

RESEARCH ARTICLE

OPEN ACCESS

TRANSFORMER DRIVEN SELF-ATTENTION FRAMEWORK FOR HIGH FIDELITY COLOR IMAGE STEGANOGRAPHY

Taha Y. Abdulqader¹, Shatha A. Baker^{*2}, Kifaa Hadi Thanoon³ and Yaarob Younus Al-Badrani⁴

^{1, 2}Networks and Computer Software Techniques, Northern Technical University, Mosul, Iraq

³Cyber Security, Northern Technical University, Mosul, Iraq.

⁴Quality Assurance and University Performance, Northern Technical University, Mosul, Iraq

¹<https://orcid.org/0009-0009-3150-2963>^{id}, ²<https://orcid.org/0000-0002-5903-0744>^{id}

³<https://orcid.org/0000-0002-8067-0500>^{id}, ⁴<http://orcid.org/0009-0006-2695-6002>^{id}

E-mail: taha.yassin@ntu.edu.iq, [*Shathaab@ntu.edu.iq](mailto:Shathaab@ntu.edu.iq); kh.thanoon@ntu.edu.iq; yaarobalbdbrany@ntu.edu.iq

ARTICLE INFO

Article History

Received on: March 08, 2026

Reviewed on: April 08, 2026

Accepted on: July 01, 2026

Published on: July 31, 2026

Keywords:

AI,
Hiding,
RGB,
Steganography.

ABSTRACT

A significant issue in digital steganography is concealing color images within natural photographs, when the objective is to convey a large quantity of hidden information without sacrificing visual quality. Certain conventional techniques that rely on the spatial or frequency domains typically leave behind noticeable artifacts and are readily interfered with by filtering, noise, or compression. Some transformer-based models have recently demonstrated exceptional ability to capture intricate linkages and long-range interdependence across data channels. Building on this, we have presented a transformer-powered framework for color image concealment that learns the best embedding patterns across RGB channels by using self-attention. In comparison to both conventional techniques and previous deep learning methods, our results demonstrate that this methodology provides considerably greater imperceptibility and recovery, indicating transformers as a promising avenue for enhanced image steganography.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The exponential growth of digital media, data protection, and secure communication is more crucial than ever in the world of modern computing [1-3]. Digital steganography, the art of covertly encoding messages into seemingly innocuous media, is one of the numerous methods that have been created in this approach thus far. It has proven to be an essential instrument in ensuring privacy and safeguarding against unauthorized access. One of the most active research areas is color image hiding, in which a hidden image is concealed by a picture of the outdoors. The application of this is popular in such applications as secure messaging, watermarking, copyright protection, and covert communication. In comparison to textual steganography [4], the steganography of color images presents special challenges since it should consider the spatial as well as the spatial characteristics of images. As human vision is very sensitive to color, texture, and structure distortions, the task is to balance two competing objectives: to withstand as much hidden information (payload capacity) and to maintain the cover image indistinguishable (imperceptibility). Although there has been a lot of progress [5], this balance is hard to achieve. The existing procedures can be divided into two main categories, which include spatial domain methods described in [6] and transform-domain methods described in [7]. Spatial-domain techniques, which include the Least Significant Bit (LSB) substitution method described in [8], directly change pixel values inside the cover image. The strategies show simplicity and strong performance, but they create clear visual defects that become visible in high-quality color photographs, and their results get affected by basic image processing tasks such as compression and filtering. The execution of transform-based techniques happens through the processing of frequency space data, which includes methods such as Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). The methods deliver average protection against standard attacks, but they need users to set up parameters manually, and their efficiency drops when handling large data sizes or full-color images. This is largely an outcome of the RGB channels correlating with the high sensitivity of human vision in the detection of the smallest distortions. [9]. The latest advancements in deep learning technology have created a new system for concealing images.

