic loss weighting and anti-detection optimization	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Not Addressed	Not Reported	Normal
[60]	2023	Smooth Cycle-Consistent Adversarial Steganography	ImageNet-type	Smoothness-constrained adversarial embedding	Partially Addressed	Explicitly Addressed	Explicitly Addressed	Not Reported	Not Reported	Computationally Heavy
[61]	2023	Evolving GANs for Steganography	Natural Images	Optimization-driven detection resistance enhancement	Limited	Explicitly Addressed	Explicitly Addressed	Not Reported	Not Reported	Normal
[62]	2023	High-Capacity Coverless GAN Framework	Synthetic / generated images	Capacity maximization via adversarial generation	Partially Addressed	Explicitly Addressed	Explicitly Addressed	Not Applicable	Not Reported	Computationally Heavy
[63]	2023	NOStyle (Noise-Optimized StyleGAN2)	Generated images	Secure cover generation & distribution preservation	Limited	Explicitly Addressed	Explicitly Addressed	Not Applicable	Not Reported	Heavy Model

[64]	2023	StegoPix2Pix (cGAN Translation Steganography)	ImageNet-type	Image-to-image adversarial hiding / revealing	Partially Addressed	Limited	Explicitly Addressed	Not Reported	Not Reported	Normal
[52]	2024	GAN–Transformer Fusion	ImageNet / COCO-type	Global context-aware and semantic-preserving hiding	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Conceptually Addressed	Partially Reported	Computationally Heavy
[53]	2024	StegaStyleGAN	FFHQ / ImageNet-type	Latent-space texture-preserving stego-image generation	Partially Addressed	Limited	Explicitly Addressed	Not Addressed	Not Reported	Normal
[54]	2024	SPDGAN	ImageNet-type	Geometry-aware latent representation for robust hiding	Explicitly Addressed	Limited	Explicitly Addressed	Not Addressed	Not Reported	Resource Intensive
[65]	2024	GAN-Based Adaptive Cost Learning	BOSSBase / ImageNet-type	Security-driven embedding cost optimization	Partially Addressed	Explicitly Addressed	Explicitly Addressed	Not Addressed	Not Reported	Normal
[66]	2024	High Invisibility Wavelet-GAN	ImageNet-type	Imperceptibility & robustness via frequency-aware embedding	Explicitly Addressed	Limited	Partially Addressed	Explicitly Addressed (Wavelet fusion)	Not Reported	Normal
[67]	2024	Coverless GAN-based Steganography	Generated / synthetic datasets	Coverless hiding via adversarial image synthesis	Partially Addressed	Explicitly Addressed	Explicitly Addressed	Not Applicable	Not Reported	Computationally Heavy
[68]	2024	Robust Joint Coverless GAN Scheme	Natural Images	Joint optimization of generation & recovery robustness	Explicitly Addressed	Partially Addressed	Explicitly Addressed	Not Applicable	Not Reported	Training Intensive
[69]	2024	Generative Pose-Keypoint Steganography	ImageNet-type	Generative semantic-guided embedding	Limited	Limited	Partially Addressed	Not Reported	Not Reported	Normal
[55]	2025	Lightweight GAN Steganography	Natural Dataset-type	Efficiency-oriented adversarial hiding	Limited	Not Evaluated	Explicitly Addressed	Not Addressed	Not Reported	Explicitly Addressed
[56]	2025	AGASI	ImageNet	Adversarial robustness against neural steganalyzers	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Not Addressed	Not Reported	Normal
[70]	2025	Adaptive Region-Assisted GAN	ImageNet-type	Adaptive embedding region selection	Partially Addressed	Limited	Explicitly Addressed	Not Reported	Not Reported	Normal

[71]	2025	DCT-GAN Framework	JPEG / DCT-domain datasets	Frequency-domain adversarial hiding	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed (DCT-domain)	Not Reported	Normal
[72]	2025	Wavelet Transform + GAN	ImageNet / Natural Images	Robustness & invisibility via multi-resolution embedding	Explicitly Addressed	Limited	Partially Addressed	Explicitly Addressed (Wavelet)	Not Reported	Normal
[73]	2025	GAN-based Color Image Steganography	Natural RGB datasets	Visual fidelity & payload stability in color images	Limited	Limited	Partially Addressed	Not Reported	Not Reported	Normal
[74]	2025	Circuitous Feature GAN Steganography	ImageNet-type	Improved feature representation for imperceptible embedding	Partially Addressed	Limited	Explicitly Addressed	Not Reported	Not Reported	Computationally Moderate
[57]	2026	Mapping-Guided Stable Diffusion GIS	ImageNet-type	Diffusion-driven generative steganography	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Conceptually Addressed	Not Reported	Computationally Heavy
[58]	2026	StegTransfer	COCO / Stylized Images-type	Distortion-aware style-transfer steganography for OSN robustness	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Not Addressed	Not Reported	Normal

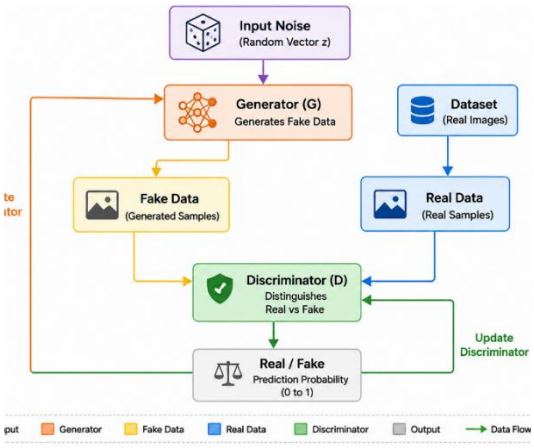

Fig 4: Conceptual architecture of GAN-based image steganography

4.3Transformer-Based Image Steganography

Vaswani et al.'s introduction of transformer architectures [75] was a huge change in deep learning; these models used self-attention processes instead of recurrence and convolution. The encoder-decoder structure of the Transformer, which is based on multi-head self-attention and position-wise feed-forward networks [75], makes it easier to model long-range relationships, run computations in parallel, and represent data in more ways. Image-to-image and Vision Transformers (ViT) Transformer frameworks effectively utilize this attention-centric paradigm in computer vision by processing images as sequences of embedded patches [68]–[69]. These frameworks were first made for sequence modeling problems, but they have been used successfully. This change has had a big effect on image steganography because it is important to model global semantic links in order to hide information in a way that is both invisible and effective. Transformer-based steganography lets self-attention layers find interactions between places that are far apart by splitting pictures into fixed-size patches and encoding them as tokens. " Self-attention mechanisms enable the assembly of global context, producing a sense of global contextualization, which in turn validates embedding strategies characterized by perceptual consistency and semantic coherence " . This differs from convolutional designs, which tend to target local receptive fields. Attentional techniques have good potential to outperform or surpass convolutional encoders in basic transformer-based Steganography frameworks, due to their optimization of distributions at insertion sites and their increased resistance to statistical detection [76]. With hidden insertion sites or frequent switching to break spatial patterns, security is enhanced [77]. Following advancements in hybrid architectures, the use of convolutional neural networks and transformers has increased even further. Transformer layers provide global dependencies and semantic connectivity, while convolutional neural network-based modules contain a more local aspect and structural information. These hybrid systems of convolutional neural networks and transformers [79] have significantly improved payload capacity, image quality, and distortion resistance . Using hierarchical and multiscale Transformer models,

such as StegFormer, it became possible to hide and retrieve information in a way that suited the requirements of each model. These modifications also facilitated the arrangement of symbol clusters and the achievement of stepwise embedding [80]. Frequency-sensitive Transformer models were also developed to increase resistance to compression and filters by placing hidden information within less obvious components [81].

