IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15750

Image Steganography: A Comprehensive Systematic Review Methods, Advances and Future Directions

Dr. Balaji K[1], Miss Dhanushree H N[2], Miss Rachana BM[3]

Department of Computer Science, Surana College, Bengaluru, Karnataka, India[1-3]

Abstract: Image steganography, which involves the embedding of data into digital images in such a way that the embedding remains undetectable, has seen dramatic changes over the past several years due to advances in deep learning, generative adversarial networks (GANs), diffusion-based generative modelling, and transformer architectures. This paper aims to synthesize recent progress within four main categories of image steganography spatial domain, transform domain, deep-learning-based, and coverless steganography via a review of 40 recent papers published from 2021 to 2025. The methods are reviewed according to a common set of metrics Peak Signal-To-Noise Ratio (PSNR), structural similarity index (SSIM), embedding capacity, invisibility, robustness against steganalysis, and computation cost. Special attention is devoted to four rapidly developing subdomains: diffusion-based generative steganography, invertible neural networks (INNs), adversarially trained hidden mechanisms, and secure methods with provable guarantees. The review additionally includes a structured taxonomy of the field, a comparative analysis based on visual data, a list of popular datasets, an overview of open challenges, and directions for future work.

Index Terms: Image steganography, deep learning, generative adversarial networks, diffusion models, invertible neural networks, Coverless steganography, steganalysis, PSNR, SSIM, systematic review.

I. INTRODUCTION

With the advancement of digital communication and cloud-based data storage taking place at a tremendous pace, the security of data when in transit has become one of the biggest challenges in computing today. The term steganography derives from two Greek words, Stefanos meaning “covered” and graphic meaning “writing,” and refers to the technique where a message is hidden within the cover message, which looks normal to an onlooker. For example, an image, an audio file or a video clip can be used as cover messages, and steganography makes sure that it becomes hard to identify that any message is hidden within it. What distinguishes it from cryptography is that it preserves the normality of the cover message, unlike cryptography, which obscures the message but not its presence.

Among all the types of carriers that exist, images have emerged as the preferred medium for steganography, mainly due to the widespread use of these images over social media, clouds, and multimedia storage systems. Traditional approaches that include the use of Least Significant Bit substitution and the frequency domain approach based on the use of DCT, DWT, and DFT have helped lay the base for what follows. However, the advent of deep learning has brought about a paradigm shift, leading to the creation of novel algorithms that perform better than traditional approaches in terms of payload capacity, fidelity, and robustness.

During the period between 2021 and 2025, there have been many developments within the domain of machine learning applied to steganography through the use of GANs, invertible neural networks (INNs), denoising diffusion probabilistic models, and vision transformers within steganographic pipelines. This results in the emergence of a new generation of methods that can make security arguments, embedding large payloads greater than 4 bits per pixel (bpp), and resist modern detection algorithms such as SRNet, XuNet, and EfficientNet [1], [2], [9], [33].

This review makes five contributions, including classification and structuring of the reviewed area, a critical overview of 40 recent papers in the area, comparison of the reviewed papers based on various metrics, presentation of the gaps in the reviewed area, and formulation of the research agenda for future works

A. Review Methodology

The candidate papers were identified by searching Google Scholar, IEEE Xplore, ScienceDirect, ACM Digital Library, and arXiv using various permutations of the keywords “image steganography,” “deep learning steganography,” “GAN steganography,” “diffusion steganography,” “coverless steganography,” “invertible neural network steganography,” and “steganalysis.”

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

For a paper to be included, the following four criteria had to be met: (i) publication during the years 2021-2025; (ii) a clear emphasis on either image steganography or a related steganalysis topic; (iii) reporting of quantifiable outcomes; and (iv) inclusion in Scopus or its equivalent.

II. BACKGROUND AND TAXONOMY

A. Fundamental Concepts

- Any image-based steganography technique will always work on three entities: the cover image C, the message image M, and the generated stego image S. An embedding function E: (C, M) → S transforms a combination of cover and message images into a stego image that needs to be similar to C both visually and statistically; an extraction function D: (S) → M extracts the secret message from the stego image. Effectiveness of such techniques is typically measured against five criteria:
- Imperceptibility — measured through PSNR (values over 38dBare normally considered good) and SSIM (values over 0.97 are considered outstanding), indicating the visual similarity between the stego image and the original cover.
- Capacity of embedding — number of secret bits embedded per pixel (bpp); large capacity is good but usually at the cost of imperceptibility.
- Security/undetectability — resistance to steganalysis; ideally secure embedding causes the steganalysis's accuracy to be close to 50% (i.e., no better than tossing a coin).
- Robustness — capability to decode the secret message even after distortion to the stego image such as compression (JPEG), resizing, cropping, or noise addition.
- Computational efficiency — computational cost for training, inference latency, etc.