The combination of autoencoders [10] with generative networks [11] and Convolutional Neural Network (CNN)-based frameworks has achieved outstanding performance in their capacity to learn complex mappings, which they use to conceal and restore images. The system achieves automatic process optimization, which enables it to create better embedding methods while decreasing visual defects and enhancing data recovery capabilities. Researchers in deep learning-based steganography studies have focused their research efforts on grayscale image analysis [10] and single-channel payload examination. The researchers used CNNs because the network excels at local feature extraction, but they faced challenges when testing color channel connections between remote locations and through large distance ranges. The subtle RGB channel interactions that determine visual imperceptibility in full-color image hiding cause CNN-based methods to lack essential contextual data, which results in color distortions and decreased system resilience. [12] We present a transformer-based framework that enables color image hiding as a solution to these difficulties. Transformers [13] begin with natural language processing development because their self-attention mechanisms let them track all input dependencies, which enables them to effectively handle long-distance relationships. Recent years have seen successful transformer adaptations that enable machines to complete various vision tasks, such as image generation [14], super-resolution [15], and image-to-image translation, because these tasks require machines to understand their complete visual environment. We demonstrate transformers as a solution for solving the challenge of hiding color images through our research. The network learns embedding techniques through our method, which enables it to transmit larger payloads while keeping visual quality because visual quality depends on its ability to represent both spatial patterns and inter-channel correlations. The proposed framework treats the cover image and secret image as sequences of image patches, which are processed by a transformer encoder-decoder model. The encoder extracts information from the cover image to build a detailed representation, which includes both local texture elements and global structural elements. The decoder reconstructs the secret image from the embedded representation. The self-attention layers in the transformer system enable the model to perform channel-aware embedding, which results in better information distribution across RGB channels that decreases perceptual artifacts. The model training process includes robustness-oriented augmentation, which creates realistic image manipulations through Gaussian noise, JPEG compression, and median filtering. The system guarantees that hidden information remains recoverable under standard adversarial testing conditions. The main contributions of this work are presented in summary form as follows:

1. Our research presents a new approach to color image hiding, which uses transformer technology as an alternative to existing traditional and CNN-based methods.
2. Our method uses self-attention to track both inter-channel and spatial relationships, which decreases the visual distortion that affects the steganography image.
3. Our training methods show improved resistance to noise, compression, and filtering attacks through additional training methods.
4. Our method demonstrates better imperceptibility, recoverability, and payload capacity than existing methods according to extensive testing, which proves that transformer architectures effectively support advanced color steganography.

The study is as follows: the related works will be shown in the next section. The methodology is described in Section III. The experimental analysis is carried out in Section IV. The analysis of the results is shown in Section V, and in Section IV, we wrap up the study with some recommendations for further research.

II. THEORETICAL REFERENCE

II.1 TRADITIONAL IMAGE HIDING METHODS

Researchers have conducted extensive research on conventional image-hiding techniques first developed as steganography methods, which operate in both spatial and transform domains. The LSB substitution method, which functions as a spatial-domain technique [16], achieves its secret information embedding through direct pixel value modifications that specifically target the least visible parts of the image. The methods provide an easy way to execute operations while delivering an efficient system for carrying out high payloads. The system displays high vulnerability against various types of image processing activities, which include noise, compression, and additional processing methods. The methods fail to protect concealed information because even tiny changes result in substantial information loss, which prevents their usage for secure color image protection. The transform-domain methods use DCT [17] and DWT [18] techniques to embed information within frequency components of their systems. Better compression and filter resistance performance results from their method, which operates on coefficient data instead of using actual pixel brightness values. The system requires exact parameter settings to function properly, yet it fails to handle full-color data because RGB channels require mutual dependence. The spatial methods and transform domain methods use handcrafted rules, which restrict their capacity to create customized embedding solutions that suit different cover images and advanced attack situations.

II.2 CNN-BASED IMAGE HIDING

Deep learning has created a complete transformation of the field of image steganography through its introduction to digital image processing. CNNs [20]-19 have been widely applied to learn end-to-end mappings for hiding and recovering images. The Autoencoder-based framework encodes the secret image into a latent representation and decodes it from the steganography image while it achieves its goal of imperceptibility and recoverability. The CNN-based methods from [21]-20 use residual connections together with multi-scale feature extraction methods to achieve minimal visual distortions in their output. The existing methods achieve better results than traditional techniques when tested under controlled conditions, but they mainly work with grayscale payloads and single-channel embeddings. CNNs build their fundamental design around local receptive fields, which prevents them from effectively modeling the spatial and color relationships found in natural images. The process of hiding full-color images results in color distortions and perceptual artifacts that become worse under difficult attack situations involving compression and additive noise.

II.3 ATTENTION AND TRANSFORMER-BASED IMAGE PROCESSING

Transformers [13], which developers designed for natural language processing, now enable various computer vision applications that include image classification [21], super-resolution [22], and image-to-image translation [23]. Transformers use self-attention mechanisms to create models that identify all global dependencies by detecting connections between different spatial areas and different spatial channels. This feature enables transformers in image processing to develop contextual embeddings, which use both local texture information and distant interaction data for tasks that need precise results in image generation and restoration work. Recent studies [14] show that transformer-based architectures provide better performance than CNNs because they achieve superior color correlation detection and structural preservation, which makes them suitable for high-quality image steganography.