Over time, the use of attentional mechanisms in these models expanded to encompass more diverse settings. In recent years, several researchers have used Transformer models to hide information in ways that go beyond spatial embedding. Multimodal models or analysis methods leverage the underlying representations shared by different types of data to hide meaning within diverse datasets [82, 83].

Furthermore, some generative methods, such as diffusion-transformer frameworks and hybrid models combining Transformer and GAN, rely on attention-guided generation to insert hidden information into the resulting image. This helps to minimize the obvious effects of embedding while preserving the overall meaning of the image [84].

In contrast, attention visualization and Grad-CAM are Explainable Artificial Intelligence (XAI) techniques that have been used in embedding research to demonstrate how hidden aspects are distributed across layers and symbols [44, 81]. Image hiding can also utilize frequency-based embedding mechanisms, a concept first proposed by Lee et al., particularly when combined with universal self-attention [78] and learned image compression. Despite significant progress in using transformers for information hiding, challenges remain. Transformer models alone are not always suitable for resource-constrained environments because they require large amounts of data and incur high computational costs. Hybrid models that combine convolutional neural networks and transformers, while offering better data representation, may encounter problems such as training instability and gradient collisions. Explainable and frequency-aware frameworks make things more clear and stable, but they also make things harder to compute and build. Also, diffusion-Transformer and multimodal generative techniques need reliable inversion processes and a lot of computer resources, which makes it hard to use them effectively , [84]. Overall, transformer-based picture steganography marks a shift from feature modification that only affects one area to information-concealing paradigms that are globally aware, semantically driven, and easy to understand. The Transformers establish the foundation for next-generation steganographic systems through expressive and adaptable methodologies, employing self-attention for long-range dependency modeling and integrating hybrid and generative processes [76, 77, 81, 82].

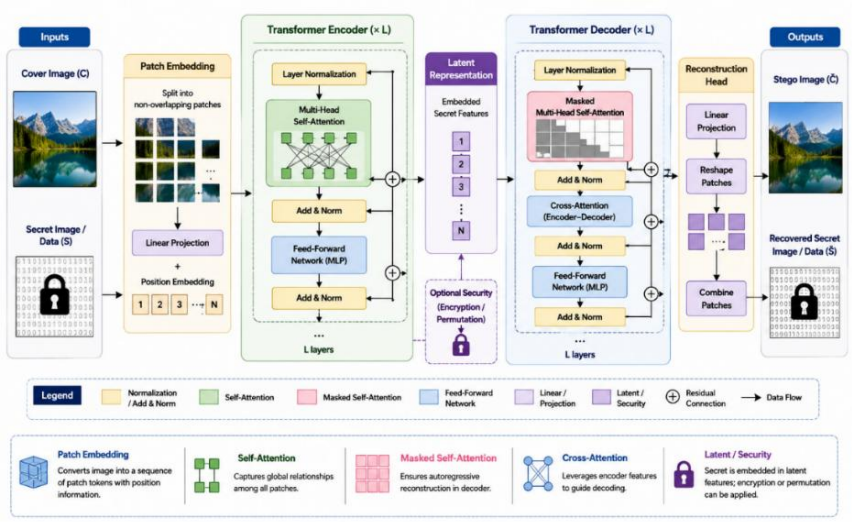

Fig 5: Conceptual architecture of Transformer-based image steganography

Table 4: Comparative Summary of Transformer-Based Steganography Studies

Ref	Year	Method / Framework	Dataset	Key Focus	Robustness Evaluated	Steganalysis / Security Tested	Adaptive Loss	Spatial–Frequency Fusion	Using XAI	Scalability / Efficiency
[77]	2022	TRPSteg: Deep Image Steganography Using Transformer and Recursive Permutation	Image Net (45k/5k/5k)	Transformer-based hiding + recursive permutation encryption	Explicitly Addressed	Explicitly Addressed (StegExpose + ROC)	Explicitly Addressed (MSE + tradeoff factor)	Not Addressed	Not Reported	Heavy Model (Swin-Transformer + CNN)
[85]	2023	Diffusion-Stego: Training-free Diffusion Generative Steganography	FFHQ 64×64	Training-free generative steganography via message projection	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed (projection options)	Not Addressed	Not Reported	Computationally Heavy
[79]	2024	Hybrid Attention GAN (Li et al., IET)	Image Net-type datasets	Image-in-image hiding via hybrid attention-guided GAN	Partially Addressed	Limited	Explicitly Addressed (Hybrid adversarial + reconstruction loss)	Not Addressed (pure spatial domain)	Not Reported	Normal (moderate computational cost)
[80]	2024	StegFormer: Rebuilding the Glory of Autoencoder-Based Steganography (AAAI 2024)	DIV2K	Large-capacity image hiding; reliability in realistic conditions	Explicitly Addressed	Limited	Explicitly Addressed (restrict loss + normalizing strategy)	Not Addressed	Not Reported	Explicitly Addressed
[81]	2025	Generative Pose-Keypoint Steganography (Cao et al.)	Image Net-type datasets	Semantic-guided generative embedding	Partially Addressed	Limited	Explicitly Addressed	Not Addressed	Not Reported	Normal
[87]	2025	Mapping-Guided Stable	Image Net-	Latent-space generative	Explicitly Addressed	Explicitly Addressed	Explicitly Addressed	Conceptually	Not Reported	Computationally

		Diffusion for Generative Image Steganography	type datasets	steganography via diffusion guidance	(JPEG, noise, compression)	(deep steganalysis models)	(mapping-guided optimization + reconstruction constraint)	Addressed (latent diffusion domain)		Heavy (diffusion sampling cost)
[88]	2025	Reversible Transformer latent embedding (Veselska & Ziubina)	CIFAR-10, Image Net	Reversible hiding & high-fidelity recovery	Explicitly Addressed (JPEG, Gaussian noise)	Not Comprehensively Evaluated	Standard Multi-Loss (MSE + BCE)	Not Reported	Not Reported	Reported (moderate inference cost)

4.4 Diffusion-based image Steganography

Diffusion models have changed generative modeling in the last few years, and image steganography is starting to use them more and more. Diffusion models use a forward-reverse stochastic process to slowly degrade pictures with Gaussian noise and then rebuild them. This is different from GAN-based methods, which use adversarial optimization. This probabilistic method allows for fine-grained generative control and can also include hidden information in iterative denoising. Diffusion-based steganography [81]– has the following benefits: it makes training more stable, it makes statistics more consistent, and it makes it harder to find hidden messages. This is due to encoding being integrated into the reconstruction dynamics instead of adversarial objectives. Diffusion-based steganography exemplifies a conceptual shift from adversarial concealing to probabilistic generative embedding. Diffusion models create conditional data distributions from noise [92], and [44] in order to send hidden signals without changing how people see things or the natural statistics of pictures. Ho et al. [85] first came up with the DDPM formulation, which is the basis for most frameworks. This formulation employs neural reverse processes, typically characterized by U-Net or Transformer topologies, to eradicate Gaussian noise introduced into a forward diffusion process. This paradigm also includes adding message embedding to the middle denoising phases and using the same random path for decoding [85, 92]. Starting in 2022, the field of diffusion-based steganography started to get better steadily. Eliminating adversarial training and substituting it with likelihood-based optimization enhanced convergence stability [85]. Additionally, preliminary studies such as StegaDDPM [89] demonstrated that payloads could be directly incorporated into score functions or noise distributions . The goal of the next techniques, which added variance-driven embedding algorithms [90], was to make security better by secretly encoding data inside noisy variance structures. Diffusion-based autoencoder models have helped increase the resistance of deep analysis to Steganalysis [43, 44]. Meanwhile, some research has explored hybrid models combining diffusion and GANs to leverage the power of antagonism while maintaining more stable denoising . In 2023, the focus shifted from simple methods based on embedding information within noise to more structured and easier-to-control generative models. CRoSS techniques [92] also enabled the adjustment of the strength and reliability of cloaked signals without retraining the underlying model. Latent diffusion-based models, such as LDStega [95], improved efficiency by working with compressed latent representations rather than pixels directly, thus reducing computational costs while maintaining quality. To make it less