B. Taxonomy of Image Steganography

1. Spatial Domain Approaches: This is based on changing the pixel intensity. The traditional LSB method replaces the least significant bits in pixels with secret data, thereby providing large capacity (maximum 8 bpp) with very little perceptual distortion. Some other forms include LSB matching, LSB matching revisited, and edge adaptive LSB, but since the statistical changes done are not complex, spatial domain methods are vulnerable to steganalysis.

2. Modify Domain Methods: As opposed to pixel manipulation, these kinds of methods use coefficients in the frequency domain for hiding information. JPEG steganography hides data into quantized coefficients of the DCT. Wavelet-based steganography utilises the multi-resolution feature of wavelet sub-bands to select proper positions for embedding the information, while Fourier-domain-based steganography embeds data into the phase or amplitude of the Fourier Transform, respectively. As a class of methods, transform domain-based steganography provides more robustness and less vulnerability to basic forms of steganalysis.

3. DL-Based Approaches: End-to-end learned embedding and extraction pipelines have transformed the whole field of study, and four sub-types should be highlighted.
   - Convolutional networks-based approaches (CNN): convolutional encoder-decoder and U-Net architecture learn to distribute the payload among regions of the cover image; it is claimed in a number of works that U-Net-based approaches have better capacity and better visual performance than simpler CNN-based ones [5].
   - GAN-based approaches: the generator network encodes the message, while the discriminator tries to maximize the imperceptibility, and it is usually combined with a steganalysis adversarial network (SAN); examples such as AGASI [34] and multi-adversarial networks [13] show very high anti-detection performance.
   - INNs approaches: embedding and extracting processes are treated as two phases of an invertible network, which theoretically should allow lossless encoding;
   - Recovery has been proved to extend PSNR beyond 40dB [1], [11], [40]; PRIS [11], in particular, has proven that improvements in robustness do not necessitate radical architectural modifications.
   - Methods utilizing denoising diffusion probabilistic models (DDPMs) or latent diffusion models (LDMs) as diffusion-based schemes, where stego images are generated through the diffusion process; the most successful approaches here include CRoSS [20], DiffStega [22] and GSD [21], which provide controllable and robust training-free coverless steganography with high perceptual quality.
   - Transformer-based methods – vision transformers and attention mechanisms [7] utilize the entire image context while choosing optimal locations for embedding instead of limiting themselves to convolutional receptive fields. Coverless Steganography: In this technique, there is no alteration to the cover image. Rather, an image with inherent characteristics that encode the secret message is chosen or created by the technique.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

- The absence of any pixel alteration makes the traditional steganalysis techniques, which search for such alterations, less effective against this technique. Cover generation by hash-based retrieval [18], GAN [4], and diffusion [22], [32] belong to this class of steganography techniques [16], [30].

III. LITERATURE REVIEW

A. Spatial and Transform Domain Advances

Despite the predominance of deep learning-based methods in recent literature, traditional approaches based on spatial and transform domains are still being improved. Spatial domain methods have been investigated extensively by Zhang et al. [24], who determined good embedding locations and developed approaches for adaptive capacity control. On the other hand, Ahmad et al. [23] used a CNN along with DCT to develop their watermarking approach, which was deployed on a cloud-hosted platform, delivering both security and quality in addition to a 2.3-second execution time. Steganography using JPEG has also been widely explored in the literature. Wang et al. [6] presented a JPEG steganography method in which content similarity analysis was conducted in a cybernetic environment, hence increasing the precision of embedding in the DCT domain. Zeng et al. [15], [19] proposed a number of robust steganography methods that resulted in "upward robust steganography," having an "overflow prevention scheme" with 40.1 dB PSNR.

B. CNN and Encoder-Decoder Architectures

The CNN-based methods have come a long way since. The HiNet approach presented by Lu et al. at CVPR in 2021 [8] illustrated efficient large-capacity image hiding with INNs achieving up to 4 bpp — a milestone in this period. Further expanding the research on this idea, Liu et al. [1] revealed that INN steganography enables lossless extraction of information with PSNR > 40 dB and a detection probability of under 52%, making it superior to LSB and other steganographic methods in this regard. Later on, Fengyong et al. [40] developed LiDiNet — an efficient invertible model with adaptive spatial attention modules that made INN steganography feasible in practice.