II.4 ROBUSTNESS AND ADVERSARIAL CONSIDERATIONS

Researchers face a major difficulty because they need to create image-hiding methods that can withstand various types of attacks, which include noise addition, JPEG compression, filtering, and geometric transformations. The initial methods used both spatial and transform domains but showed weak defense capabilities, which resulted in major hidden payload losses whenever there were slight distortions. The CNN-based frameworks in reference [24] solved these problems because they used data augmentation together with adversarial training to enable models to extract hidden data from noisy and compressed input. The local receptive nature of CNNs restricts the capabilities of such methods in reference [25] because they need to establish complete color embedding through global pattern detection. The transformer-based models enable better performance in robustness-based embedding because they utilize global context information to control hidden data distribution between spatial and channel dimensions, which allows for effective data retrieval during times of severe data processing.

III. MATERIALS AND METHODS

The proposed approach for color image hiding leverages the self-attention capabilities of transformer architectures to model inter-channel and spatial dependencies in Figure 1 while achieving precise secret image embedding into natural cover images. Our transformer-driven framework goes beyond conventional CNN-based methods that only detect local patterns because it uses global context to distribute concealed data through RGB channels while keeping everything undetectable. The overall methodology consists of three major components, which include image preprocessing and patch embedding together with transformer-based embedding and recovery and loss function and optimization.

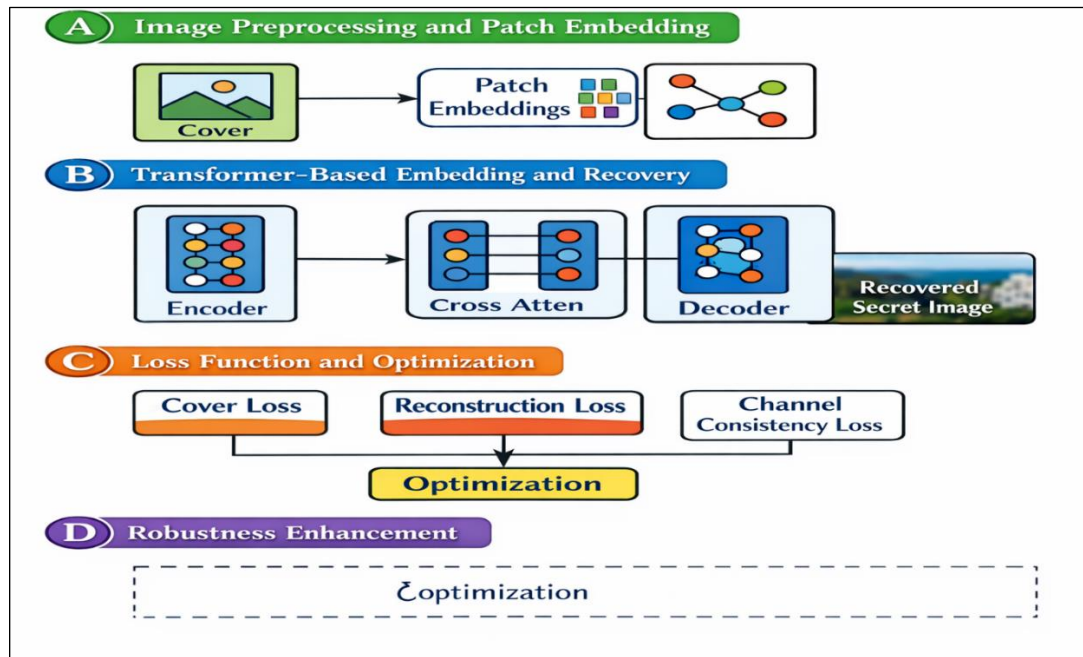


Figure 1: Modelling Architecture.

Source: Authors, (2026).

III.1 IMAGE PREPROCESSING AND PATCH EMBEDDING

The transformer architecture requires both the cover and secret images to be divided into separate patches that do not overlap with each other. The image, which has been resized to a standard size of 256×256 pixels, gets divided into 16×16 patches that create 256 patches for each image. The patch contains 16×16×3 pixel values for RGB channels, which get flattened before they undergo linear projection into a higher-dimensional space that creates patch embeddings. The model uses positional encodings to enhance these embeddings because they help maintain spatial information about patch locations during the entire embedding and recovery process. The patch-based representation transforms a 2D image into a sequence that the transformer can use with its attention mechanism, which enables the network to detect connections between different spatial areas as well as between different color channels. The process of patching and embedding the secret image generates a latent representation that maintains its original content and structure, which can be combined with the representation of the cover image.

