Deep Learning Techniques for Image Steganography: A Comprehensive Review

Jannies Varghese
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kottayam, India
janniesvarghese2026@cs.ajce.in

Hariprasad Prasanth
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kottayam, India
hariprasadprasanth2026@cs.ajce.in

Blessy Mariam Babu
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kottayam, India
blessymariambabu2026@cs.ajce.in

Chris Joseph
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kottayam, India
chrisjoseph2026@cs.ajce.in

Ms. Bini M Issac
Assistant Professor
Dept. of Computer Science & Engineering
Amal Jyothi College of Engineering
Kanjirappally, Kottayam, India
binimissac@amaljyothi.ac.in

***Abstract*—Image steganography is an important aspect of secure communication that hides confidential messages within digital images in a way that escapes detection. Conventional steganographic algorithms, including Least Significant Bit (LSB) and frequency domain-based approaches, have been found to have low embedding capacity, susceptibility to steganalysis attacks, and rigidity in terms of rule-based embedding processes. However, with the evolution of deep learning concepts, especially in the realms of convolutional neural networks (CNNs) and generative networks, image steganography has moved towards adaptive and data-driven models that have improved imperceptibility and robustness to a great extent. This review paper provides a detailed examination of the existing state-of-the-art deep learning-based models for image steganography, including CNN-based encoder-decoder models, GAN-based adaptive cost learning models, hybrid CNN-frequency domain models, and multi-layered steganographic models. The reviewed papers are critically compared in terms of embedding capacity, visual distortion, robustness to steganalysis attacks, computational complexity, and applicability. Based on this comparative analysis, the important research gaps in the existing models are identified. The review aims to act as a reference for researchers and students who would like to gain insight into the current developments in deep learning-based image steganography.In addition, the architectural trade-offs highlighted in this review provide practical guidance for selecting suitable steganographic frameworks under different application constraints, including capacity, security, and deployment efficiency.**

***Index Terms*—Image Steganography, Deep Learning, Convolutional Neural Networks (CNN), Encoder–Decoder Architecture, GAN-Based Steganography, Data Hiding, Information Security, PSNR, SSIM**

I. Introduction

As the growth of digital communication and multimedia data transfer accelerates, the confidentiality and integrity of sensitive information have become a pressing issue. Although cryptographic methods aim to protect the content of information, they are not capable of hiding the fact that communication exists. Image steganography is a method that overcomes this problem by hiding secret information in digital images in a way that is imperceptible to human observation.

In traditional image steganography, methods like Least Significant Bit (LSB) substitution and frequency domain approaches involving transforms like Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) follow predefined rules for embedding data. Although these approaches are easy to implement and require less computation, they are often associated with low embedding capacity, susceptibility to advanced steganalysis attacks, and a lack of flexibility to accommodate varying characteristics of images.As steganalysis methods have become increasingly sophisticated, particularly with the use of machine learning, the limitations of conventional rule-based steganography have become more apparent.

In recent years, deep learning has become a strong method for solving complex image processing problems because it can automatically learn features from data. This progress has greatly impacted the field of image steganography. It has

led to learning-based approaches that replace manual embedding rules with flexible neural network models. Convolutional Neural Networks (CNNs), encoder-decoder architectures, and Generative Adversarial Networks (GANs) have been widely studied to improve imperceptibility, embedding capacity, and resistance to steganalysis.

Several deep learning based steganography techniques have been proposed in the literature. These include CNN based encoder decoder frameworks for end-to-end learning, GAN based adaptive cost learning models that use adversarial training, hybrid approaches that combine deep learning with frequency domain transforms, and multi-layered frameworks that blend compression with learning based embedding. While these methods show significant improvements over traditional approaches, they also bring new challenges. These challenges include increased computational complexity, training instability, and limited practicality for real-world use.

Given the rapid and varied advancements in deep learning-based image steganography, a systematic review and comparison of current techniques is necessary to understand trends, evaluate strengths and weaknesses, and identify gaps in research. This review examines and compares key deep learning-based image steganography methods based on important performance factors such as visual quality, embedding capacity, resistance to steganalysis, and computational efficiency. Additionally, this paper discusses a simplified encoder-decoder method that balances security and practicality, offering insights into future research directions in secure image data hiding. Despite the growing number of deep learning-based steganography techniques, there is still a lack of clear comparative analysis that highlights the architectural trade-offs among different approaches. Many studies focus on improving specific performance metrics such as embedding capacity or imperceptibility, but fewer works systematically examine how these design choices affect robustness, security against steganalysis, and computational efficiency. Therefore, a comprehensive review and comparison of recent architectures is necessary to better understand their strengths, limitations, and suitability for different application scenarios. The main contributions of this review paper are summarized as follows:

• A comprehensive review of recent deep learning–based image steganography techniques, including GAN-based frameworks, hybrid CNN–DCT approaches, multi-layered architectures, and adaptive cost learning models.