Moreover, attention mechanisms became even more beneficial for CNN-based approaches. Ma et al. [13] combined multiple adversarial models with channel-attention modules, thus significantly increasing the robustness against SRNet-based steganalysis. Zhang et al. [29] incorporated style transfer into a conditional image-hiding network

C. GAN-Based Generative Steganography

GAN-based steganography has proven to be one of the best ways of ensuring imperceptibility while having high capacity. The work done by Peng et al. [4] in IEEE Transactions on Circuits and Systems for Video Technology in 2022 has developed a strong coverless model using GAN and gradient-descent approximation, proving the ability of generating images to embed secrets without having any cover at all. Zhu et al. [2] developed Resen-Hi-Net involving a residual network and pixel shuffling along with image encryption, providing 0.98 for SSIM, plus good resistance to steganalysis attacks.

Yang et al. [14] developed a model called ACGIS where an adversarial cover generator is used for preserving noise-residual features for defeating noise-sensitive steganalysis. Layerwise Adversarial Learning was suggested by Chen et al. [12], where adversarial perturbation is added to the network at different levels in order to make the model less detectable by several steganalysis attacks. Chahar et al. [34] developed AGASI, where the GAN model is developed with the encoder as the generator and discriminator enforcing naturalness to the stego image.

D. Invertible Neural Network (INN) Steganography

Steganography using INNs has been a hot research area due to its guarantee for the exactness of recovery. Liu et al. [1] proposed the steganographic approach using INNs where the forward pass hides the message, and the backward pass reveals it, resulting in exact or nearly exact recovery. The researchers in [11] (PRIS) have made INNs practically robust by adding a lightweight mechanism to handle perturbation without affecting the invertible model structure and achieved a PSNR of 41.2 dB.

The modified INN steganography algorithm through super-resolution [24], where steganography is viewed as the upscaling process, applies embedding a smaller image into a bigger one through INN and learns the inverse transformation to retrieve the data. The modified INN steganography technique not only increases the capacity of INN steganography but also solves the problem of dimensionality mismatch in INN steganography. Fan et al. [35] proposed a generator for hiding data into deep network layers.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

E. Diffusion Model-Based Steganography

Among the recent advancements in steganography, diffusion-based steganography can be deemed the most significant one. Starting from the Latent Diffusion Model (LDM) framework put forward by Rombach et al. [9] at CVPR 2022, a number of other research groups managed to design steganographic schemes using the noise-injecting property of diffusion models.

Generative Steganography Diffusion (GSD) technique was presented by Wei et al. [21]; it was the first technique that implemented embedding of secrets into the diffusion generation procedure without altering the existing cover object. Controllable, Robust and Secure Diffusion-Based Framework (CRoSS) was suggested by Yu et al. [20], at NeurIPS 2023. Universal, training-free coverless steganography was performed using DiffStega [22] (Yang et al., IJCAI 2024) with pre-trained diffusion models not requiring expensive fine-tuning. Message-driven MDDM [36] (ICML 2025) extended the direction.

F. Transformer-Based Steganography

Vision transformers are also slowly finding their place in image steganography. Zhou et al. [7] suggested an approach to deep image steganography based on transformer blocks utilizing recursive permutation with performance (in terms of PSNR) of 38.7 dB and strong security features. Thanks to the global attention property of transformers, the network is capable of recognizing the right places for embeddings all over the whole image, unlike CNN architectures that work in a local fashion by design. Luo et al. [31] (StegMamba, IEEE T-CSVT 2025) offered the Mamba state-space model (SSM) for multi-image steganography, demonstrating increased throughput and capacity compared to transformer baselines.

G. Coverless Steganography

Uncovered steganography has evolved from primitive hash-based image retrieval to significantly advanced generative models. The state of uncovered steganography was reviewed by Meng et al. [16] in Neurocomputing (2024). Meng et al. [30] in IEEE T-CSVT (2025) suggested a universal approach using image restoration that aims at improving the robustness of any uncovered steganography algorithm. Li et al. [18] used image retrieval based on SURF features for robust uncovered steganography.

H. Robustness Against Social Media and Lossy Channels

One such challenge is to ensure the robustness of the steganography technique even after the stego images have undergone processing by social media sites like WhatsApp and X/Twitter, where images get compressed and resized in a standardized way through JPEG encoding, resizing, and colour normalisation. Tang et al. [27] suggested joint learning of cost and payload with attention to image-wise batch steganography. Yang et al. [28] IEEE Trans. Multimedia (2024) proposed a provably secure and robust steganography technique using an information-theoretic secure model with error correction.