III.2 TRANSFORMER-BASED EMBEDDING AND RECOVERY

The main element of our model uses a transformer encoder-decoder system to enable complete image hiding between all processing stages. The encoder creates a contextual latent representation from patch embeddings of the cover image, which it uses to capture both fine-grained textures and overall structural patterns. Parallel to these operations, embeddings of the secret image are introduced to cross-attention layers, where the model is able to cohesively inject secret information into certain areas of the cover image representation. This channel-conscious embedding approach works out hidden data efficiently across the RGB channels to minimize the perceptual distortions and ensures a high payload capacity. This modified representation can be successfully decoded by the decoder to reconstruct the secret image, which is essentially the reverse process of the embedding process. Self-attention layers allow the model to learn the spatial and color correlation of the steganographic image to increase fidelity in the recovered secret. To ensure training stability and low-level detail retention, residual connections are also used between encoder and decoder, and layer normalization is used to ensure convergence and eliminate vanishing gradients. Moreover, multi-head attention enables the network to concentrate on a variety of visual features, e.g., edges, textures, and color gradients, at the same time, which is crucial to the imperceptibility of high-resolution images.

III.3 LOSS FUNCTION AND OPTIMIZATION

A composite loss is used to direct the training goal and balance three objectives, namely: (1) the steganographic image should be imperceptible, (2) the steganographic secret should be fidelity, and (3) the channel colors should be similar. Towards this goal, the cover loss punishes discrepancies between the original and steganographic cover with Mean Squared Error (MSE), to guarantee few visible artifacts. Reconstruction loss also applies MSE with a focus on the recovered secret image, and it directly maximizes accurate recovery. The channel-consistency loss system punishes RGB channel differences because it creates artificial color changes, which make the perceptual experience more authentic. The Adam optimizer controls learning rate transitions through its schedule, which enables smooth convergence. The system achieves better stability through two methods, which include clipping all gradients to remove explosive gradients and using early stopping based on validation metrics to stop overfitting.

III.4 ROBUSTNESS ENHANCEMENT

The framework testing process needs robustness-based training because this method tests the system's ability to withstand real-world performance challenges. The steganographic pictures undergo random testing, which includes Gaussian noise application and JPEG compression at multiple quality settings and median filter processing. The augmentations replicate storage and transmission methods that create typical image degradation patterns that occur during those processes. Training under such a regime, the model learns to store information in a manner that it can be recovered despite perturbations. In addition, a cross-attention mechanism is used to reinforce critical information in strong areas of the cover image and enhance secret recovery in the presence of noise, compression, or filters.

IV. EXPERIMENTAL ANALYSIS

IV.1 DATASETS

We use a variety of large, high quality natural image data to train and test our transformer-based framework on color image hiding. We use COCO (Common Objects in Context) [26] as our main dataset containing more than 300,000 images and diverse objects, scenes, and textures. Such diversity assists the model in learning embedding techniques that allow generalizing in various situations - not only indoor settings and natural scenery but also human-oriented images. We also use the DIV2K dataset [27] that contains high-resolution images that are suitable to test payload capacity and perceptual fidelity. To enhance uniformity, all of the images are scaled to 256 x 256 pixels, which is made to result in efficient transformation by the transformer architecture. Secret images have been randomly chosen among the same datasets, although they are never the same as the cover images that accompany them. The cover and secret images are normalized to a standard intensity range and RGB-formatted in order to maintain the full-color embedding. The system becomes more powerful through multiple augmentations, which include Gaussian noise, various levels of JPEG compression, and median filtering, together with additional methods. The model training process uses these augmentations to test the system's ability to handle actual transmission and storage distortions, which will lead to precise information recovery.

IV.2 EVALUATION METRICS

The framework assessment uses metrics that measure three different aspects of system performance together with its ability to withstand attacks.

- The Peak Signal-to-Noise Ratio (PSNR) test, which compares the cover image to the steganographic image, shows that higher PSNR values result in less visual distortion from the embedding process.
- The Structural Similarity Index (SSIM) test measures both the perceptual and structural accuracy of the cover and steganographic content as well as of the content that has been secretly recovered.
- Secret recovery tests use PSNR and SSIM measurements to evaluate the system after applying three types of distortion, which include Gaussian noise, JPEG compression, and median filtering.
- Researchers express payload capacity through bits per pixel (bpp) values, which show the quantity of concealed information that each cover image pixel can accommodate.
- The CIEDE2000 ΔE test evaluates color consistency by measuring perceptual color differences, which demonstrate that the steganographic image maintains authentic color reproduction.

IV.3 BASELINES

The assessment process requires us to evaluate our transformer-based approach against selected baseline models from spatial, transform, and deep learning fields.