• A comparative analysis of these architectures based on key performance metrics such as embedding capacity, imperceptibility (PSNR and SSIM), robustness against steganalysis, and computational complexity.

• Identification of major research challenges and open problems in modern deep learning-based steganography systems.

• Practical recommendations for selecting suitable steganographic architectures based on application requirements such as capacity, security, and deployment constraints.

II. Background And Preliminaries

A. Classical Image Steganography

Classical image steganography hides secret data inside images using fixed rules. Common spatial-domain techniques include Least Significant Bit (LSB) substitution and Pixel Value Differencing (PVD). LSB is easy to implement and works well in simple scenarios, but it is highly vulnerable to statistical steganalysis due to predictable pixel modifications. PVD improves embedding by using pixel intensity differences, allowing more data to be hidden in edge regions while maintaining better visual quality.

Transform-domain methods, such as Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), embed information into frequency components of images. These techniques generally provide better robustness against compression and noise but increase computational complexity. Overall, classical methods are effective in controlled environments but lack adaptability to modern detection techniques.

B. Deep Learning–Based Steganography

Deep learning has introduced a more flexible and adaptive approach to image steganography. Most deep learning–based methods fall into two categories: encoder–decoder architectures and adversarial (GAN-based) frameworks. Encoder–decoder models learn an end-to-end mapping that embeds secret data into cover images and later recovers it with minimal visual distortion. These methods can achieve high embedding capacity when trained on paired image data.

GAN-based approaches add a discriminator or steganalyzer during training, forcing the generator to produce stego images that closely resemble natural images. This adversarial process improves imperceptibility and resistance to steganalysis but often comes with higher training complexity and stability issues. Performance is typically evaluated using metrics such as bits per pixel (bpp), PSNR, SSIM, and robustness under common image manipulations.

III. Literature Review

Recent advancements in image steganography have been greatly influenced by deep learning techniques, which aim to address the shortcomings of traditional rule-based embedding methods. Many studies have looked into convolutional neural networks (CNNs), generative adversarial networks (GANs), and hybrid frameworks to enhance embedding capacity, imperceptibility, and resistance to steganalysis.

Ahmad et al. [1] proposed a hybrid CNN and DCT steganography framework designed for cloud-based environments. In their method, CNNs identify suitable embedding regions within cover images, while the Discrete Cosine Transform (DCT) embeds secret data in the frequency domain. This combination improves visual quality and robustness against common image processing tasks, such as compression. Experimental results showed higher embedding capacity, better PSNR and SSIM values, and strong resistance to steganalysis. However, using frequency-domain transformations increases

computational complexity and limits adaptability when compared to fully learning-based spatial methods.

Sanjalawe et al. [2] introduced a deep learning-driven multi-layered steganographic framework that combines Huffman coding, Least Significant B it (LSB) e mbedding, a nd a deep learning-based encoder-decoder network. Huffman coding is used for both payload compression and statistical obfuscation, while the deep learning model improves imperceptibility and robustness. Their results showed exceptionally high SSIM values (above 99%) and strong resistance to noise and compression attacks. However, the multi-layered design adds complexity to the system and may reduce scalability for lightweight or real-time applications.

Wang et al. [3] proposed a GAN-based adaptive cost learning approach to strengthen steganographic security. Their method focuses on learning embedding probability maps using a dual-stream U-Net architecture, enhanced with an attention mechanism. By concentrating on texture-rich and edge regions, the model reduces embedding distortion and boosts resistance against modern steganalysis attacks. Experimental evaluations confirmed b etter s ecurity p erformance c ompared t o existing GAN-based methods. Still, GAN-based training brings challenges including training instability, high computational demands, and longer convergence times.

Ramandi et al. [4] presented VidaGAN, an adaptive GAN-based image steganography framework that includes an encoder, decoder, and critic network. VidaGAN achieves a high hiding capacity of up to 3.9 bits per pixel while maintaining acceptable visual quality. The authors introduced adaptive loss functions to balance capacity and imperceptibility and assessed robustness with steganalysis tools like StegExpose. Although VidaGAN offers state-of-the-art capacity, the trade-off between transparency and robustness continues to be a challenge, and the model requires careful tuning and large datasets for stable training.