I. Steganalysis: The Adversary Perspective

Steganalysis awareness is important in the development of a secure and robust steganography system. Modern steganalysis is heavily dependent on deep learning algorithms, where detectors like SRNet, XuNet, and EfficientNet almost perfectly detect classical steganography. In the comparison study carried out by Wani and Sultan [38], they discovered that the EfficientNet architecture with the addition of squeeze-and-excitation layers has the best detection robustness. Song et al. [10] in their review paper in Expert Systems with Applications have pointed out that deep-learning-based steganalysis has improved the security level so much that only distributional or generative approaches may effectively hide data from detection. The framework for automatic steganographic cost learning in GANs by reinforcement learning signals from steganalysis was proposed by Huang et al. [39]

Method / Technique Used	Referenced Paper Numbers
Invertible neural network (INN) steganography	[1], [8], [11], [40]
GAN-based steganography	[2], [4], [5], [13], [14], [34], [39]
Diffusion model-based steganography	[9], [20], [21], [22], [36]
CNN / encoder-decoder architecture	[5], [12], [13], [23], [29]
Transformer / attention-based steganography	[7], [31]
Spatial domain (LSB / pixel modification)	[3], [17], [24]

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15750

Transform domain (DCT / DWT / DFT)	[6], [15], [19], [23]
Coverless steganography (hash / retrieval-based)	[4], [16], [18], [22], [30], [32]
Robust steganography (social media / JPEG)	[15], [19], [27], [28], [30]
Adversarial steganography (anti-steganalysis)	[12], [13], [14], [34], [39]
Layerwise / multi-level adversarial learning	[12], [13]
Lightweight / edge-efficient steganography	[11], [40]
Multi-image steganography	[26], [31]
Provably secure steganography (info-theoretic)	[28]
Bio-inspired optimization for steganography	[25]
Latent-space embedding with smart decoder	[37]
Reinforcement learning (RL) for cost learning	[39]
Generative steganography (cover generation)	[4], [5], [21], [22], [35], [36]
Super-resolution / up-sampling for hiding	[26], [24]
Steganalysis/detection methods (adversary)	[3], [10], [17], [33], [38]
Survey / systematic review	[3], [10], [16], [17], [32], [33]
Mamba state-space model (SSM) for steganography	[31]
Style-transfer integration in hiding	[29]
Message-driven diffusion steganography	[36]
Latent diffusion model (foundation)	[9], [20], [21], [22], [36]

IV. COMPARATIVE ANALYSIS OF METHODS

Reading through the 40 papers together reveals the following four recurring trends. First, deep learning-based techniques tend to outperform classical spatial- and transform-domain approaches on all fronts. Second, GAN and diffusion-based approaches obtain the highest imperceptibility and security results compared to all other types of methods. Third, coverless techniques are distinct in that they obtain the highest possible security and robustness results since no cover is ever altered. Fourth, for those techniques that alter the cover, invertible neural networks yield the best compromise between embedding capacity and imperceptibility.

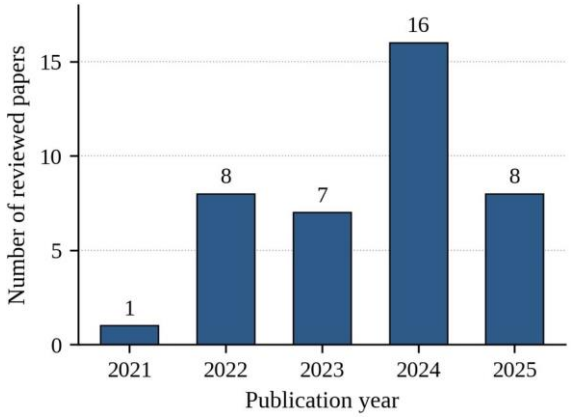


Fig. 1. Distribution of the 40 reviewed papers by publication year (2021– 2025), showing the sharp rise of deep-learning-based and diffusion-model approaches after 2023.

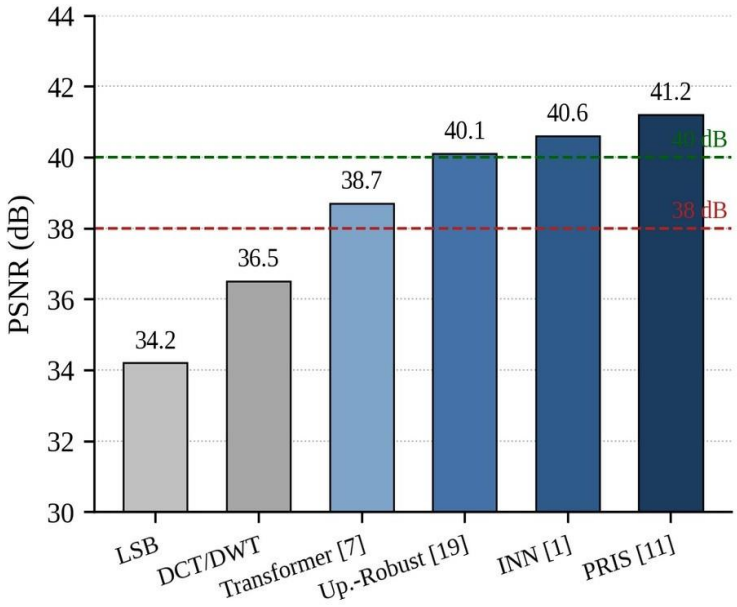
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15750

The accelerated development of the field can be seen from Fig. 1, where about 60% of the reviewed works were published in 2024 or 2025, which reflects the rapid adoption of diffusion and transformer techniques since the introduction of latent diffusion models in 2022 [9].