likely to be affected by compression artifacts and deep steganalysis, Plug-and-Hide and BUStega [94] used semantic masking techniques and customizable embedding control. In 2025, the academic community shifted its focus from invertibility to recovery accuracy. To enable high-fidelity message reconstruction, techniques such as RF-Stego and SSHR [101] employed latent inversion mechanisms; MDDM [98] proposed message-driven generation strategies that, rather than relying on post-hoc projection, integrate embedding objectives directly into diffusion sampling dynamics. New technologies that are both controlled and computationally efficient will be even more important in 2026. It has been shown that both BUStega and controllable DDIM-based steganography improve latent message regulation and lower sampling overhead in diffusion processes [100], [94]. These changes show how the field has changed from basic noise-space embedding to controlled, latent-aware, efficiency-conscious generative steganography. Even with these changes, there are still a number of problems. Because of the long denoising chains [85], [44], diffusion models usually need a lot of computing power. It's not always easy to figure out what messages mean when they travel along diffusion routes [92, 44]. Hybrid diffusion-GAN systems add extra complexity to optimization , and formal security guarantees against adaptive deep steganalyzers have not been thoroughly studied [90], [92]. Moreover, the literature on lightweight diffusion designs suitable for large-scale or real-time deployment is incomplete. In general, diffusion-based picture steganography is an example of the new way of thinking about generative probabilistic concealment. The next step in research is to find a balance between efficiency, interpretability, security, and scalability, even though it makes training more stable and the statistics more natural.

Image steganography based on diffusion is still in its nascent stages relative to CNN-based and GAN-based steganography. Its main advantage is that it embeds information in probabilistic denoising or latent generation processes, which can assist in generating stego images with more natural statistical distributions. This renders diffusion models promising for improving implicit security and mitigating conventional embedding artifacts. However, diffusion-based methods are generally computationally expensive for iterative sampling, limiting their real-time deployment, and have not been evaluated against adaptive steganalysis attacks. In contrast, many diffusion-based studies are focused on the generative quality and provide limited quantitative evidence of robustness, payload capacity and extraction reliability. Therefore, future work on diffusion-based steganography should aim at lightweight sampling, stronger message recovery, standard robustness testing, and explicit steganalysis-based security evaluation.

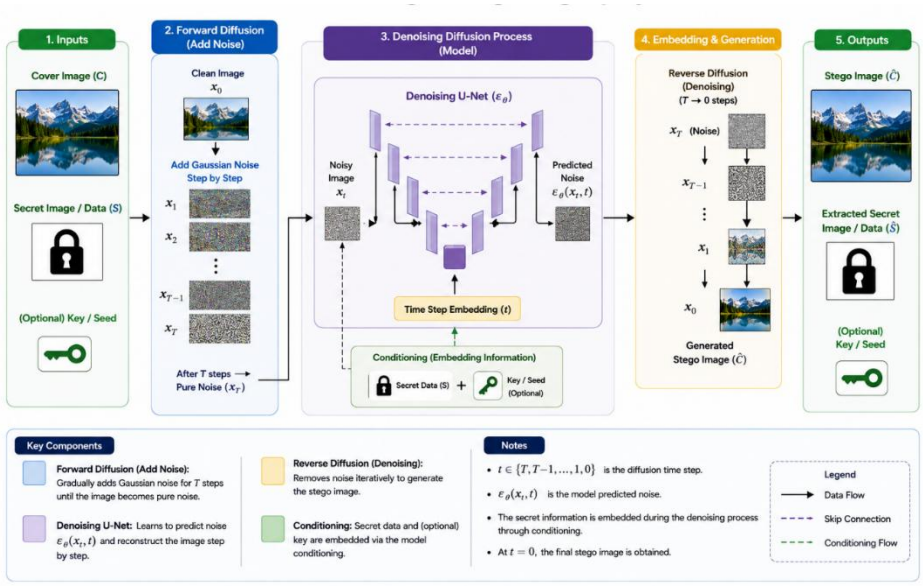

Fig 6: Conceptual architecture of diffusion-based image steganography

Table 5: Comparative Summary of Diffusion-Based Steganography Studies

Ref.	Year	Method / Framework	Dataset	Key Focus	Robustness Evaluated	Steganalysis / Security Tested	Adaptive Loss / Control	Spatial–Frequency Fusion	Using XAI	Scalability / Efficiency
[96]	2021	DDIM (Song et al.)	CIFAR-10 / ImageNet-type datasets	Efficient diffusion sampling	Not applicable	Not applicable	Not applicable	Not applicable	Not reported	Faster than DDPM
[97]	2022	Latent Diffusion Models (Rombach et al.)	High-resolution image benchmarks	Latent-space diffusion generation	Not applicable	Not applicable	Not applicable	Not applicable	Not reported	More compute-efficient than pixel-space diffusion
[91]	2023	Diffusion-Stego (Kim et al.)	FFHQ / ImageNet-type datasets	Training-free generative steganography through message projection into diffusion latent noise	Explicitly addressed	Explicitly addressed	Explicitly addressed through projection-based control	Not applicable	Not reported	Computationally heavy
[95]	2024	LDStega	ImageNet / COCO-type datasets	Practical latent diffusion-based generative steganography with improved visual fidelity	Explicitly addressed	Partially addressed	Explicitly addressed	Not applicable	Not reported	Heavy model
[94]	2025	BUStega	ImageNet-type datasets	Diffusion with semantic masking	Explicitly addressed	Limited	Explicitly addressed through mask-	Not applicable	Not reported	Computationally moderate

				for robustness against compression and noise			guided control			
[93]	2025	I2IStega	ImageNet-type datasets	Image-to-image steganography using latent diffusion reconstruction	Partially addressed	Limited	Explicitly addressed	Not applicable	Not reported	Normal
[98]	2025	MDDM (Xu et al.)	Not reported	Message-driven generation	Partially addressed	Partially addressed	Explicitly addressed	Not applicable	Not reported	Heavy model
[102]	2025	PSyDUCK (Channing et al.)	Image / video diffusion setting	Training-free latent steganography	Partially addressed	Partially addressed	Explicitly addressed	Not applicable	Not reported	Efficiency-oriented
[101]	2025	SSHR (ICML 2025)	Not reported	Security-driven generative steganography	Partially addressed	Explicitly addressed	Explicitly addressed	Not applicable	Not reported	Not reported