From the reviewed literature, it is clear that deep learning-based steganography methods significantly o utperform traditional techniques in terms of adaptability and security. However, GAN-based approaches often face high computational costs and training complexity, while hybrid and multi-layered systems add architectural overhead. These observations underscore the need for simpler encoder-decoder-based models that balance imperceptibility, recovery accuracy, and practical deployment, which motivates the approach discussed in this work. In addition to the individual strengths of these approaches, a comparative view of the reviewed studies reveals that different architectural designs prioritize different objectives in image steganography. GAN-based frameworks such as VidaGAN emphasize maximizing embedding capacity while maintaining acceptable visual quality through adversarial training. Adaptive cost learning approaches focus on improving resistance to steganalysis by learning embedding probability maps that concentrate modifications i n texture-rich regions. Hybrid CNN–DCT architectures provide robustness against compression and cloud-related distortions, whereas multi-layered frameworks combine compression and deep learning to achieve high visual fidelity and reliable message recovery. These differences highlight the importance of selecting steganographic architectures based on specific application requirements such as payload capacity, security against detection, computational efficiency, and deployment environment.

IV. Methodology of this Review

A. VidaGAN, Adaptive GAN for Image Steganography [4]

1) Main Idea and Contributions: VidaGAN introduces an adaptive generative adversarial framework for image steganography made up of three main parts: an encoder, a decoder, and a critic network. The main goal of this framework is to find a good balance between embedding capacity and visual quality while being strong against steganalysis.

The authors present a modified CSPNet-inspired backbone for the encoder and decoder. This design cuts down on computational costs while still allowing for dense feature reuse. To avoid overfitting and make training more stable, they add a soft-labeling loss into the adversarial learning process. They also use a mean squared error (MSE) targeting strategy to manage the differences between the cover and stego images. This approach provides a practical balance between capacity and imperceptibility. To ensure reliable message recovery despite any errors from the decoder, they apply Reed–Solomon error-correcting codes as an external reliability layer.

The main contributions of VidaGAN include: (a) a new CSPNet-based encoder-decoder structure; (b) the use of soft-label loss to improve generalization; (c) an MSE-targeted training method to manage capacity and transparency trade-offs; and (d) thorough robustness tests under JPEG compression, additive noise, and cropping attacks. The experimental results show a peak hiding capacity of 3.9 bits-per-pixel (bpp) on the DIV2K dataset and an area-under-curve (AUC) value of 0.60 using the StegExpose steganalysis tool, demonstrating a strong balance between embedding capacity and detectability.

2) Architecture and Training Strategy: In the VidaGAN framework, the encoder converts a cover image C and a binary secret message M into a stego image S. The decoder then reconstructs the hidden message M' from the stego image. A critic network, which acts as both a discriminator and steganalyzer, ensures that S and C have similar distributions through adversarial training.

The CSPNet-based architectural changes reduce unnecessary computation while keeping effective information flow through dense connections. Training uses a compound loss function that combines reconstruction accuracy, decoding loss, MSE between cover and stego images, adversarial loss, and a soft-label regularization term. For reliable end-to-end performance, Reed–Solomon coding is used as an outer error-correction layer, which enhances robustness but slightly lowers effective payload capacity.

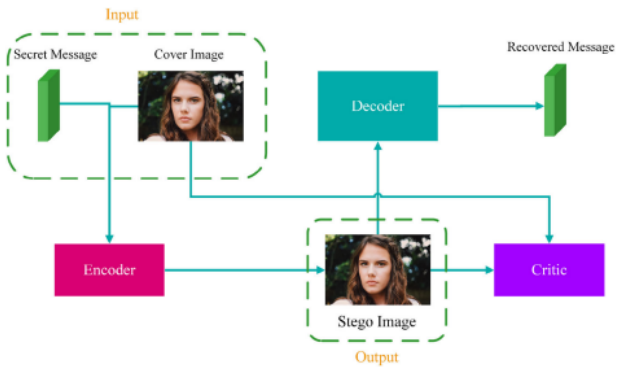

Fig. 1: Architecture of the VidaGAN framework showing the encoder, decoder, and critic.