As is evident from Fig. 2, LSB, the classical technique of embedding in the spatial domain, lags behind other techniques in image quality, whereas the techniques proposed using INN architecture like PRIS [11] and the framework proposed by Liu et al. [1] cross the 40 dB limit of visual losslessness.

The comparison presented in Fig. 3 illustrates the following trade-offs in brief: classical techniques are efficient but behind regarding security and invisibility; INN approaches are top in terms of capacity while near the top in terms of invisibility; and diffusion-based and coverless schemes excel in security and robustness while requiring much more computational resources.

V. DATASETS AND EVALUATION METRICS

A. Benchmark Datasets

- In terms of benchmarking datasets that have been referenced throughout the reviewed literature, only a few recurrent collections are being mentioned:
- ImageNet/ILSVRC – a collection of approximately 1.2 million images from 1,000 categories used to train deep steganography models.
- MS-COCO – a collection of approximately 330,000 annotated images used not only to train but also to evaluate steganographic capacity and quality.
- CelebA/FFHQ – face image collections containing around 202,000 and 70,000 images, respectively, used to evaluate generative steganography.
- BOSSBase – 10,000 grayscale images at the resolution of 512x512, which are the long-established benchmarking datasets for spatial domain steganography and steganalysis research.
- BOWS2 – a complement to BOSSBase used for evaluation of steganalysis performance.
- PASCAL VOC 2012 – around 17,000 images with annotations; it was used for the assessment of some INN-based steganography approaches [1].
- DIV2K – high-resolution dataset (800 training and 100 validation images), favoured by researchers.

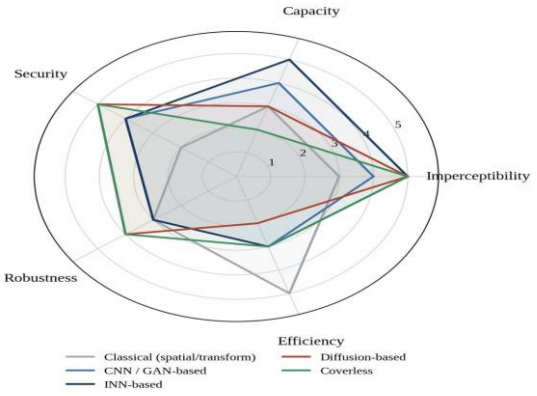
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026

DOI: 10.17148/IJARCCE.2026.15750

B. Performance Metrics



- In the analyzed papers, there are seven metrics used most often:
- PSNR (Peak Signal-to-Noise Ratio) – it measures pixel-level distortion between cover and stego images; a value higher than 38 dB is considered high, and INN-based methods often have more than 40 dB.
- SSIM (Structural Similarity Index) – perceptual metric, which takes into account luminance, contrast, and structure similarity; the value higher than 0.97 is considered excellent.
- Embedding capacity (bpp) – number of secret bits embedded into one pixel; classical methods usually reach 1-3 bpp, and deep learning ones – 3-4 bpp+.
- Detection accuracy – this metric evaluates the ability of the steganalysis (for instance, SRNet) to distinguish cover and stego images; the closer the values to 50%, the better the security is provided.
- Secret Recovery Accuracy (SRA) – the percentage of secret bits successfully recovered; practically, the value must be higher than 99%.
- Fréchet Inception Distance (FID) – metric used in generative and diffusion-based steganography to measure the distance between generated and real-world images.

VI. EMERGING TRENDS AND NEW DIRECTIONS

A. Diffusion Model Integration

However, the implementation of diffusion models for steganography is definitely one of the greatest breakthroughs in the domain recently. Compared to GANs, diffusion models have a more stable training process and greater sample diversity, which in turn leads to more realistic images with better security features. Moreover, the increasing prevalence of pre-trained diffusion models (Stable Diffusion, DALL-E, Imagen) allows for building steganographic methods without training at all [22], [36].

It can be observed that the increase in capacity for INN-based techniques was particularly pronounced once HiNet [8] and similar architectures were introduced, even though diffusion and coverless steganography techniques, despite having lower capacity to start with, are gradually catching up with the development of their backbone networks.

