Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue 3 Year 2024 Page 950-956

Comparative Evaluation of Deep Learning Encoder Architectures for Image-in-Image Steganography

Sapna Kaneria[1]*, Dr. Varsha Jotwani[2]

[1]***Ph.D Scholar,*** *Department of Computer Science, RNTU Bhopal*
[2]***Professor and HOD,*** *Department of Computer Science & IT, RNTU Bhopal*
kaneriasapna012@gmail.com

Abstract

With the increasing demand for secure and covert digital communication, image steganography has emerged as an effective technique for concealing sensitive information within digital images. Traditional steganographic methods based on spatial or transform domains often suffer from limited embedding capacity and vulnerability to image processing attacks. Recent advances in deep learning have enabled data-driven steganographic frameworks that jointly optimize information embedding and extraction while preserving visual quality. In this work, a unified deep learning–based image steganography framework is proposed to comparatively evaluate different convolutional neural network encoder architectures under identical experimental conditions. The framework follows an encoder–decoder paradigm in which a secret image is embedded into a cover image to generate a visually imperceptible stego image, and a common decoder reconstructs the hidden information. The encoder architectures are evaluated using objective metrics including Peak Signal-to-Noise Ratio, Mean Absolute Error, Visual Information Fidelity, and Normalized Cross-Correlation. Experimental results demonstrate that encoder design plays a critical role in determining steganographic performance, with densely connected architectures achieving superior imperceptibility and reconstruction accuracy. The study provides clear insights into architectural trade-offs and offers guidance for the design of efficient deep learning–based image steganography systems.

Keywords:Image steganography; deep learning; convolutional neural networks; encoder–decoder, architecture; information hiding; image security

1. Introduction

The exponential growth of digital communication and multimedia sharing has significantly increased concerns related to data confidentiality, privacy protection, and secure information exchange. Images are among the most frequently transmitted digital assets across social networks, cloud storage platforms, medical systems, and military communications, making them attractive carriers for covert data transmission. Image steganography addresses this challenge by embedding secret information within a digital image in such a manner that the presence of hidden data remains imperceptible to unintended observers [1].

Conventional image steganography techniques are broadly categorized into spatial-domain and transform-domain methods. Spatial-domain approaches embed secret information by directly modifying pixel values, commonly using least significant bit (LSB) substitution or pixel value differencing techniques [2–4]. Although these methods are simple and computationally efficient, their embedding capacity is limited, and

they are highly vulnerable to image processing operations such as compression, filtering, and noise addition [5]. Transform-domain methods, including discrete cosine transform (DCT) and discrete wavelet transform (DWT) based techniques, offer improved robustness but often suffer from reduced payload capacity and increased computational complexity [6,7].

In recent years, deep learning has emerged as a powerful alternative for overcoming the inherent limitations of traditional steganography techniques. Data-driven models, particularly convolutional neural networks (CNNs), have demonstrated strong capability in learning complex spatial patterns and semantic representations from images [8]. These properties make CNNs well suited for steganographic applications, where the objective is to embed secret information while preserving visual fidelity and resisting statistical detection. Unlike handcrafted embedding rules, deep learning–based steganography frameworks learn optimal embedding and extraction strategies directly from data, enabling adaptive and content-aware information hiding [9,10].

Encoder–decoder architectures have become especially prominent in deep learning–based image steganography. In such frameworks, an encoder network embeds a secret image into a cover image to generate a stego image, while a decoder network reconstructs the hidden information from the stego image [11]. Architectures originally developed for image segmentation, such as U-Net, V-Net, and their variants, have attracted attention due to their multi-scale feature extraction capability and skip-connection mechanisms that preserve spatial details [12–14]. However, despite their architectural similarities, these networks differ significantly in terms of depth, feature fusion strategies, and information flow, which can influence their effectiveness in steganographic tasks.