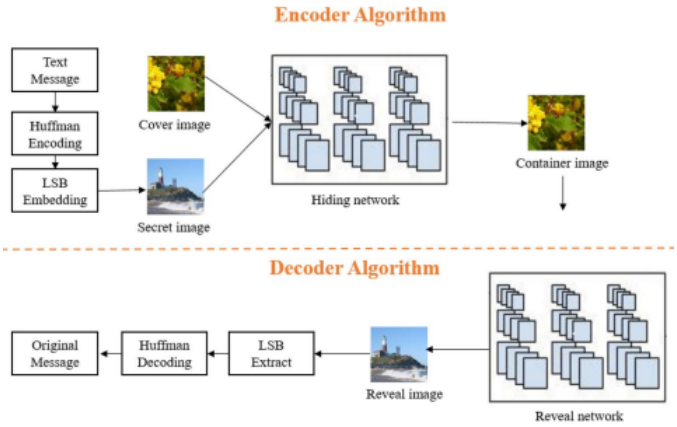

Fig. 2: Framework of the proposed steganographic method.

B. Deep Learning–Driven Multi-Layered Steganographic Approach [2]

1) Core Idea and Contributions: Sanjalawe et al. propose a hybrid, multi-layered steganographic framework that combines lossless compression, deep learning–based adaptive embedding, and lightweight spatial-domain techniques to enhance security and robustness. The core idea is to exploit the complementary strengths of different stages: Huffman coding for payload compression and statistical obfuscation, a deep learning encoder–decoder for adaptive image-in-image hiding, and Least Significant Bit (LSB) embedding for computational efficiency when appropriate.

The primary contributions of this work include: (a) the integration of Huffman coding as both a compression and obfuscation mechanism; (b) a deep learning–based encoder–decoder network that adaptively embeds image payloads while preserving visual quality; (c) a flexible design that supports lightweight LSB embedding for small payloads; and (d) extensive experimental validation demonstrating high imperceptibility and robustness against noise and compression attacks. The authors report consistently high SSIM values (above 99%) and near-perfect recovery accuracy under standard conditions, highlighting the effectiveness of the multi-layered design.

2) Method and Processing Pipeline: The proposed framework operates through a three-stage pipeline designed to balance imperceptibility, robustness, and computational efficiency.

- **Huffman Coding Stage:** Secret payloads are first analyzed for compressibility and, when beneficial, encoded using Huffman coding. This stage reduces the payload size and introduces additional statistical randomness, making the hidden data less predictable to steganalysis techniques.
- **Deep Encoder–Decoder Stage:** A CNN-based hiding network adaptively embeds the compressed payload into a cover image. The encoder learns to place information in statistically safe, texture-rich regions, while the decoder reconstructs the payload from the stego image. The network is trained end-to-end using reconstruction and perceptual loss functions to ensure high visual fidelity and accurate recovery.
- **Optional LSB Embedding Stage:** For small payloads, the framework optionally employs LSB insertion to reduce computational overhead. This fallback mechanism enables fast embedding while maintaining imperceptibility in scenarios where deep learning–based embedding is unnecessary.

The multi-layered architecture enables flexible adaptation to different payload sizes and security requirements, at the cost of increased system complexity due to the sequential processing stages.

C. GAN-Based Adaptive Cost Learning with Attention U-Net [3]

1) Main Idea and Contributions: Wang et al. introduce a GAN-based framework for embedding cost learning that aims to create precise pixel-wise embedding probability maps for secure image steganography. The main contribution is a new generator built on a dual-stream U-Net architecture that includes a convolutional spatial attention (CSA) module. One stream processes the original cover image, while the second stream uses an enhanced version of the image generated through Laplacian filtering to highlight edges and contours.

This architecture helps the network focus embedding probability in texture-rich and edge areas, which offer more safety for data hiding. Important contributions include the attention-augmented dual-stream U-Net, the edge-enhanced auxiliary input, unique skip connections to preserve structural information, and experimental results showing better security against modern steganalysis compared to ASDL-GAN and UT-GAN.

2) Architecture and Training Strategy: The generator takes two inputs: the original cover image and an edge-enhanced image. Convolutional spatial attention modules are used along the contracting path of the original-image stream to emphasize important regions. When upsampling, specific skip-connection strategies are applied to combine information from both streams. The learned embedding probability map is turned

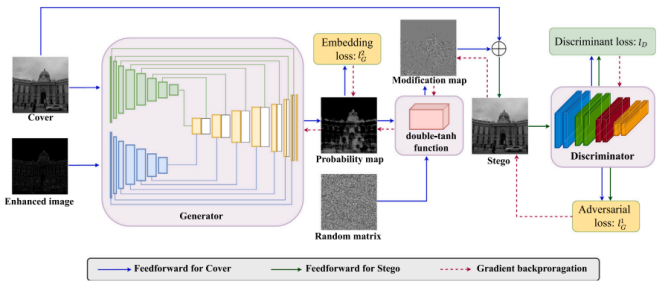

Fig. 3: Architecture of the GAN-based adaptive cost learning framework with dual-stream U-Net and CSA modules.

into pixel modification maps using a double-tanh embedding simulator inspired by UT-GAN.