B. Multimodal and Cross-Modal Steganography

There is an ever-growing number of studies related to cross-modal steganography, where messages in one modality (such as text/audio) are embedded into the image. Vision and language multimodal models, like CLIP and BLIP-2, are widely used nowadays for the generation of steganographic images, which are semantic and hence hard to detect by perceptual methods. This is a relatively new area but still offers many opportunities for making meaningful contributions.

C. Secure Steganography for 3D and Medical Images

In addition to 2D images, there are also applications of steganography to other objects. Medical image steganography, aimed at securing patient medical data, mainly applies watermarking techniques, such as DFT and DWT, that allow lossless embedding. Steganography for 3D Gaussian splatting (such as SecureGS) is also a promising area of investigation.

D. Bio-Inspired and Quantum Resistant Approaches

Bio-inspired optimization techniques, such as genetic algorithms, particle swarm and ant colony optimization have been applied in order to create the steganographic key and determine optimal embedding regions. As Rezaei and Javadpour

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

[25] demonstrated, bio-inspired algorithms help in improving both security and the quality of stego images transmitted. The field of quantum-resilient steganographic key management represents another promising research area, motivated by the quantum computing threats to traditional cryptographic schemes protecting steganographic keys

E. Steganography in Social Network Environments

However, social media sites use powerful image processing techniques – re-compressing JPEG, resizing, colour normalisation – that can totally destroy hidden information. As a result, social-media-aware steganography has turned out to be a field on its own. Indeed, algorithms designed specifically to deal with these kinds of distortions [27], [28] proved their robustness.

F. Adversarial Robustness and Anti-Forensics

The intersection between adversarial machine learning and steganography has resulted in methods that can fool steganalysis networks in order to misguide them. In these techniques, perturbations are applied to the cover image in such a way that the error rate of the steganalysis increases, without any degradation in perceptual quality. Some examples of this technique include ACGIS [14] and AGASI [34].

G. Large Language Model (LLM) Assisted Steganography

Language models have been used recently for text steganography, while the same technology can be employed for image steganography indirectly via text-to-image generation pipelines. Merging the two implies that a new direction can be explored where the steganographic message is embedded into the semantic prompt space of the text-to-image model.

VII. OPEN CHALLENGES AND RESEARCH GAPS

A. Capacity–Security Tradeoff

Notwithstanding the many advancements that have been made, the basic conflict between capacity and security has not yet been resolved. Any technique beyond 4 bpp will result in statistical anomalies, which will be detectable via sophisticated steganalysis tools, while the most secure techniques are limited to 1-2 bpp. There is a need for theoretical analysis based on the rate-distortion model.

B. Robustness Under Unknown Distortions

The steganography techniques that have already been proposed for robustness in practice have been trained for a specific type of distortion, such as JPEG, Gaussian noise, and resizing. In actual practice, the image will undergo unknown and compound distortions, and current approaches cannot handle them. Zero-shot robustness and domain adaptation from transfer learning could solve this problem.

C. Standardised Benchmarking

A lack of widely accepted benchmarks such as those used in classification research with ImageNet is also evident in this research domain. Differing sets of datasets and measures have been utilized in differing research papers, making cross-comparison challenging. There is a need for community-accepted benchmarks in this regard.

D. Real-Time Performance

GAN-based and diffusion models require heavy computations, which make them difficult to deploy in real time. However, lightweight models like LiDiNet [40] and PRIS [11] solve some aspects of the above challenge; however, model compression, knowledge distillation, and neural architecture search are relatively under-exploited in steganography.

E. Formal Security Proofs

Most deep learning-based steganography techniques rely on empirical evaluation and not proofs of security. Very few papers provide any information-theoretic security guarantees. One of the earlier works in this space that proves security is by Yang et al. [28]; however, a theory for computational security of deep steganography is still in its nascent stages.

F. Multi-Party and Authenticated Steganography

All the present approaches operate on the assumption that there is only one source and only one destination. Cases where there are multiple parties involved, either by having multiple receivers of a stego-image or by having multiple sources in a cover image, have been relatively untouched.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

VIII. FUTURE RESEARCH DIRECTIONS

- In light of the above-identified trends and gaps, the following directions seem particularly well-suited for impactful contributions indexed in Scopus:
- Unified Diffusion Steganography – a universal framework that provides for controlled secret concealment, security, and robustness of the steganographic method using only one architecture by taking advantage of the mathematical nature of the diffusion process.
- Federated Steganography – steganographic model training using a federated learning approach and, thus, not requiring sharing of any raw images, which is vital in sensitive areas like medicine and finance.
- Quantum-resistant Steganographic Keys – management of the steganographic keys based on post-quantum cryptography primitives that are resilient to future attacks by quantum computers.
- Neuromorphic Steganography – embedding approaches that target neuromorphic hardware, e.g., Intel Loihi, thus allowing for ultra-low-power covert communication in Internet-of-Things devices.
- Steganalysis-Resistant Universal Architecture – meta-learning framework that keeps adapting the embedding approach to avoid steganalysis attacks using new methods without needing to be retrained.
- 3D and Video Steganography Using Diffusion Models – extension of the latent diffusion steganography to 3D point clouds, meshes, and video.
    - Steganography for AI-generated image watermarking—imperceptible watermarking and steganographic techniques for AI-generated images (deepfakes and synthetic images) to enable authentication of origin and content.