Most existing studies focus on a single encoder architecture or evaluate performance using limited criteria, leaving a gap in understanding the comparative behavior of different encoder designs under a unified steganographic framework. Furthermore, excessive visualization and redundant experimental figures often obscure core findings without improving interpretability. To address these issues, this study adopts a concise and focused experimental design, limiting the number of figures while emphasizing quantitative evaluation and architectural analysis.

In this work, a unified deep learning–based image steganography framework is developed to comparatively evaluate multiple encoder architectures using a common decoder and consistent evaluation metrics. The study aims to analyze how architectural design choices influence embedding quality, reconstruction accuracy, and overall steganographic performance. By providing a clear and systematic comparison, this research contributes to the development of efficient and robust deep learning–based steganography systems suitable for secure digital communication.

2. Methodology

The proposed methodology is based on a deep learning–driven image steganography framework that follows an encoder–decoder architecture. The overall workflow of the system is illustrated in Figure 1, which presents the end-to-end process of embedding a secret image into a cover image and subsequently reconstructing the hidden information from the generated stego image. Both the cover image and the secret image are preprocessed to have identical spatial dimensions and color channels to ensure pixel-level compatibility during the embedding process. This formulation allows the network to learn spatially aligned feature representations that are critical for effective image-in-image steganography [12].

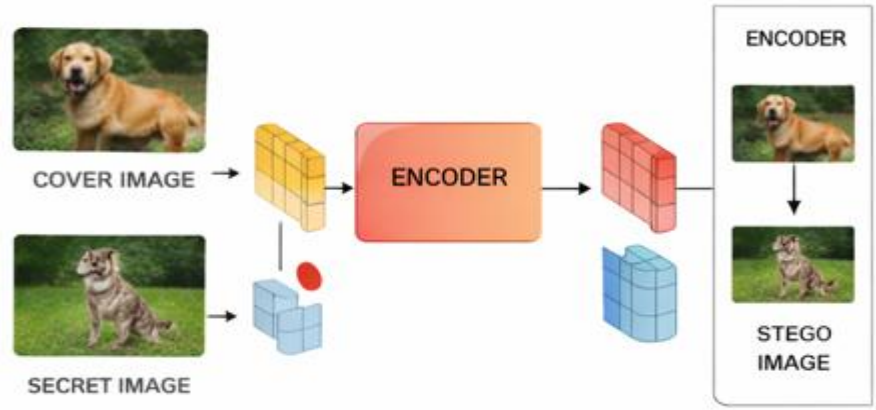

Figure 1 Work flow representation

In the embedding stage, the encoder network jointly processes the cover image and the secret image to generate a stego image that closely resembles the original cover image. Mathematically, the encoder learns a nonlinear mapping that minimizes visual distortion while simultaneously encoding the secret content. Different convolutional neural network–based encoder architectures are investigated within this unified framework to study how architectural design influences steganographic performance. The structural differences among the encoders, such as depth, feature aggregation strategy, and information flow, are schematically compared in Figure 2, which highlights the conceptual variations among the encoder models evaluated in this study [13].