Training occurs in an adversarial setting, where a CNN-based discriminator serves as a steganalyzer. The overall loss function merges adversarial loss, reconstruction consistency, and regularization terms. This setup allows the generator to learn embedding cost maps that strengthen resistance to steganalysis.

D. Enhanced CNN-DCT Steganography for Cloud Environments [1]

1) Main Idea and Contributions: Ahmad et al. introduce a hybrid CNN-DCT steganography framework for secure image storage and transmission in cloud environments. This method combines the feature-learning ability of convolutional neural networks with the strength of transform-domain embedding. A CNN analyzes the cover image and finds suitable areas for data embedding, while the Discrete Cosine Transform (DCT) embeds secret information into selected frequency coefficients.

The main contributions of this work include merging CNN-based region selection with DCT-based frequency-domain embedding, improving imperceptibility during common image processing tasks, and making it suitable for cloud applications where images often undergo compression and transmission distortions.

2) Method and Embedding Process: In this framework, the CNN is trained to extract image features and predict blocks suitable for embedding. For each selected block, a DCT is applied, and secret bits are added to mid-frequency coefficients based on a set embedding strategy. The inverse DCT is then performed to create the stego image.

This hybrid approach combines CNN-driven region selection with frequency-domain embedding, leading to better visual quality and increased resistance to compression. This makes it effective for steganography in cloud settings.

V. Performance Evaluation and Comparative Analysis

This section compares the reported performance of representative deep learning–based image steganography architectures reviewed in this paper. Since the evaluated methods differ in datasets, payload settings, and evaluation protocols, the results are presented as reported highlights rather than direct experimental comparisons.

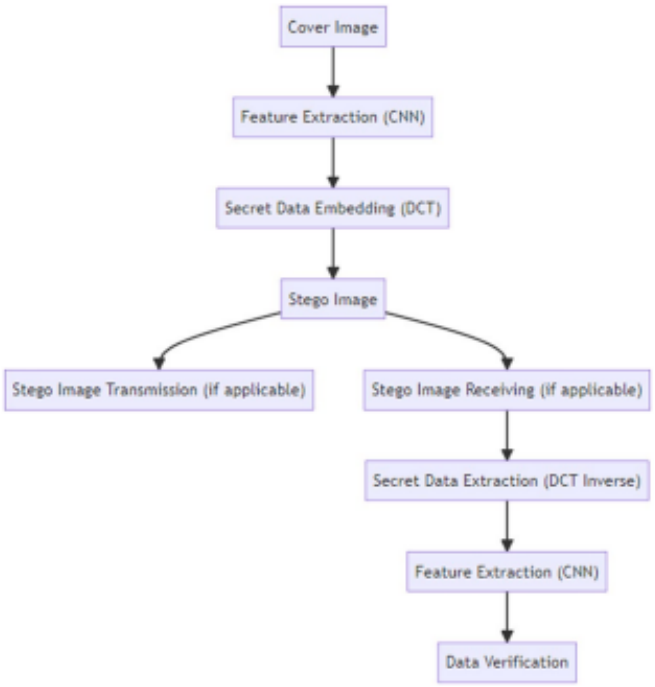

Fig. 4: Block diagram of the proposed CNN–DCT steganographic approach.

A. Adaptive GAN-Based Encoder–Decoder Architecture (VidaGAN)

The Adaptive GAN-based Encoder–Decoder architecture emphasizes a controllable trade-off between embedding capacity and detectability through adversarial learning and explicit MSE targeting [4].

TABLE I: Reported Performance of Adaptive GAN-Based Encoder–Decoder Architecture [4]

Metric	Reported Value
Embedding Capacity (bpp)	3.9
PSNR (dB)	∼35
SSIM	∼0.95
Steganalysis AUC (StegExpose)	0.60
Robustness (JPEG/Noise/Crop)	Evaluated

This architecture demonstrates strong capacity–imperceptibility balance, though it incurs high computational cost due to adversarial training.

B. Multi-Layered Encoder–Decoder with Compression and Lightweight Embedding

The Multi-Layered Encoder–Decoder architecture integrates Huffman coding, deep learning–based hiding, and optional LSB embedding to enhance imperceptibility and robustness through layered security [2].