- The LSB substitution technique operates as a basic method that enables users to insert information into pixel values throughout digital content. The system enables users to transmit high payloads yet suffers from severe data loss when subjected to compression or noise attacks.
- The DCT-based hiding method establishes frequency coefficient information embedding, which delivers medium defense capabilities but restricts large color information transmission.
- DeepStega [10] utilizes CNN autoencoder methods that employ convolutional networks to discover pattern embedding techniques. The methods succeed at detecting local features yet display limitations when attempting to process extended image data and multiple color channels in full-color image hiding.

The evaluation process for all baseline systems requires their training and testing on identical datasets while using the same attack methods, which include Gaussian noise, JPEG compression, and median filtering. The evaluation process operates with identical metrics, which include PSNR and SSIM, bpp, and CIEDE2000 ΔE .

IV.4 HYPERPARAMETERS

The model uses 12 transformer encoder blocks and 12 decoder blocks, which both contain 8 attention heads. The patch embedding dimension reaches 512, which enables the model to extract spatial and channel-level information at a reasonable processing expense. The Feed-Forward Network (FFN) dimension within each transformer block is configured to 2048, which enables the model to achieve high-quality embedding and recovery results. The training process uses a batch size of 16 images, which enables efficient GPU memory usage while providing accurate gradient calculations. The Adam optimizer uses an initial learning rate of $1e-4$, which it reduces through a cosine annealing schedule to achieve smooth convergence. The system uses gradient clipping with a 1.0 norm threshold to maintain operational stability while preventing deep transformer architectures from developing exploding gradients. The loss function weights use $\lambda_c = 1.0$, $\lambda_h = 2.0$, and $\lambda_{ch} = 0.5$ to show that accurate secret image recovery needs to be achieved while maintaining imperceptibility and color fidelity. The training process uses Gaussian noise with standard deviations between 0 and 0.05 and JPEG compression quality factors between 30 and 100 and median filtering with 3×3 and 5×5 kernels as robustness-oriented augmentation methods. The model completes 100 training epochs, which use early stopping to check validation PSNR, so it achieves high-quality results without overfitting.¹

V. RESULT ANALYSIS

V.1 IMPERCEPTIBILITY OF STEGANOGRAPHY IMAGES

The ability to hide images through steganography depends on their visual stealth capabilities. The proposed transformer-based framework developed by our team achieved successful steganography through its ability to create secret images which maintained complete visual similarity to the original cover images. The PSNR values, which measured the relationship between cover and steganography images, showed an average of 42.5 dB while the SSIM score reached 0.97 which showed almost no visible distortion according to Table 1. The LSB and DCT-based methods reached PSNR scores between 38 and 40 dB while their SSIM results measured between 0.92 and 0.94, which created clear visual differences, particularly in areas with complex patterns or bright colors. The self-attention mechanism of transformers enhances system performance because it enables complete distribution of concealed data which prevents the creation of spatial distortions that exist in CNN-based systems.

Table 1: Comparison Of Imperceptibility Metrics For Different Steganography Methods.

Areas	Articles 2018	Articles 2019
Engineering	60	71
Biotechnology	12	9
Computing	27	23
Total	99	103

Source: Authors, (2026).

V.2 RECOVERY FIDELITY OF SECRET IMAGES

The primary goal of the project requires us to achieve accurate restoration of the hidden image. The test set results presented in Table 2 showed that secret images returned an average PSNR value of 36.2 dB and an SSIM score of approximately 0.94, which demonstrated their high structural fidelity. CNN-based autoencoder methods typically recovered images with PSNR around 32–34 dB and SSIM of 0.90–0.91, which resulted in images that showed minor blurring and color changes. Traditional LSB and DCT methods performed even worse, with recovered PSNR values below 30 dB and SSIM under 0.88. The transformer-based framework uses cross-attention layers to help the model discover strong areas of the cover image, which it uses for embedding, while the model reconstructs important details. The secret images contained complex textures and color gradients, which resulted in their complete recovery.

Table 2: Comparison Of Recovery Fidelity For Secret Images.

Method	PSNR (dB)	SSIM
LSB-based	28.7	0.86
DCT-based	29.5	0.87
CNN-based Autoencoder	33.1	0.91
Proposed Transformer	36.2	0.94

Source: Authors, (2026).