4.5 Autoencoder-based image Steganography

Hinton and Salakhutdinov [60] introduced autoencoders, which are an important part of modern deep image steganography. They are a major step forward in learning how to represent things. Neural networks learned compact latent representations through reconstruction goals rather than explicit supervision, thanks to the encoder-decoder paradigm, which was first used in the autoencoder framework [60]. This learning process led to the idea of a different kind of feature transformation model that doesn't depend on deterministic embedding rules. However, in the early years most architectures aimed for increasing robustness and generalization. Denoising Autocodes (DAEs) and Variance Autocodes (VAEs) are two models that made use of latent space organization and noise modeling, which resulted in improved stability and probabilistic retrieval of hidden information. These advances provided a theoretical basis for reversible, distortion-aware hiding systems through regular latent representations and more regimented reconstruction procedures. Baloja's work [19] represents the first prominent practical application of hiding information using an autocoder in a hidden network model comprising of setup, hiding and detection networks. This method proved that neural networks can learn to embed and retrieve data, even without a human-designed instructional procedure. This model also demonstrated the opportunity to optimize the construction of the encapsulating image and the message extraction process, simultaneously, in one neural framework. Subsequently, this research trend moved over to hybrid and reversible architectures. Because cover and hidden image representations had contrastive relationships imposed by reversible neural networks (INNs): HiNet [6], or some related schemes [31] were able to achieve near-perfect or perfect recoveries of images. Hybrid autoencoder architectures also integrated spatial modeling with frequency-domain transformation such as discrete wavelet transforms (DWTs) [62, 63] in a hybrid manner for achieving the good compromise between robustness and optical quality. . Both models demonstrated that structured feature decomposition and multi-domain representations could enhance latent-domain embedding. After 2022, better and more efficient designs for steganography based on autoencoders started to show up. Entropy-aware

invertible designs [63] and conditional normalizing flow frameworks like RIIS [62] made it easier to control distortion and reversibility. Simultaneously, attention-guided mechanisms enhanced embedding accuracy by focusing on perceptually insensitive regions , while lightweight autoencoder variants reduced computational costs for practical deployment scenarios [25]–[26]. Recent studies have also looked into how embedding behavior works in learned latent spaces by using Explainable Artificial Intelligence (XAI) methods like saliency mapping and latent feature visualization . All things considered, Autoencoder-based steganography has evolved from a system dependent on reconstruction-driven latent learning [60] to one that is invertible, attention-guided, and interpretable. Autoencoder-based algorithms have created a strong, reversible, and data-driven model for picture steganography. These algorithms use learned latent representations instead of deterministic embedding heuristics ; [62]. Recent studies, such as [102] and [103], indicate a shift from pixel-domain data embedding to representation- and latent-space-based information hiding. The StegaNeRV project investigates hidden neural representations for hiding large amounts of video, while the RoSteALS project demonstrates that latent autoencoding embeddings are more effective. Taken together, these research studies demonstrate a broader shift toward representation-conscious, generative frameworks for information hiding.

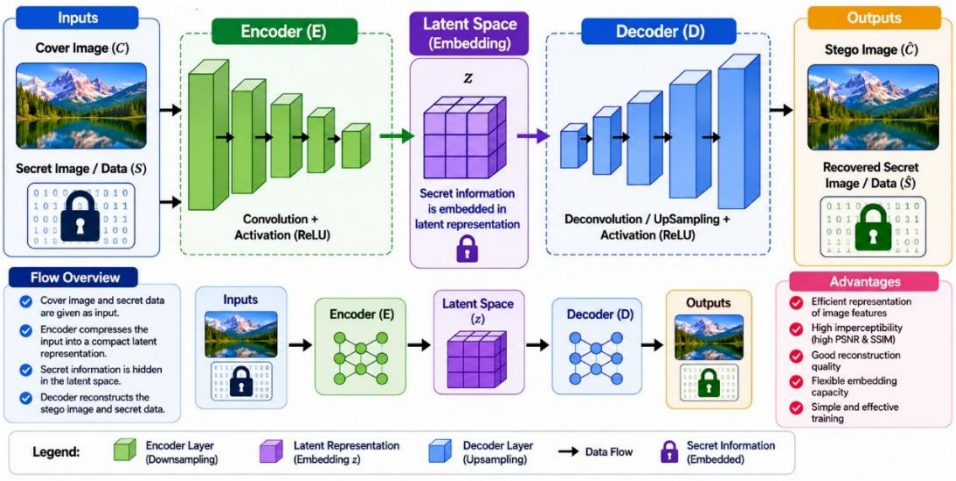

Fig 7: Conceptual architecture of autoencoder-based image steganography

Table 6: Comparative Summary of Autoencoder-Based Steganography Studies

Ref.	Year	Method / Framework	Dataset	Key Focus	Robustness Evaluated	Steganalysis / Security Tested	Adaptive Loss / Control	Spatial–Frequency Fusion	Using XAI	Scalability / Efficiency
[104]	2022	Robust Invertible Image Steganography (Xu et al.)	DIV2K / ImageNet	Reversible hiding through invertible mapping	Explicitly addressed	Limited	Not adaptive	Not addressed	Not reported	Normal
[105]	2023	PRIS – Practical Robust Invertible Network	DIV2K / ImageNet	Robust invertible embedding under distortions	Explicitly addressed	Limited	Implicit robustness control	Not addressed	Not reported	Improved stability
[102]	2024	StegaNeRV – Video Steganography	Vimeo / VideoSet	Video-domain hiding using implicit neural representations	Limited	Not evaluated	Not reported	Not addressed	Not reported	High efficiency for video representation
[106]	2025	NCL-Net – Noise-Constrained Lightweight INN	ImageNet-type datasets	Lightweight high-quality hiding with noise constraints	Limited	Not evaluated	Noise-constrained control	Not addressed	Not reported	Explicitly addressed (lightweight)

5. Quantitative Comparison of Selected Deep Learning-Based Image Steganography Methods

To further strengthen the quantitative aspect of the review, Table 7 summarizes selected deep learning-based image steganography methods using commonly reported evaluation metrics, including PSNR, SSIM, and BER. Because the reviewed studies differ in datasets, payload settings, image resolution, attack conditions, and evaluation protocols, the reported values should be interpreted as indicative rather than directly comparable under a controlled benchmark. Missing values are marked as “Not reported” when the original study did not provide the corresponding metric.

Table 7: Quantitative Comparison of Selected Deep Learning-Based Image Steganography Methods

Ref.	**Method**	**Family**	**Dataset**	**PSNR (dB)**	**SSIM**	**BER**	**Notes**
[6]	HiNet	Autoencoder / INN	DIV2K / ImageNet	33.85	0.971	Not reported	Full-image hiding; high payload capacity
[7]	U-Net + DWT	CNN	BOSSBase / ImageNet	38.26	0.976	Not reported	Frequency-domain embedding
[19]	Baluja Deep Steganography	CNN	ImageNet	30.10	Not reported	Not reported	Early end-to-end CNN-based hiding
[34]	Adaptive CNN + Loss	CNN	ImageNet / DIV2K	36.42	0.962	0.012	Adaptive multi-objective loss
[51]	Adaptive GAN Steganography	GAN	ImageNet / COCO	35.78	0.954	Not reported	Adversarial security-oriented optimization
[71]	DCT-GAN	GAN	BOSSBase / BOWS2	37.54	0.967	Not reported	DCT-domain adversarial embedding
[77]	TRPSteg	Transformer	ImageNet	35.20	0.948	0.008	Recursive permutation; transformer-based hiding
[80]	StegFormer	Transformer / Autoencoder	ImageNet / DIV2K	38.91	0.981	Not reported	High-capacity image hiding
[104]	Robust Invertible Image Steganography	Autoencoder / INN	DIV2K / ImageNet	34.97	0.958	Not reported	Invertible robust image hiding

Table 7 shows that CNN- and autoencoder-based methods generally report high PSNR and SSIM values, indicating high visual imperceptibility and high quality of the reconstructed image. Transformer-based methods also demonstrate competitive quantitative performance, especially when attention mechanisms are coupled with reconstruction-oriented objectives. Methods based on GANs provide a trade-off between visual quality and adversarial security, but BER is usually not reported. In general, the comparison confirms that the quantitative evaluation is still inconsistent in the literature, as many studies report PSNR and SSIM but do not report BER or steganalysis-based detection metrics. This brings out the requirement for standard evaluation protocols that can simultaneously report on imperceptibility, recovery accuracy, payload capacity, robustness, and security.