TABLE II: Reported Performance of Multi-Layered Encoder–Decoder Architecture [2]

Metric	Reported Value
PSNR (dB)	>38
SSIM	>0.99
Bit Error Rate (BER)	≈0
Payload Recovery Accuracy	100%
Robustness (Noise/Compression)	High

The layered design ensures high reliability and visual fidelity but increases system complexity due to multiple processing stages.

C. Attention-Enhanced Dual-Stream U-Net with Adaptive Cost Learning

The Attention-Enhanced Dual-Stream U-Net architecture focuses on improving steganographic security by learning embedding probability maps using adversarial training and spatial attention mechanisms [3].

TABLE III: Reported Performance of Attention-Enhanced Dual-Stream U-Net Architecture [3]

Metric	Reported Value
Embedding Distortion (MSE)	Reduced vs. baselines
Steganalysis Detection Error	Improved
Security vs. ASDL-GAN	Superior
Security vs. UT-GAN	Superior
Dataset	BOSSBase Variants

This architecture achieves strong resistance to modern steganalysis, though at the cost of high training complexity.

D. Hybrid CNN–DCT Frequency-Domain Embedding Architecture

The Hybrid CNN–DCT architecture combines CNN-based region selection with frequency-domain embedding to improve robustness and imperceptibility in cloud-oriented scenarios [1]. This hybrid strategy provides a practical balance between robustness and computational efficiency, making it suitable for cloud-based applications.

TABLE IV: Reported Performance of Hybrid CNN–DCT Architecture [1]

Metric	Reported Value
Mean Squared Error (MSE)	1.25
Bit Error Rate (BER)	0.028
PSNR (dB)	37.4
SSIM	0.921
False Positive Rate (Steganalysis)	2.1%
Execution Time per Image	2.3 s

E. Overall Architectural Comparison

The comparison in Table V shows the trade-offs among recent deep learning image steganography architectures. Adaptive GAN-based encoder and decoder models achieve high embedding capacity and strong resistance to steganalysis through adversarial learning, but they come with increased training complexity and a reliance on error-correction methods. Multi-layered encoder and decoder frameworks focus on visual quality and strength by integrating compression, learning-based embedding, and lightweight spatial techniques. This results in reliable recovery but with higher complexity in the pipeline. Attention-enhanced dual-stream architectures target resistance to detection by learning adaptive embedding cost maps using attention methods. These offer strong security against modern steganalyzers but require significant computational resources. In contrast, hybrid CNN and DCT architectures focus on practical use by combining learned region selection with embedding in the frequency domain. They provide strength against compression and cloud-related distortions while maintaining moderate complexity. Overall, the analysis shows that increasing capacity and security often raises the architectural and computational costs. This drives the need for balanced encoder and decoder designs that achieve imperceptibility, strength, and practicality at the same time.

F. Trade-Offs and Selection Guidance

The comparative analysis shows that the choice of architecture depends heavily on application requirements. Encoder-decoder GAN-based frameworks work well in situations that value high embedding capacity. For example, VidaGAN achieves payloads of up to 3.9 bpp, but this comes with

TABLE V: Comprehensive Comparison of Reviewed Steganographic Architectures

Architecture	Approach	Datasets	Capacity	Imperceptibility	Robustness	Notes
Adaptive GAN Encoder–Decoder (VidaGAN) [4]	GAN encoder–decoder–critic, CSPNet backbone, MSE targeting, Reed–Solomon coding	DIV2K	3.9 bpp	MSE-targeted; StegExpose AUC ≈ 0.60	JPEG, noise, cropping	High capacity; relies on error correction; complex training
Multi-Layered Encoder–Decoder with Compression [2]	Huffman coding + DL encoder–decoder + optional LSB fallback	Tiny ImageNet, COCO, CelebA	Moderate	SSIM > 0.99; near-perfect recovery	Compression, noise, resizing	Layered security; high fidelity; increased pipeline complexity
Attention-Enhanced Dual-Stream U-Net (Adaptive Cost GAN) [3]	Dual-stream U-Net with CSA attention and Laplacian enhancement	BOSSBase variants	App.-dep.	Improved vs ASDL-GAN and UT-GAN	Strong vs CNN steganalyzers	Optimized for detectability; high training cost
Hybrid CNN–DCT Frequency-Domain Architecture [1]	CNN-based region selection with DCT-domain embedding	Custom images	0.45 bpp	MSE ≈ 1.25; PSNR ≈ 37.4 dB; SSIM ≈ 0.921	Compression, cloud distortions	Cloud-oriented; moderate complexity; lower capacity

increased detectability and a dependence on external error-correction methods for reliable decoding [4]. On the other hand, hybrid architectures like CNN-DCT and multi-layered Huffman-LSB-deep learning frameworks focus on imperceptibility and robustness against practical cloud changes. These include compression, resizing, and noise, making them suitable for secure multimedia sharing applications [1], [2]. When the main goal is to resist modern deep learning-based steganalyzers, adaptive cost learning approaches with attention mechanisms provide better security by focusing embedding changes in areas rich in texture and edges [3]. Lastly, in resource-limited environments, layered designs that can switch to lightweight embedding techniques offer a flexible balance between computational efficiency and security, allowing deployment across different hardware platforms [2].