IX. CONCLUSION

The present review explores the state-of-the-art in image steganography, which is an emerging technique within the area of data hiding, based on 40 papers from prominent conferences and journals IEEE, ACM, Springer, and Elsevier, covering the period from 2021 to 2025. In doing so, the review has tracked the development of the discipline from traditional methods up to deep learning-based approaches.

Four key points can be mentioned. First, deep learning has revolutionised image steganography in that there are now several methods that simultaneously produce a high PSNR of more than 40 dB and embedding capacity of more than 4 bpp. Second, diffusion-based steganography such as CRoSS [20], DiffStega [22], and MDDM [36] is probably the most important recent development, which provides training-free, highly robust, and even provenly secure embedding. Third, coverless steganography has evolved from simple retrieval-based approaches to advanced generative models that are highly resistant to pixel-level attacks. Finally, the competition between steganography and steganalysis keeps going, and efficient net-based detectors [38] are now the latest achievement in this field.

Some open questions that need attention include the formulation of a theoretical explanation of the capacity and security trade-off, resilience to distortions that are unknown or mixed, benchmarking, real-time performance, proof of security, and the ability to handle multiple parties. Being an interesting blend of information security, computer vision, and machine learning, image steganography continues to offer many avenues for research. This review serves as a state-of-the-art basis for future development of steganographic systems.

REFERENCES

[1] L. Liu, L. Tang, and W. Zheng, “Lossless image steganography based on invertible neural networks,” Entropy, vol. 24, no. 12, p. 1762, 2022.
[2] X. Zhu, Z. Lai, N. Zhou, and J. Wu, “Steganography with high reconstruction robustness: Hiding of encrypted secret images,” Signal Processing, vol. 201, p. 108722, 2022.
[3] P. C. Mandal, I. Mukherjee, G. Paul, and B. N. Chatterji, “Digital image steganography: A literature survey,” Information Sciences, Elsevier, vol. 609, pp. 1451–1488, 2022.
[4] 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, 2022.
[5] P. Wei, S. Li, X. Zhang, G. Luo, Z. Qian, and Q. Zhou, “Generative steganography network,” in Proc. 30th ACM Int. Conf. Multimedia (MM), Lisbon, Portugal, 2022, pp. 1621–1629.
[6] Z. Wang et al., “JPEG steganography with content similarity evaluation,” IEEE Trans. Cybernetics, vol. 53, no. 8, pp. 5082–5093, 2022.
[7] Z. Zhou, M. Wang, B. Liu, and T. Li, “Deep image steganography using transformer and recursive permutation,” Entropy, vol. 24, no. 7, p. 878, 2022.