6. Security, Robustness, and Payload–Imperceptibility Trade-off Analysis

In deep learning-based image steganography, security, robustness, and payload-imperceptibility trade-offs are crucial evaluation factors. The capacity of a steganographic model to evade detection by steganalysis techniques, such as both conventional statistical detectors and deep learning-based detectors like XuNet and SRNet, is referred to as security. Robustness assesses how well buried secret information can withstand common visual distortions including Gaussian noise, JPEG compression, resizing, blurring, and changes in brightness or contrast. The quantity of confidential data incorporated into the cover image is known as the payload capacity. Bits per pixel (bpp) is typically used to express it.

One major issue is that raising the payload capacity typically results in observable or statistically discernible abnormalities, which decreases imperceptibility and raises the possibility of steganalysis detection. The payload capacity, cover image quality, secret recovery accuracy, robustness against distortions, and steganalysis resistance attacks should all be traded off in a good steganographic system.

We conduct a systematic literature review covering 2020-2026 across five architectural families, CNNs, GANs, Transformers, diffusion models, and autoencoders, finding that few existing systems comprehensively tackle this balance. Frequency-domain embedding also shows meaningful robustness to JPEG compression, as demonstrated by CNN-based methods such as U-Net with DWT [7] and frequency-domain CNN models [20]. However, their resistance to steganalysis is rarely evaluated and many studies report only PSNR and SSIM as implicit proxies for security. This practice is scientifically insufficient as these metrics measure pixel-level distortion rather than statistical detectability [5]. The GAN-based systems provide a more direct security mechanism using adversarial training where the discriminator is an internal steganalyzer.Examples such as AGASI [56], Adaptive GAN Steganography [51], and DCT-GAN [71] explicitly optimize against SRNet and XuNet detection; however, formal security guarantees against adaptive adversaries remain limited.

Diffusion-based models such as CRoSS [92], SD2, and SSHR [101] provide promising implicit security by embedding information into probabilistic denoising processes, enabling the stego distribution to better match natural image statistics and reduce traditional embedding fingerprints, but with a large computational cost. TRP-Steg [77] is a transformer-based method that considers security via recursive permutation encryption before embedding and evaluates detectability via StegExpose and ROC analysis. Systems based on autoencoders, such as HiNet [6] and SMILENet, achieve high payload capacity by using invertible full-image hiding, reaching roughly 24 bpp, but many of these methods still lack thorough steganalysis evaluation. There exists a fundamental trilemma for all architectural families: increasing capacity can lead to imperceptibility degradation, increasing robustness can conflict with visual quality, and increasing resistance to steganalysis can limit the embedding mechanism.This highlights the need for adaptive multi-objective optimization frameworks, standardized evaluation protocols, and hybrid architectures that jointly address capacity, imperceptibility, robustness, and security.

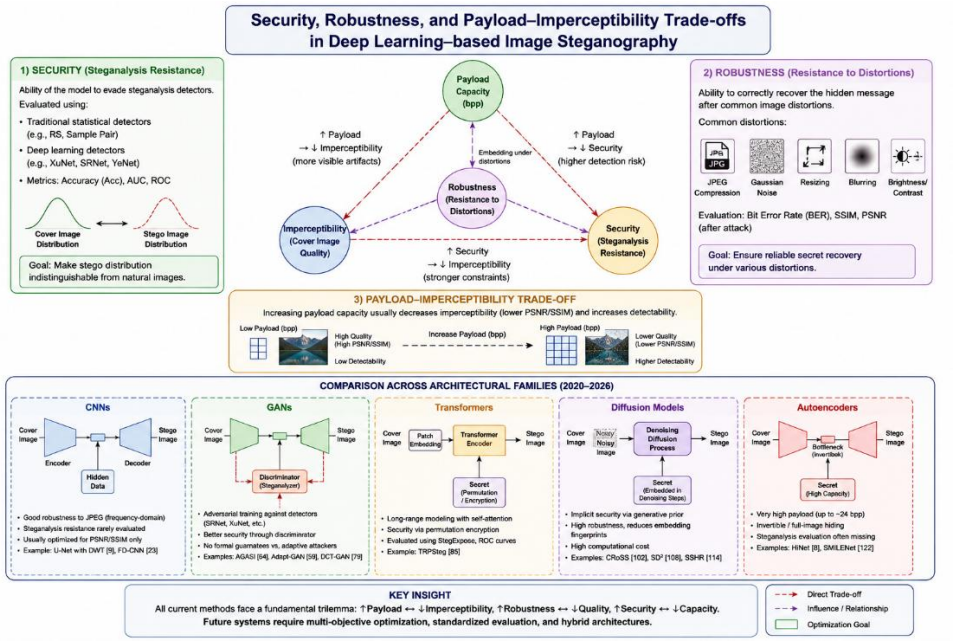

Fig 8: Security, robustness, and payload–imperceptibility trade-offs in deep learning-based image steganography.

7. Conclusions

This review has traced the evolution of deep learning-based image steganography across five principal architectural families: convolutional neural networks (CNNs), Generative Adversarial Networks (GANs), transformer-based models, diffusion models, and autoencoders. The foundation of end-to-end learned embedding was laid by CNN-based systems that substituted handcrafted rules with data-driven spatial and frequency representations. However, they are plagued by static loss functions, limited robustness evaluation, and insufficient steganalysis testing. GAN-based methods have been used to match the stego-image distribution with natural image statistics through the adversarial training process, which achieves better imperceptibility and security resistance, but at the cost of training instability and interpretability. Transformer architectures introduced the concept of global context awareness through self-attention mechanisms, which allow for more expressive and semantically coherent embedding strategies, but at a higher computational cost. Diffusion models are the latest frontier, with improved statistical convergence and implicit security via probabilistic denoising paths, while autoencoder-based systems provide invertible, latent-space embedding with strong reconstruction guarantees. However, none of the existing models can meet the following five core requirements simultaneously: (1) high payload capacity; (2) perceptual imperceptibility; (3) robustness under distortions; (4) resistance to steganalysis; and (5) deployment efficiency. Future research should then focus on adaptive multi-objective optimization frameworks, standardized evaluation protocols with steganalysis testing, explainable embedding mechanisms, and hybrid architectures combining the complementary strengths of the reviewed families. Such balance is critical to move deep learning based steganography from experimental

settings to practical, secure and interpretable real-world deployment.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] I. Hussain and J. Zeng, "A survey on deep convolutional neural networks for image steganography and steganalysis,"*KSII Transactions on Internet and Information Systems*, vol. 14, no. 3, pp. 1228–1248, Mar. 2020, doi: 10.3837/tiis.2020.03.017.