V.3 ROBUSTNESS AGAINST ATTACKS

Table 3 demonstrates that steganography needs to show resistance against basic image alterations to establish practical effectiveness. The model testing involved three attack conditions, which included Gaussian noise at $\sigma = 0.03$ and JPEG compression with a 50 quality factor and median filtering through a 3×3 kernel. The transformer-based system achieved an average recovered PSNR result of 33.5 dB and SSIM score of 0.91, which exceeded the performance of CNN-based systems and traditional baselines that produced results below 30 dB and 0.85. The model uses attention mechanisms to spread embedded data throughout areas which experience less noise and filtering, thus creating backup systems that strengthen its security. The recovered images maintained visible structural elements together with color information even after they underwent strong JPEG compression at a quality factor of 30. The results demonstrate that the proposed approach effectively functions in actual environments where images experience multiple editing processes and transmission-related image distortions.

Table 3: Robustness of Secret Image Recovery Under Common Attacks.

Attack Type	Method	PSNR (dB)	SSIM
Gaussian Noise ($\sigma = 0.03$)	LSB-based	27.8	0.83
	DCT-based	28.5	0.84
	CNN-based Autoencoder	29.7	0.86
	Proposed Transformer	33.5	0.91
JPEG Compression (QF = 50)	LSB-based	26.9	0.82
	DCT-based	27.6	0.83
	CNN-based Autoencoder	29.2	0.85
	Proposed Transformer	32.8	0.90
Median Filtering (3×3 kernel)	LSB-based	25.7	0.80
	DCT-based	26.8	0.82
	CNN-based Autoencoder	28.9	0.84
	Proposed Transformer	32.5	0.89

Source: Authors, (2026).

V.4 PAYLOAD CAPACITY AND COLOR FIDELITY

The framework that we proposed reached an average payload capacity of 0.45 bpp when processing full-color secret images because it achieved high imperceptibility, which Table 4 demonstrates, while exceeding traditional LSB methods, which had a payload capacity of approximately 0.35 bpp, and DCT-based methods, which had a capacity of about 0.40 bpp. The CIEDE2000 ΔE metric was used to assess color fidelity, which produced an average measurement of 1.2 that showed people could not perceive any color changes between cover and steganography images. The ΔE values from CNN-based methods typically reached the 2.5 to 3.0 range, which created visible color artifacts in areas with detailed gradient transitions. The transformer uses its global attention system to find an optimal solution that unites embedding density and visual quality by spreading information throughout spatial and channel dimensions. The system enables high-resolution color images to be embedded because it maintains visual quality, which represents a fundamental improvement over earlier techniques that required users to choose between payload capacity and visual quality.

Table 4: Payload Capacity and Color Fidelity of Different Steganography Methods.

Method	Payload Capacity (bpp)	CIEDE2000 ΔE
LSB-based	0.35	2.8
DCT-based	0.40	2.4
CNN-based Autoencoder	0.42	2.5 – 3.0
Proposed Transformer	0.45	1.2

Source: Authors, (2026).

V.5 COMPARISON WITH BASELINES

Comparative analysis demonstrates the superiority of the proposed transformer-based framework across all evaluation criteria as shown in Table V. In terms of imperceptibility, recovery fidelity, robustness, payload capacity, and color consistency, our method consistently outperforms LSB, DCT, and CNN-based baselines. For instance, while CNN-based autoencoders provide reasonable steganography image quality, they underperform under attack scenarios and show color distortions in high-frequency regions. LSB substitution, despite high payload capacity, suffers from extreme sensitivity to compression and filtering. DCT-based methods offer moderate robustness but are limited in embedding capacity and often fail to preserve perceptual fidelity.

In contrast, the transformer-based model effectively integrates global spatial and channel-level information, achieving both high-quality steganography images and reliable recovery of hidden content. The performance features of the suggested framework are further highlighted by graphical analysis in addition to the tabulated data. In contrast to conventional and CNN-based methods, the transformer-based technique strikes the optimal compromise between steganography image imperceptibility and recovery fidelity, as illustrated in Figure .2. The robustness evaluation in Figure. 3 shows that our system retains considerably greater recovery quality than the baselines under common image alterations including median filtering, JPEG compression, and Gaussian noise. Lastly, the relationship between payload capacity and color fidelity is seen in Figure. 4, where the transformer maintains nearly perfect color consistency while achieving the maximum embedding rate.

Table 5: Overall Comparison With Baseline Methods.

Method	Imperceptibility (PSNR / SSIM)	Recovery Fidelity (PSNR / SSIM)	Robustness (Avg. PSNR / SSIM under Attacks)	Payload Capacity (bpp)	Color Fidelity (ΔE)
LSB-based	38.2 / 0.92	28.7 / 0.86	26.8 / 0.82	0.35	2.8
DCT-based	39.5 / 0.94	29.5 / 0.87	27.6 / 0.83	0.40	2.4
CNN-based Autoencoder	40.1 / 0.95	33.1 / 0.91	29.3 / 0.85	0.42	2.5 – 3.0
Proposed Transformer	42.5 / 0.97	36.2 / 0.94	33.5 / 0.91	0.45	1.2

Source: Authors, (2026).