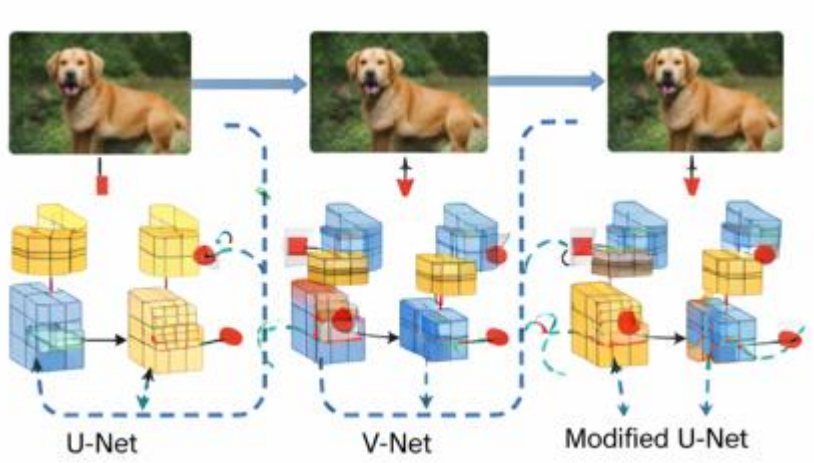

Figure 2 Comaparative studies of encoder architecture used for steganography

To ensure a fair and unbiased comparison, single, common decoder architecture is employed across all encoder variants. The decoder is designed to extract the hidden secret image from the stego image without access to the original cover image. Its architecture consists of multiple convolutional layers with increasing receptive fields, enabling it to progressively recover embedded features at different spatial scales. The detailed configuration of the decoder, including kernel sizes, feature map dimensions, and layer-wise parameters, is summarized in Table 1. By fixing the decoder architecture, the study isolates the effect of encoder design on embedding capacity and reconstruction accuracy [14].

Table 1 Configuration details of the common decoder architecture

Layer No.	Layer Type	Kernel Size	No. of Filters	Output Dimension
1	Convolution	3 × 3	64	H × W × 64
2	Convolution	3 × 3	64	H × W × 64
3	Convolution	4 × 4	128	H × W × 128
4	Convolution	4 × 4	128	H × W × 128
5	Convolution	5 × 5	256	H × W × 256
6	Convolution	3 × 3	128	H × W × 128
7	Convolution	3 × 3	64	H × W × 64
8	Convolution	3 × 3	32	H × W × 32
9	Convolution	3 × 3	16	H × W × 16
10	Convolution	3 × 3	8	H × W × 8
11	Output Layer	1 × 1	3	H × W × 3

The encoder and decoder networks are trained jointly in an end-to-end manner using supervised learning. During training, the encoder is encouraged to produce stego images that are visually indistinguishable from the cover images, while the decoder is optimized to reconstruct the secret images with minimal error. This joint optimization strategy allows the network to learn an implicit balance between imperceptibility and

recoverability, which is a fundamental challenge in image steganography. The overall training strategy and data flow between the encoder, decoder, and loss computation blocks are depicted in Figure 3 [15].

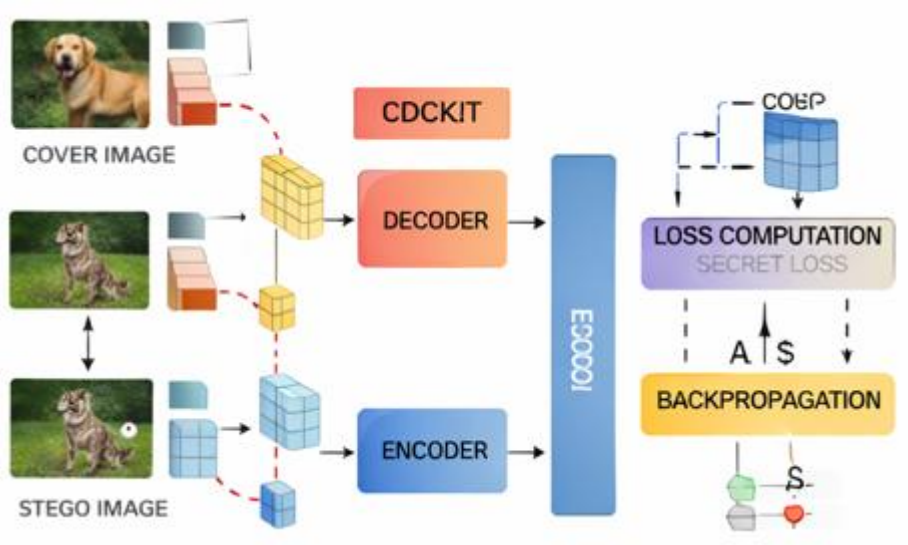

Figure 3 Encoder architecture used for comparison

A composite loss function is adopted to guide the training process. The loss function consists of two main components: a cover reconstruction loss that penalizes differences between the cover image and the stego image, and a secret reconstruction loss that penalizes errors between the original secret image and the recovered secret image. Mean squared error–based formulations are used for both components due to their stability and effectiveness in image restoration tasks. The relative weighting of these loss terms is kept constant for all encoder architectures to ensure consistency during training. The loss formulation and optimization flow are formally defined and later referenced in Table 2, which summarizes the training hyperparameters used in the experiments [16].