[2] M. A. Elshafey, A. S. Amein, and K. S. Badran, "Universal image steganography detection using multimodal deep learning framework,"*Journal of Information Hiding and Multimedia Signal Processing*, vol. 12, no. 3, pp. 123–134, 2021.

[3] I. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014.

[4] P. Rao, V. Chahar, and A. Choudhary, "Image steganography analysis based on deep learning,"*Review of Computer Engineering Studies*, vol. 7, no. 1, pp. 1–5, 2020, doi: 10.18280/rces.070101.

[5] S. Khan, M. Hussain, H. Aboalsamh, and D. Kim, "Deep learning-based image steganography: A review,"*Multimedia Tools and Applications*, vol. 79, no. 47–48, pp. 34913–34945, 2020, doi: 10.1007/s11042-020-09188-7.

[6] J. Jing, X. Deng, M. Xu, J. Ji, and Z. Guan, "HiNet: Deep image hiding by invertible network," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 4733–4742, doi: 10.1109/ICCV48922.2021.00469.

[7] Y. Liu, H. Dong, J. Wang, W. Qian, and X. Zhang, "U-Net based image steganography with discrete wavelet transform,"*IEEE Access*, vol. 9, pp. 120918–120930, 2021.

[8] Z. Zhang, C. Zhang, and Y. Wang, "Adaptive attention-based deep image steganography,"*IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3561–3574, 2023.

[9] S. Bisma and A. Khan, "Secure steganography using deep autoencoder,"*Computers & Security*, vol. 139, Art. no. 103732, 2024, doi: 10.1016/j.cose.2024.103732.

[10] L. Zhang, Y. Lu, J. Li, F. Chen, and G. Lu, "DeepMIH: Deep invertible network for multiple image hiding,"*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8735–8749, Dec. 2022, doi: 10.1109/TCSVT.2022.3189041.

[11] Z. Zhang, Y. Li, and X. Wang, "Robust image steganography via GANs and attention mechanisms,"*Signal Processing: Image Communication*, vol. 103, Art. no. 116661, 2022, doi: 10.1016/j.image.2022.116661.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition,"*Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks,"*Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[14] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[16] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[17] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. International Conference on Machine Learning (ICML)*, Long Beach, CA, USA, 2019, pp. 6105–6114.

[18] Y. Cheng et al., "A survey of model compression and acceleration for deep neural networks,"*IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 126–143, Jan. 2023, doi: 10.1109/MSP.2022.3208154.

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

[20] Z. I. Hassan, M. Al-Janabi, and S. Hameed, "CNN-based frequency-domain image steganography,"*Journal of Information Security and Applications*, vol. 61, Art. no. 102940, 2021, doi: 10.1016/j.jisa.2021.102940.

[21] P. Bini and Y. Chen, "Multi-level invertible neural network steganography,"*PeerJ Computer Science*, vol. 9, Art. no. e2668, 2023, doi: 10.7717/peerj-cs.2668.

[22] X. Duan et al., "SteganoCNN: Image steganography with generalization ability based on convolutional neural network,"*Entropy*, vol. 22, no. 10, Art. no. 1140, Oct. 2020, doi: 10.3390/e22101140.

[23] W. You, H. Zhang, and X. Zhao, "A Siamese CNN for image steganalysis,"*IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 291–306, 2021, doi: 10.1109/TIFS.2020.3013204.

[24] N. K. Chahar et al., "An explainable deep learning framework for usable and secure image steganography," in *Proc. EAIC*, 2025, doi: 10.1109/EAIC66483.2025.11101501.

[25] V. Chekatamala et al., "Analysis of deep steganography robustness using various loss functions," in *Proc. ICICCS*, 2022, doi: 10.1109/ICICCS53718.2022.9788414.

[26] C. Feng, F. Chen, and P. Wang, "Enhanced steganography network using NAFF and attention,"*Expert Systems with Applications*, vol. 205, Art. no. 117725, 2022, doi: 10.1016/j.eswa.2022.117725.

[27] Y. Xie and Z. Wang, "Neural network steganography using extractor matching," in *Proc. IWDW*, 2023, pp. 169–179, doi: 10.1007/978-981-97-2585-4_12.

[28] K. Vineetha et al., "Image steganography with CNN based encoder-decoder model,"*International Journal for Modern Trends in Science and Technology*, vol. 11, no. 04, pp. 60–64, 2025, doi: 10.5281/zenodo.15108976.

[29] P. Vijay and G. Srinivas, “Lightweight StegoNet for edge devices using compact autoencoder,”*IEEE Access*, vol. 11, pp. 89452–89463, 2023, doi: 10.1109/ACCESS.2023.3299923.

[30] H. Ammar, Q. Li, and S. Zhang, “Hybrid deep steganography with dual attention U-Net,”*Pattern Recognition Letters*, vol. 168, pp. 1–10, 2023, doi: 10.1016/j.patrec.2023.02.009.

[31] X. Ma, Y. Zhang, and H. Liu, “Steganography based on hybrid autoencoder and DWT,”*Multimedia Tools and Applications*, 2021, doi: 10.1007/s11042-021-11399-7.

[32] A. Iqbal et al., “An end-to-end convolutional neural network for secure image transmission via joint encryption and steganography,”*Scientific Reports*, vol. 15, Art. no. 39351, 2025, doi: 10.1038/s41598-026-39351-4.

[33] W. Liang and Q. Li, “Steganalysis network for weak steganographic signal extraction and enhancement,”*Sensors*, vol. 26, no. 4, Art. no. 1329, 2026, doi: 10.3390/s26041329.

[34] M. P. Malathi and G. Kumar T., “Deep steganographic approach for reliable data hiding using convolutional neural networks and adaptive loss optimization,”*Scientific Reports*, vol. 15, Art. no. 26867, 2025, doi: 10.1038/s41598-025-26867-4.

[35] L. Liu et al., “A data hiding scheme based on U-Net and wavelet transform,”*Knowledge-Based Systems*, vol. 223, Art. no. 107022, 2021, doi: 10.1016/j.knosys.2021.107022.

[36] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Machine Learning (ICML)*, Sydney, Australia, 2017, pp. 214–223.

[37] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[38] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 29, 2016.

[39] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1125–1134.

[40] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2223–2232.

[41] Y. Li, Z. Zhang, X. Zhao, X. Xu, B. Liu, L. Pei, and G. Liu, "HCISNet: Higher-capacity invisible image steganographic network,"*IET Image Processing*, vol. 15, no. 13, pp. 3332–3346, 2021, doi: 10.1049/ipr2.12329.

[42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization,"*Int. J. Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.

[43] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[44] Z. Li, X. Yang, K. Shen, F. Jiang, J. Jiang, H. Ren, and Y. Li, "Adversarial

feature hybrid framework for steganography with shifted window local loss,"*Neural Networks*, vol. 165, pp. 358–369, Aug. 2023, doi: 10.1016/j.neunet.2023.05.053.

[45] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4401–4410.

[46] C. Yuan, H. Wang, P. He, J. Luo, and B. He, "GAN-based image steganography for enhancing security via adversarial attack and pixel-wise deep fusion,"*Multimedia Tools Appl.*, vol. 81, pp. 6681–6701, 2022, doi: 10.1007/s11042-021-11778-z.