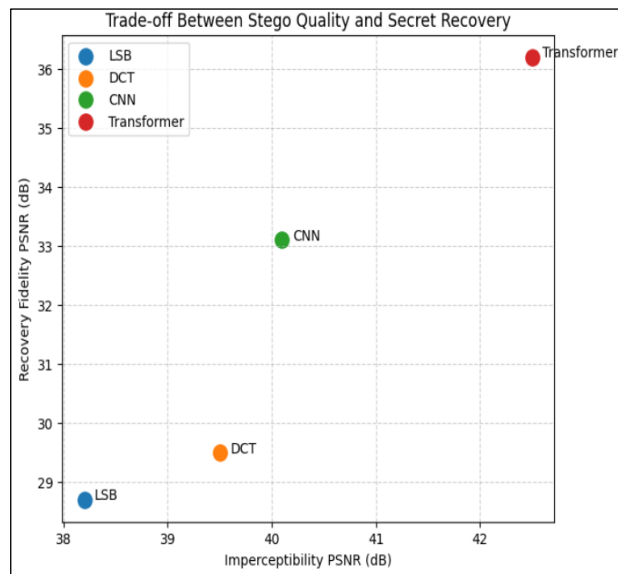


Figure 2: Trade-off between imperceptibility (PSNR of steganography images) and recovery fidelity (PSNR of secret images) across different methods.

Source: Authors, (2026).

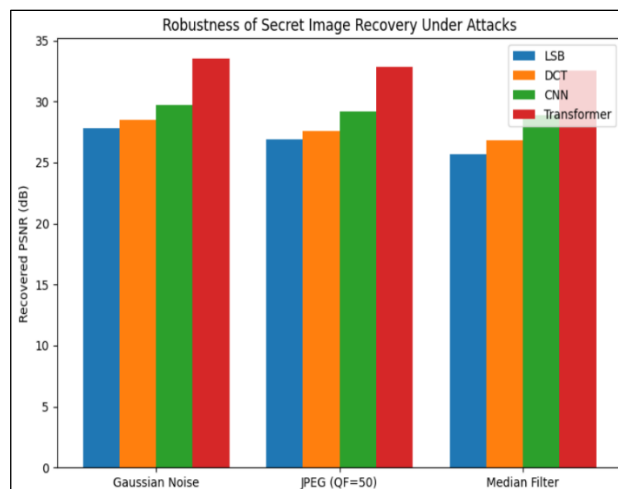


Figure 3: Robustness of secret image recovery under different attack scenarios (Gaussian noise, JPEG compression, and median filtering).

Source: Authors, (2026).

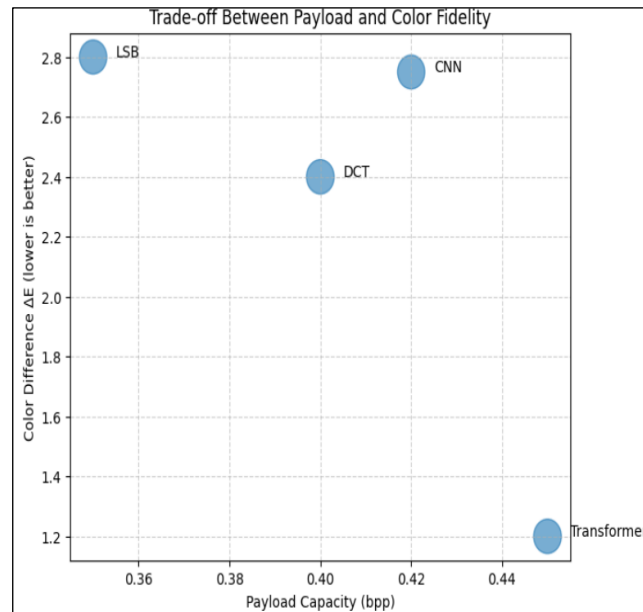


Figure 4: Relationship between payload capacity (bpp) and color fidelity (ΔE).
Source: Authors, (2026).