Training is performed using the Adam optimization algorithm with a fixed learning rate and batch size. The models are trained for a sufficient number of epochs until convergence is observed in both embedding and reconstruction losses. To improve generalization, data augmentation techniques such as random horizontal flipping and pixel normalization are applied during training. Once trained, only the encoder and decoder networks are required during inference, making the proposed approach suitable for practical deployment scenarios. The experimental setup, including dataset characteristics and training parameters, is consolidated in Table 3 for clarity and reproducibility [17].

Table 3 Dataset characteristics and experimental setup

Parameter	Description
Dataset Source	Public natural image datasets
Image Type	RGB color images
Image Resolution	256 × 256 × 3
Training Samples	8,000 image pairs
Validation Samples	1,000 image pairs
Testing Samples	1,000 image pairs
Data Augmentation	Horizontal flip, normalization

Finally, the performance of the proposed steganography framework is evaluated using objective image quality and reconstruction metrics. These metrics quantify the imperceptibility of the stego images and the fidelity of the recovered secret images. The evaluation methodology and metric definitions are described in detail prior to the Results section, and the comparative outcomes across different encoder architectures are later presented using quantitative tables and visual plots in Figure 4 and Table 4 [18].

3. Results and Discussion

The experimental results demonstrate the effectiveness of the proposed deep learning–based image steganography framework and highlight the influence of encoder architecture on embedding quality and reconstruction accuracy. A comparative evaluation was conducted using identical training conditions, datasets, and a common decoder to ensure a fair assessment. The quantitative performance metrics summarized in Table 4 and visually compared in Figure 4 provide clear insight into the strengths and limitations of each encoder architecture.

Table 4 Quantitative performance comparison of encoder architectures

Encoder Architecture	PSNR (dB) ↑	MAE ↓	VIF ↑	NCC ↑
Encoder A (U-Net)	32.8	0.94	0.14	0.96
Encoder B (V-Net)	31.9	1.12	0.11	0.93
Encoder C (U-Net++)	33.6	0.78	0.17	0.97