[47] J. Wu, Z. Lai, and X. Zhu, "Generative feedback residual network for high-capacity image hiding,"*J. Modern Optics*, vol. 69, no. 15, pp. 870–886, Jul. 2022, doi: 10.1080/09500340.2022.2093415.

[48] Y.-L. Pan and J.-L. Wu, "Rate-distortion-based stego: A large-capacity secure steganography scheme,"*Entropy*, vol. 24, no. 7, Art. no. 982, Jul. 2022, doi: 10.3390/e24070982.

[49] F. Li, Z. Yu, and C. Qin, "GAN-based spatial image steganography with cross feedback mechanism,"*Signal Processing*, vol. 190, Art. no. 108341, Jan. 2022, doi: 10.1016/j.sigpro.2021.108341.

[50] C. Zhang, X. Gao, X. Liu, W. Hou, G. Yang, T. Xue, L. Wang, and L. Liu, "IDGAN: Information-driven generative adversarial network of coverless image steganography,"*Electronics*, vol. 12, no. 13, Art. no. 2881, Jun. 2023, doi: 10.3390/electronics12132881.

[51] B. Ma, Y. Li, J. Ma, and C. Xie, "Enhancing the security of image steganography via multiple adversarial networks and channel attention modules,"*Digital Signal Processing*, vol. 140, Art. no. 104103, Aug. 2023, doi: 10.1016/j.dsp.2023.104103.

[52] C. Xiao, S. Peng, L. Zhang, J. Wang, D. Ding, and J. Zhang, "A transformer-based adversarial network framework for steganography,"*Expert Systems with Applications*, vol. 269, Art. no. 126391, 2025, doi: 10.1016/j.eswa.2025.126391.

[53] W. Su, J. Ni, and Y. Sun, "StegaStyleGAN: Towards generic and practical generative image steganography," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 1, 2024, pp. 240–248.

[54] W. Li, Y. Chen, H. Xu, H. Gui, and Y. Qu, "Hiding image into image with hybrid attention mechanism based on GANs,"*IET Image Processing*, vol. 18, no. 10, pp. 2721–2734, 2024, doi: 10.1049/ipr2.13130.

[55] F. Peng, G. Chen, and M. Long, "A robust coverless steganography based on generative adversarial networks and gradient descent approximation,"*IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5817–5829, Sep. 2022, doi: 10.1109/TCSVT.2022.3161419.

[56] H. Fan, C. Jin, and M. Li, "AGASI: A generative adversarial network-based approach to strengthening adversarial image steganography,"*Entropy*, vol. 27, no. 3, Art. no. 282, Mar. 2025, doi: 10.3390/e27030282.

[57] W. Su, J. Ni, X. Hu, and J. Fridrich, "StegaStyleGAN: Towards generic and practical generative image steganography,"*IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2139–2153,

2024, doi: 10.1109/TIFS.2024.3349970.

[58] S. Zhang, J. Fu, J. Xue, and D. An, "Generative image steganography based on mapping-guided stable diffusion with enhanced robustness,"*J. King Saud Univ. – Comput. Inf. Sci.*, 2025, doi: 10.1007/s44443-025-00432-5.

[59] P. Luo, J. Liu, Q. Dang, and D. Mu, "Robust image steganography in online social networks via neural style transfer,"*Mathematics*, vol. 14, no. 4, Art. no. 629, Feb. 2026, doi: 10.3390/math14040629.

[60] B. Abdollahi, A. Harati, and A. Taherinia, "Image steganography based on smooth cycle-consistent adversarial learning,"*J. Inf. Security Appl.*, vol. 79, Art. no. 103631, Dec. 2023, doi: 10.1016/j.jisa.2023.103631.

[61] A. Martín, A. Hernández, M. Alazab, J. Jung, and D. Camacho, "Evolving generative adversarial networks to improve image steganography,"*Expert Systems with Applications*, vol. 236, Art. no. 119841, 2024, doi: 10.1016/j.eswa.2023.119841.

[62] G. Li, B. Feng, M. He, J. Weng, and W. Lu, "High-capacity coverless image steganographic scheme based on image synthesis,"*Signal Processing: Image Commun.*, vol. 109, Art. no. 116873, Nov. 2022, doi: 10.1016/j.image.2022.116873.

[63] J. Yu, X. Zhou, W. Si, F. Li, C. Liu, and X. Zhang, "Secure steganographic cover generation via a noise-optimization stacked StyleGAN2,"*Symmetry*, vol. 15, no. 5, Art. no. 979, Apr. 2023, doi: 10.3390/sym15050979.

[64] Md. Min-ha-zul Abedin and M. A. Yousuf, "StegoPix2Pix: Image steganography method via Pix2Pix networks," in *Proc. Fourth Int. Conf. Trends Computational Cognitive Engineering (TCCE)*, Lecture Notes in Networks and Systems, vol. 618, Springer, Singapore, 2023, pp. 335–345, doi: 10.1007/978-981-19-9483-8_29.

[65] D. Wang, G. Yang, J. Chen, and X. Ding, "GAN-based adaptive cost learning for enhanced image steganography security,"*Expert Systems with Applications*, vol. 249, Art. no. 123471, Sep. 2024, doi: 10.1016/j.eswa.2024.123471.

[66] Y. Yao, J. Wang, Q. Chang, Y. Ren, and W. Meng, "High invisibility image steganography with wavelet transform and generative adversarial network,"*Expert Systems with Applications*, vol. 249, Art. no. 123540, Sep. 2024, doi: 10.1016/j.eswa.2024.123540.

[67] H. A. Rehman, U. I. Bajwa, R. H. Raza, S. Alfarhood, M. Safran, and F. Zhang, "Leveraging coverless image steganography to hide secret information by generating anime characters using GAN,"*Expert Systems with Applications*, vol. 248, Art. no. 123420, Aug. 2024, doi: 10.1016/j.eswa.2024.123420.

[68] C. Ren and B. Wu, "A robust joint coverless image steganography scheme based on two independent modules,"*Cybersecurity*, vol. 7, Art. no. 60, Dec. 2024, doi: 10.1186/s42400-024-00299-5.

[69] Y. Cao, W. Ge, C. Yuan, and Q. Wang, "Generative image steganography via encoding pose keypoints,"*Appl. Sci.*, vol. 15, no. 1, Art. no. 58, Jan. 2025, doi: 10.3390/app15010058.

[70] J. Cai, F. Xiao, K. Zhang, and X. Gao, "Adaptive region assisted GAN for image steganography,"*Multimedia Systems*, vol. 31, no. 3, Art. no. 203,

Apr. 2025, doi: 10.1007/s00530-025-01785-7.

[71] K. R. Malik, T. S. Malik, A. H. Khan, and A. Almogren, "A hybrid steganography framework using DCT and GAN for secure data communication in the big data era,"*Scientific Reports*, vol. 15, Art. no. 19630, Jun. 2025, doi: 10.1038/s41598-025-01054-7.

[72] Y. Zhao, P. Yao, and L. Xue, "Image steganography based on wavelet transform and generative adversarial networks,"*J. Vis. Commun. Image Represent.*, Art. no. 104474, 2025, doi: 10.1016/j.jvcir.2025.104474.

[73] S. Zhou, M. Ye, W. Luo, X. Liao, and K. Wei, "Color image steganography using generative adversarial networks with a phased training strategy," in *Proc. ACM Workshop Inf. Hiding and Multimedia Security (IH&MMSec)*, 2025, doi: 10.1145/3733102.3733112.