DOI: 10.17148/IJARCCE.2026.15750

[8] J. Jing, X. Deng, M. Xu, J. Wang, and Z. Guan, "HiNet: Deep image hiding by invertible network," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 4713–4722.
[9] 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.
[10] B. Song et al., "A survey on deep-learning-based image steganography," Expert Systems with Applications, vol. 254, p. 124390, 2024.
[11] H. Yang et al., "PRIS: Practical robust invertible network for image steganography," Engineering Applications of Artificial Intelligence, vol. 133, p. 108419, 2024.
[12] B. Chen, L. Shi, Z. Cao, and S. Niu, "Layerwise adversarial learning for image steganography," Electronics, vol. 12, no. 9, p. 2080, 2023.
[13] B. Ma et al., "Enhancing the security of image steganography via multiple adversarial networks and channel attention modules," Digital Signal Processing, vol. 141, p. 104121, 2023.
[14] J. Yang and X. Liao, "ACGIS: Adversarial cover generator for image steganography with noise residuals features-preserving," Signal Processing: Image Communication, vol. 113, p. 116927, 2023.
[15] K. Zeng et al., "Robust steganography for high quality images," IEEE Trans. Circuits Syst. Video Technol., vol. 33, pp. 4893–4906, 2023.
[16] L. Meng, X. Jiang, and T. Sun, "A review of coverless steganography," Neurocomputing, vol. 566, p. 126945, 2024.
[17] D. R. I. M. Setiadi et al., "Digital image steganography survey and investigation," Signal Processing, vol. 206, p. 108908, 2023.
[18] F. Li et al., "A robust coverless image steganography algorithm based on image retrieval with SURF features," Security and Communication Networks, vol. 2024, Art. no. 5527837, 2024.
[19] K. Zeng et al., "Upward robust steganography based on overflow alleviation," IEEE Trans. Multimedia, vol. 26, pp. 299–312, 2024.
[20] 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.
[21] P. Wei, Q. Zhou, Z. Wang, Z. Qian, X. Zhang, and S. Li, "Generative steganography diffusion," arXiv preprint arXiv:2305.03472, 2023.
[22] Y. Yang et al., "DiffStega: Towards universal training-free coverless image steganography with diffusion models," in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), 2024.
[23] Ahmad et al., "Hybrid CNN-DCT approach for image steganography in cloud-based environments," Journal of Cloud Computing: Advances, Systems and Applications, vol. 13, 2024.
[24] C. Zhang, "Research on key technologies of spatial domain image steganography," Transactions on Computer Science and Intelligent Systems Research, vol. 6, 2024.
[25] S. Rezaei and A. Javadpour, "Bio-inspired algorithms for secure image steganography," Multimedia Tools and Applications, Springer, 2024.
[26] P. Wei, Q. Zhou, Z. Wang, Z. Qian, X. Zhang, and S. Li, "Generative steganography diffusion," arXiv preprint arXiv:2305.03472, 2023.
[27] Y. Yang et al., "DiffStega: Towards universal training-free coverless image steganography with diffusion models," in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), 2024.
[28] Ahmad et al., "Hybrid CNN-DCT approach for image steganography in cloud-based environments," Journal of Cloud Computing: Advances, Systems and Applications, vol. 13, 2024.
[29] C. Zhang, "Research on key technologies of spatial domain image steganography," Transactions on Computer Science and Intelligent Systems Research, vol. 6, 2024.
[30] S. Rezaei and A. Javadpour, "Bio-inspired algorithms for secure image steganography," Multimedia Tools and Applications, Springer, 2024.
[31] P. Ping et al., "Hiding multiple images into a single image using up-sampling," IEEE Trans. Multimedia, vol. 26, pp. 4401–4415, 2024.
[32] W. Tang et al., "Joint cost learning and payload allocation with image-wise attention for batch steganography," IEEE Trans. Information Forensics and Security, vol. 19, pp. 2826–2839, 2024.
[33] Z. Yang et al., "Provably secure robust image steganography," IEEE Trans. Multimedia, vol. 26, pp. 5040–5053, 2024.
[34] F. Zhang et al., "Conditional image hiding network based on style transfer," Information Sciences, vol. 662, p. 120225, 2024.
[35] L. Meng et al., "A universal framework for improving the robustness of coverless image steganography," IEEE Trans. Circuits Syst. Video Technol., vol. 35, no. 1, pp. 922–937, 2025.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 7, July 2026
DOI: 10.17148/IJARCCE.2026.15750

[36] T. Luo et al., "StegMamba: Distortion-free immune-cover for multi-image steganography with state space model," IEEE Trans. Circuits Syst. Video Technol., vol. 35, no. 5, pp. 4576–4591, 2025.
[37] A. Z. Abadin, R. Sulaiman, and M. K. Hasan, "Review: Embedding strategies emphasizing robustness, security, payload capacity, and visual quality," Computers, Materials & Continua, vol. 79, 2024.
[38] O. Alrusaini, "Current methods for hiding information in images alongside steganalysis methods," Frontiers in Artificial Intelligence, vol. 8, 2025.
[39] N. K. Chahar et al., "AGASI: GAN-based approach for adversarial image steganography," in Proc. Int. Conf. Electronics, AI & Technology (ICEAIT), 2025.
[40] H. Fan, C. Jin, and M. Li, "Generator model for image steganography boosting capacity without complex architectures," Entropy, vol. 27, 2025.
[41] Z. Xu et al., "MDDM: Practical message-driven generative image steganography based on diffusion models," in Proc. Int. Conf. Machine Learning (ICML), 2025.
[42] F. Li et al., "Deep learning-based image steganography with latent space embedding and smart decoder selection," Applied Sciences (MDPI), Dec. 2025.
[43] M. A. Wani and S. Sultan, "Deep learning for steganalysis: Evaluating model robustness against image transformations," Frontiers in Artificial Intelligence, vol. 8, 2025.
[44] Z. Huang et al., "GAN and reinforcement learning automatic cost learning framework for steganography," Pattern Recognition, Elsevier, vol. 150, p. 110327, 2024.
[45] L. Fengyong et al., "LiDiNet: Lightweight deep invertible steganographic network with adaptive spatial attention," Engineering Applications of Artificial Intelligence, vol. 133, 2024.