VI. Challenges and Open Problems

Despite the advancements shown by recent deep learning-based steganography architectures, the reviewed works highlight several unresolved challenges and open research problems:

1) **Reliable decoding without heavy outer coding:** High-capacity encoder-decoder GAN frameworks such as VidaGAN achieve impressive payload rates. However, decoding errors still occur in practice, requiring the use of external error-correction schemes like Reed-Solomon coding to ensure reliable recovery [4]. While effective, this reduces the net usable payload and complicates end-to-end system design. Achieving reliably decoding without heavy outer coding remains an open challenge.
2) **Security against adaptive deep steganalysis:** Adaptive cost learning and attention-enhanced architectures improve resistance to existing steganalyzers by learning embedding probability maps that focus on texture- and edge-rich areas [3]. However, steganalysis methods are using deep neural networks that can adapt to new embedding strategies. Maintaining long-term security against evolving, data-driven steganalyzers remains an unresolved arms race.
3) **Architectural complexity and training efficiency:** GAN-based frameworks and attention-driven dual-stream networks offer good security benefits but require computationally intensive and sometimes unstable adversarial training [3], [4]. In contrast, multi-layered and hybrid designs reduce training instability but create complex processing pipelines that are hard to optimize end-to-end [2]. Designing architectures that balance security, stability, and training efficiency remains an open problem.
4) **Adaptability versus fixed embedding rules:** Hybrid CNN-DCT architectures show robustness under compression and cloud-related distortions, but their dependence on predefined frequency-domain embedding rules limits adaptability to different image content and evolving detection methods [1]. Bridging the gap between rule-based robustness and fully adaptive learning-based embedding in the frequency domain is a promising research direction.
5) **Evaluation consistency and reproducibility:** The reviewed studies use different datasets, payload settings, and evaluation metrics, making direct comparisons difficult. While some works emphasize capacity and detectability [3], [4], others focus on visual fidelity and robustness [1], [2]. Establishing standardized benchmarks, shared datasets, and common steganalysis protocols is essential for reproducible and meaningful evaluation.

VII. Practical Recommendations

Based on the architectural analysis and reported experimental results of the reviewed works, the following practical recommendations can guide the selection of steganographic frameworks for real-world applications:

1) **High-capacity data hiding scenarios:** For applications that need maximum payload capacity within a single image, encoder, decoder GAN-based architectures are suggested. Frameworks like VidaGAN show a much higher embedding capacity than hybrid methods, as long as external error-correction mechanisms are acceptable to ensure reliable decoding [4].
2) **Cloud-based image storage and transmission:** In settings where images go through compression, resizing, and transmission noise, hybrid CNN, DCT and multi-layered encoder, decoder frameworks provide a more reliable and practical solution. Their focus on making changes undetectable and resisting common distortions makes them suitable for secure cloud-based multimedia applications [1], [2].
3) **Security-critical covert communication:** When it is crucial to resist modern deep learning-based steganalyzers, adaptive cost learning architectures with attention mechanisms should be preferred. Dual-stream attention-based models effectively focus embedding changes in statistically safe areas, enhancing security against advanced detectors [3].
4) **Resource-constrained or heterogeneous deployments:** For systems that operate with limited computational resources or across different hardware platforms, layered designs that allow fallback to lightweight embedding methods provide a good balance between efficiency and security. Multi-layered frameworks that combine compression, deep learning, and simple spatial embedding are particularly suitable in these situations [2].

VIII. Conclusion

This review provided an in-depth look at recent deep learning-based image steganography architectures, focusing on encoder-decoder GAN frameworks, adaptive cost learning models with attention mechanisms, multi-layered hybrid designs, and CNN-DCT frequency-domain approaches. By comparing different designs, the review explored how these choices affect embedding capacity, imperceptibility, robustness, security against steganalysis, and practical use.