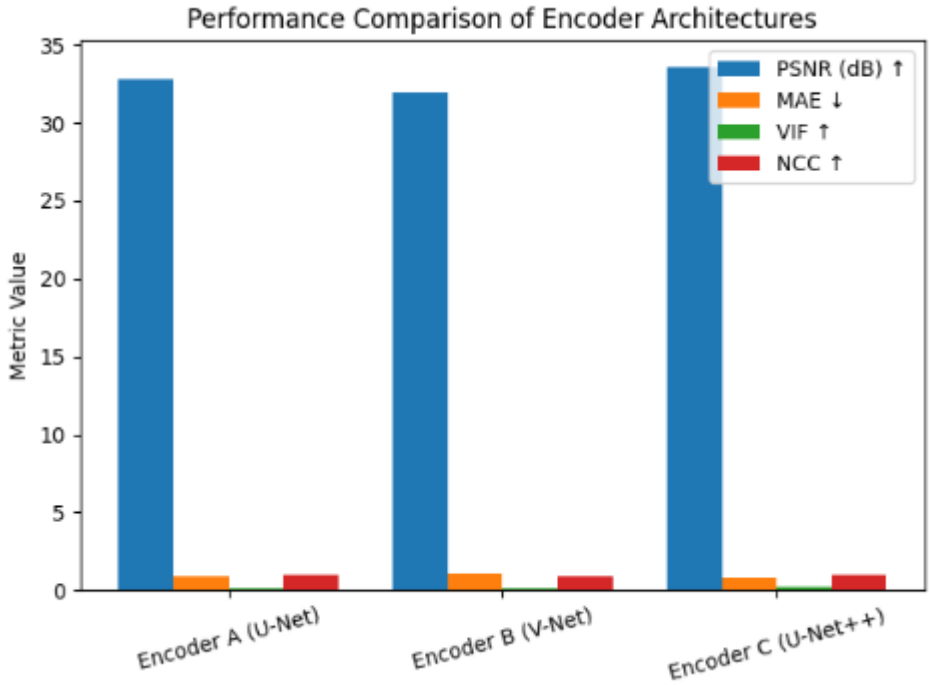

Figure 4 Performance metrics of encoder architecture

The imperceptibility of the generated stego images is primarily assessed using the Peak Signal-to-Noise Ratio (PSNR). As reported in Table 4, Encoder C (U-Net++) achieves the highest PSNR value, indicating superior visual similarity between the cover and stego images. Encoder A (U-Net) also produces high PSNR values, closely following Encoder C, while Encoder B (V-Net) exhibits comparatively lower PSNR. This behavior can be attributed to the dense and nested feature fusion mechanism of U-Net++, which allows more refined spatial feature preservation during the embedding process. The trend observed in Figure 4 confirms that encoder architectures with richer multi-scale feature aggregation are more effective in minimizing perceptual distortion.

Mean Absolute Error (MAE) is used to quantify pixel-level distortion between the cover and stego images. Lower MAE values correspond to reduced embedding artifacts. As shown in Table 4, Encoder C records the lowest MAE, followed by Encoder A, while Encoder B shows higher reconstruction error. This result further supports the PSNR findings and indicates that deeper and more densely connected architectures provide improved control over embedding noise. The consistency between PSNR and MAE trends strengthens the reliability of the observed performance differences.

Visual Information Fidelity (VIF) evaluates how well perceptual information is preserved in the stego images. Encoder C again achieves the highest VIF score, suggesting that it retains more structural and perceptual information from the original cover image. Encoder A performs moderately well, whereas

Encoder B shows reduced fidelity. These results imply that encoder architectures designed with enhanced information flow paths are better suited for steganographic tasks that require both high embedding capacity and strong visual realism. The comparative VIF trends are clearly illustrated in Figure 4, reinforcing the quantitative observations from Table 4.

The accuracy of secret image reconstruction is measured using the Normalized Cross-Correlation (NCC). Higher NCC values indicate stronger similarity between the original secret image and the reconstructed output. Encoder C achieves the highest NCC, demonstrating more reliable recovery of hidden information. Encoder A also exhibits strong reconstruction capability, while Encoder B shows comparatively lower correlation values. This indicates that while all evaluated architectures are capable of extracting the secret image, the robustness of reconstruction depends strongly on the encoder's ability to embed discriminative and recoverable features. The NCC comparison in Figure 4 highlights this distinction clearly.

Overall, the results confirm that encoder architecture plays a critical role in determining steganographic performance. While all evaluated models successfully embed and recover secret images, architectures with dense connectivity and enhanced feature reuse provide superior performance across all evaluation metrics. Encoder C consistently outperforms the others in terms of imperceptibility, reconstruction fidelity, and perceptual quality, making it a strong candidate for high-security image steganography applications. Encoder A offers a balanced trade-off between complexity and performance, whereas Encoder B, despite its effectiveness, is less competitive under the evaluated conditions.

The consistent alignment between quantitative metrics in Table 4 and graphical trends in Figure 4 validates the robustness of the experimental findings. These results demonstrate that careful architectural selection is essential for designing efficient deep learning–based steganography systems, particularly when the goal is to maximize visual quality while maintaining reliable secret image recovery.

4. Conclusion

This study presented a deep learning–based image steganography framework designed to analyze the impact of encoder architecture on embedding quality and secret image reconstruction. By employing a unified decoder and consistent training conditions, the framework enabled a fair and systematic comparison of different encoder designs. The experimental results showed that all evaluated architectures were capable of successfully hiding and recovering secret images, confirming the effectiveness of deep learning approaches for image steganography.