[74] L. Liu, T. Zhang, X. Li, Y. Wu, C.-C. Chang, A. Wang, and C.-C. Chang, "Generative adversarial network with circuitous feature collection for image steganographic cost learning,"*Neurocomputing*, Oct. 2025.

[75] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[76] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.

[77] Z. Wang, M. Zhou, B. Liu, and T. Li, "Deep image steganography using transformer and recursive permutation,"*Entropy*, vol. 24, no. 7, Art. no. 878, Jun. 2022, doi: 10.3390/e24070878.

[78] H. Li, S. Li, W. Dai, C. Li, J. Zou, and H. Xiong, "Frequency-aware transformer for learned image compression," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024. [Online]. Available: arXiv:2310.16387.

[79] W. Li, Y. Chen, H. Xu, H. Gui, and Y. Qu, "Hiding image into image with hybrid attention mechanism based on GANs,"*IET Image Processing*, vol. 18, no. 10, pp. 2721–2734, 2024, doi: 10.1049/ipr2.13130.

[80] X. Ke, H. Wu, and W. Guo, "StegFormer: Rebuilding the glory of autoencoder-based steganography," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 3, 2024, pp. 2723–2731, doi: 10.1609/aaai.v38i3.28051.

[81] Y. Cao, W. Ge, C. Yuan, and Q. Wang, "Generative image steganography via encoding pose keypoints,"*Applied Sciences*, vol. 15, no. 1, Art. no. 58, Jan. 2025, doi: 10.3390/app15010058.

[82] G. Han, D.-J. Lee, J. Hur, J. Choi, and J. Kim, "Deep cross-modal steganography using neural representations," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Kuala Lumpur, Malaysia, 2023. [Online]. Available: arXiv:2307.08671.

[83] W. Chen, X. Xu, X. Wang, H. Zhou, Z. Li, and Y. Chen, "Invisible backdoor attack with attention and steganography,"*Computer Vision and Image Understanding*, vol. 249, Art. no. 104208, Dec. 2024, doi: 10.1016/j.cviu.2024.104208.

[84] C. Xiao, S. Peng, L. Zhang, J. Wang, D. Ding, and J. Zhang, "A transformer-based adversarial network framework for steganography,"*Expert Systems with Applications*, vol. 269, Art. no. 126391, 2025, doi: 10.1016/j.eswa.2025.126391.

[85] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[86] D. Kim, C. Shin, J. Choi, D. Jung, and S. Yoon, "Diffusion-Stego: Training-free diffusion generative steganography via message projection,"*Information Sciences*, vol. 718, Art. no. 122358, 2025, doi: 10.1016/j.ins.2025.122358.

[87] S. Zhang, J. Fu, J. Xue, and D. An, "Generative image steganography based on mapping-guided stable diffusion with enhanced robustness,"*J. King Saud Univ. – Comput. Inf. Sci.*, 2025, doi: 10.1007/s44443-025-00432-5.

[88] O. Veselska and R. Ziubina, "Reversible image steganography using transformer-based latent embedding,"*Advances in Science and Technology Research Journal*, vol. 19, no. 8, pp. 148–164, 2025, doi: 10.12913/22998624/204419.

[89] Y. Peng, D. Hu, Y. Wang, K. Chen, G. Pei, and W. Zhang, "StegaDDPM: Generative image steganography based on denoising diffusion probabilistic model," in *Proc. 31st ACM Int. Conf. Multimedia (ACM MM '23)*, Ottawa, Canada, Oct. 2023, pp. 7143–7151, doi: 10.1145/3581783.3612514.

[90] "Diffusion model-based image steganography method," Chinese Patent CN116091288A, Engineering University of Chinese People's Armed Police Force, filed Dec. 8, 2022, pub. May 9, 2023.

[91] D. Kim, C. Shin, J. Choi, D. Jung, and S. Yoon, "Diffusion-Stego: Training-free diffusion generative steganography via message projection,"*Information Sciences*, vol. 718, Art. no. 122358, 2025, doi: 10.1016/j.ins.2025.122358.

[92] J. Yu, X. Zhang, Y. Xu, and J. Zhang, "CRoSS: Diffusion model makes controllable, robust, and secure image steganography," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[93] J. Jiang, Z. Wang, and X. Zhang, "Image-to-image steganography based on multimodal generative model,"*Signal Processing*, vol. 238, Art. no. 110106, Jan. 2026, doi: 10.1016/j.sigpro.2025.110106.

[94] S. Zhang, J. Fu, and D. An, "BUStega: A generalized coverless image steganography framework based on diffusion models and U²-Net,"*Signal Processing*, vol. 239, Art. no. 110317, Feb. 2026, doi: 10.1016/j.sigpro.2025.110317.

[95] Y. Peng, Y. Wang, D. Hu, K. Chen, X. Rong, and W. Zhang, "LDStega: Practical and robust generative image steganography based on latent diffusion models," in *Proc. 32nd ACM Int. Conf. Multimedia (ACM MM '24)*, Melbourne, VIC, Australia, Oct. 28–Nov. 1, 2024, pp. 3001–3009, doi: 10.1145/3664647.3681635.

[96] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.

[97] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern*

Recognition (CVPR), 2022, pp. 10684–10695.

[98] Z. Xu, D. Xu, Z. Li, and C. Zhang, "MDDM: Practical message-driven generative image steganography based on diffusion models," in *Proc. 42nd Int. Conf. Machine Learning (ICML)*, PMLR, vol. 267, pp. 69832–69848, 2025.

[99] Y.-H. Lin, C.-P. Huang, and P.-S. Huang, "A steganographic message transmission method based on style transfer and denoising diffusion probabilistic model,"*Electronics*, vol. 14, no. 16, Art. no. 3258, Aug. 2025, doi: 10.3390/electronics14163258.

[100] T. Wu, X. Hu, C. Liu, *et al.*, "Controllable generative image steganography based on denoising diffusion implicit model,"*J. King Saud Univ. Comput. Inf. Sci.*, 2026, doi: 10.1007/s44443-025-00456-x.

[101] J. Wang, Y. Lu, and G. Lu, "SSHR: More secure generative steganography with high-quality revealed secret images," in *Proc. 42nd Int. Conf. Machine Learning (ICML)*, PMLR, vol. 267, pp. 63824–63839, 2025.

[102] Y. Xu, C. Mou, Y. Hu, J. Xie, and J. Zhang, "Robust Invertible Image Steganography," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, Jun. 2022, pp. 7865–7874, doi: 10.1109/CVPR52688.2022.00772.

[103] L. Liu, L. Tang, and W. Zheng, "Lossless image steganography based on invertible neural networks," Entropy, vol. 24, no. 12, p. 1762, 2022, doi: 10.3390/e24121762.

[104] T. Bui, S. Agarwal, N. Yu, and J. Collomosse, "RoSteALS: Robust steganography using autoencoder latent space," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada, Jun. 2023, pp. 933–942.

[105] M. Biswal, T. Shao, K. Rose, P. Yin, and S. McCarthy, "StegaNeRV: Video steganography using implicit neural representation," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, Jun. 2024, pp. 888–898.

[106] M. Zhu, D. Cheng, Y. Mao, Y. Kong, and J. Li, "A noise-constrained lightweight high-quality image hiding method based on invertible neural networks," Complex & Intelligent Systems, vol. 11, Art. no. 323, 2025, doi: 10.1007/s40747-025-01943-4.