The analysis shows that encoder-decoder GAN frameworks excel at maximizing payload capacity through end-to-end learning and adversarial optimization. However, they often depend on external error-correction methods and involve high training complexity. Adaptive cost learning architectures enhance resistance to modern deep learning-based steganalyzers by creating embedding probability maps that focus modifications in statistically safe areas. Still, their computational needs and training instability limit real-time use. Multi-layered frameworks provide strong imperceptibility and robustness by combining compression, learning-based embedding, and lightweight spatial techniques, but they increase pipeline complexity. Hybrid CNN-DCT approaches find a practical balance by using learned region selection and transform-domain robustness, making them especially suitable for cloud-oriented scenarios, although they have limited adaptability due to fixed embedding rules.

A common theme across all the reviewed works is the trade-off between capacity, security, robustness, and computational efficiency. No single architecture optimizes all these goals at once, emphasizing the need for application-driven design choices. The review also points out ongoing challenges related to reliable decoding without heavy outer coding, resilience against adaptive steganalysis, efficient training and deployment, architectural scalability, and reproducible evaluation.

In summary, this review highlights that the future of image steganography relies on creating streamlined architectures that balance learning-based adaptability with practical limitations, supported by standardized benchmarks and thorough evaluation against evolving steganalysis methods. By gathering insights from recent developments and pinpointing key research gaps, this work aims to provide a useful reference for researchers and practitioners working on the next generation of secure and deployable steganographic systems.

REFERENCES

[1] S. Ahmad, J. O. Ogala, F. Ikpotokin, M. Arif, J. Ahmad, and S. Mehfuz, "Enhanced CNN-DCT Steganography: Deep Learning-Based Image Steganography Over Cloud," *SN Computer Science*, vol. 5, no. 408, 2024, DOI: 10.1007/s42979-024-02756-x.

[2] Y. Sanjalawe, S. Al-E'mari, S. Fraihat, M. Abuallraj, and E. Alzubi, "A deep learning-driven multi-layered steganographic approach for enhanced data security," *Scientific Reports*, vol. 15, no. 4761, 2025, DOI: 10.1038/s41598-025-89189-5.

[3] 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, p. 123471, 2024.

[4] V. Y. Ramandi, M. Fateh, and M. Rezvani, "VidaGAN: Adaptive GAN for image steganography," *IET Image Processing*, vol. 18, pp. 3329-3342, 2024, DOI: 10.1049/ipr2.13177.

[5] M. M. Mahmoud and H. T. Elshoush, "Enhancing LSB using binary message size encoding for high capacity, transparent and secure audio steganography–An innovative approach," *IEEE Access*, vol. 10, pp. 29954-29971, 2022.

[6] I. Kich and Y. Taouil, "CNN auto-encoder network using dilated inception for image steganography," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 21, no. 4, pp. 358-368, 2021.

[7] M. Dalal and M. Juneja, "A secure video steganography scheme using DWT based on object tracking," *Information Security Journal: A Global Perspective*, vol. 31, no. 2, pp. 196-213, 2022.

[8] R. Meng, W. Chen, W. Wang, Q. Chen, and Z. Cai, "A Fusion Steganographic Algorithm Based on Faster R-CNN," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1-16, 2018.

[9] D. Polisetty and S. W. A. Rizvi, "GAN-based Adaptive Image Steganography," *International Journal of Computer Applications*, vol. 187, no. 6, pp. 45-50, May 2025.

[10] N. Subramanian, O. Elharrouss, S. Al-Maadeed, and A. Bouridane, "Image Steganography: A Review of the Recent Advances," *IEEE Access*, vol. 9, pp. 23409-23424, 2021.

[11] S. Baluja, "Hiding Images in Plain Sight: Deep Steganography," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[12] X. Yuan, Y. Yang, W. Zhang, H. Chen, J. Huang, and N. Yu, "IAAE-Stega: Generic Blockchain-based Steganography Framework via Invertible Adversarial Autoencoder," *IEEE Transactions on Network Science and Engineering*, 2025.

[13] I. Hussain, J. Zeng, and S. Tan, "A Survey on Deep Convolutional Neural Networks for Image Steganography and Steganalysis," *KSII Transactions on Internet & Information Systems*, vol. 14, no. 3, pp. 1219-1240, 2020.

[14] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang, "CNN-based adversarial embedding for image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 8, pp. 2074-2087, 2019.

[15] B. Ray, J. Mukhopadhyay, S. Hossain, S. B. Goswami, A. K. Bhuiyan, and M. H. Rahman, "Image steganography using deep learning-based edge detection," *Multimedia Tools and Applications*, vol. 80, no. 24, pp. 33475-33503, 2021.

[16] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303-38314, 2018.