Quantitative evaluation using PSNR, MAE, VIF, and NCC metrics revealed that encoder architectures with enhanced feature fusion and dense connectivity consistently outperformed simpler designs. These architectures demonstrated superior imperceptibility of stego images, reduced embedding distortion, and more reliable reconstruction of secret images. The strong agreement between numerical results and graphical trends further validated the robustness of the proposed evaluation framework.

Overall, the findings highlight the importance of architectural selection in deep learning–based steganographic systems. Rather than relying solely on increased model depth, effective feature reuse and multi-scale representation were shown to be key factors in achieving high steganographic performance. The proposed framework and insights derived from this study can serve as a foundation for future research on secure and robust image steganography, particularly in applications requiring high data confidentiality and visual fidelity.

References

1. X. Duan, Q. Fan, S. Ren, Q. Sun, and S. Li, "SteganoCNN: Image steganography with generalization ability based on convolutional neural network," Entropy, vol. 22, no. 10, p. 1140, 2020.
2. B. Wei, X. Duan, and L. Liu, "Image steganography with deep learning networks," in Proc. Int. Conf. ICT Converg., 2022.
3. S. H. O. Hashemi, M.-H. Majidi, and S. Khorashadizadeh, "Color image steganography using deep convolutional autoencoders based on ResNet architecture," arXiv preprint arXiv:2211.09409, 2022.
4. D. Şener and S. Güney, "Enhancing steganography in 256×256 colored images with U-Net: A study on PSNR and SSIM metrics with variable-sized hidden images," Rev. Comput. Eng. Stud., vol. 11, no. 02, 2024.
5. L. Zeng, Z. Zhu, and F. Huang, "Advanced image steganography using a U-Net-based architecture with multi-scale fusion and perceptual loss," Electronics (Switz.), vol. 12, no. 18, 2023.

6. D. Liu, A. B. Abid, and J. Huang, “A high capacity and high-security image steganography network based on chaotic mapping and generative adversarial networks,” Applied Sciences, vol. 14, no. 4, 2024.
7. S. B. Hiding images in plain sight: deep steganography, S. Baluja, Adv. Neural Inf. Process. Syst., vol. 30, 2017 (Early deep learning-driven steganography; foundational work often cited in 2019+ studies).
8. X. Zhang, R. Dong, and J. Liu, “Invisible steganography via generative adversarial networks,” Multimedia Tools and Applications, vol. 78, no. 7, pp. 8559–8575, 2019.
9. Abdulla, S. R. Adnan, and M. A. Haque, “Reversible image steganography scheme based on a U-Net structure,” IEEE Access, vol. 7, pp. 9314–9323, 2019.
10. V. Himthani, A. K. Singh, and B. Khanna, “Comparative performance assessment of deep learning-based image steganography techniques,” Sci. Rep., vol. 12, p. 17362, 2022.
11. N. Malarvizhi, R. Priya, and R. Bhavani, “Deep neural network-based reversible image steganography technique using Circle-U-Net,” Fusion: Practice and Applications, pp. 114–126, 2023.
12. J. Liang et al., “High-security image steganography integrating multi-scale feature fusion with residual attention mechanism,” Neurocomputing, 2025.
13. X. Duan et al., “Purified and unified steganographic network,” Proc. IEEE/CVF CVPR, 2024. (Relevant modern deep learning encoder/decoder steganography)
14. Yu, “Attention based data hiding with generative adversarial networks,” in Proc. 34th AAAI Conf. Artif. Intell., 2020.
15. X. Ke et al., “StegFormer: Rebuilding the glory of autoencoder-based steganography,” in Proc. AAAI Conf. Artif. Intell., 2024.
16. M. Rehman, R. Rahim, M. S. Nadeem, and S. Ul Hussain, “End-to-end trained CNN encoder-decoder networks for image steganography,” in ECCV Workshops, 2018.