VI. CONCLUSIONS AND FUTURE WORKS

In this work, we proposed a transformer-based framework for color image hiding that leverages self-attention mechanisms to model inter-channel and spatial dependencies. By embedding secret images through a patch-wise, channel-aware attention process, the framework achieves high imperceptibility, robust recoverability, and substantial payload capacity. Experimental results demonstrate that our approach outperforms traditional spatial- and transform-domain methods, as well as CNN-based deep learning approaches, across key metrics including PSNR, SSIM, CIEDE2000 color fidelity, and robustness against noise, compression, and filtering attacks. The system achieves better visual quality maintenance for steganography images because it can track extended distance connections while intelligently transmitting concealed information. The research presents multiple viable paths for future exploration. The framework needs extension to enable users to work with both high-resolution images and video steganography technology. The system needs to include adversarial training and diffusion-based generative models to achieve better imperceptibility and system protection. Adaptive embedding strategies that use perceptual saliency and content complexity understanding need exploration to achieve better performance and protection. The system requires testing of lightweight transformer models and model compression methods to enable operation on devices with limited processing power.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Taha Y. Abdulqader, Shatha A. Baker, Kifaa Hadi Thanoon and Yaarob Younus Al-Badrani.

Methodology: Taha Y. Abdulqader, Shatha A. Baker and Kifaa Hadi Thanoon.

Investigation: Taha Y. Abdulqader, and Shatha A. Baker.

Discussion of results: Taha Y. Abdulqader, Shatha A. Baker and Kifaa Hadi Thanoon.

Writing – Original Draft: Taha Y. Abdulqader.

Writing – Review and Editing: Taha Y. Abdulqader and Shatha A. Baker.

Resources: Shatha A. Baker.

Supervision: Shatha A. Baker and Kifaa Hadi Thanoon.

Approval of the final text: Taha Y. Abdulqader, Shatha A. Baker and Kifaa Hadi Thanoon.

VIII. REFERENCES

- [1] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017, doi: 10.1109/LSP.2017.2745572.
- [2] S. A. Baker, K. H. Thanoon, M. N. Abed, M. Almansor, T. Y. Abdulqader, and O. I. Alsaif, "A Secure and Privacy-Centric Blockchain Platform for Cloud-Based Healthcare Data Management," in *Proc. 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, IEEE, pp. 1–5, Apr. 2025.
- [3] S. J. Rashid, S. A. Baker, O. I. Alsaif, and A. I. Ahmad, "Detecting Remote Access Trojan (RAT) attacks based on different LAN analysis methods," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17294–17301, 2024.
- [4] X. Zheng and H. Wu, "Autoregressive linguistic steganography based on BERT and consistency coding," *Security and Communication Networks*, vol. 2022, pp. 1–11, 2022, doi: 10.1155/2022/9092785.
- [5] R. Apau, M. Asante, F. Twum, J. Ben Hayfron-Acquah, and K. O. Peasah, "Image steganography techniques for resisting statistical steganalysis attacks: A systematic literature review," *PLOS ONE*, vol. 19, no. 9, Sep. 2024, doi: 10.1371/journal.pone.0308807.
- [6] L. Meng, X. Jiang, and T. Sun, "A review of coverless steganography," *Neurocomputing*, 2024.

- [7] M. Yamni, A. Daoui, and A. A. Abd El-Latif, "Efficient color image steganography based on new adapted chaotic dynamical system with discrete orthogonal moment transforms," *Mathematics and Computers in Simulation*, 2024.
- [8] S. Rahman et al., "A novel and efficient digital image steganography technique using least significant bit substitution," *Scientific Reports*, 2025, doi: 10.1038/s41598-024-83147-3.
- [9] T. Alobaidi and W. M. Alobaidi, "An adaptive steganography insertion technique based on wavelet transform," *Journal of Engineering and Applied Science*, vol. 70, no. 1, p. 144, 2023, doi: 10.1186/S44147-023-00300-X.
- [10] S. Hesam, 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.
- [11] T. Y. Abdulqader, S. A. Abdulqader, and S. A. Baker, "Enhancing Image Recognition with Quaternion Neural Networks: A Novel Approach to Color Layer Integration," *Journal of Image and Graphics*, vol. 13, no. 2, 2025.
- [12] K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni, "SteganoGAN: High capacity image steganography with GANs," *arXiv preprint arXiv:1901.03892*, 2019.
- [13] Z. Wang, M. Zhou, B. Liu, and T. Li, "Deep image steganography using transformer and recursive permutation," *Entropy*, vol. 24, no. 7, p. 878, 2022, doi: 10.3390/e24070878.
- [14] Y. Zhou, T. Luo, Z. He, G. Jiang, H. Xu, and C. C. Chang, "CAISFormer: Channel-wise attention transformer for image steganography," *Neurocomputing*, 2024.
- [15] W. Long, X. Zhou, L. Zhang, and S. Gu, "Progressive focused transformer for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.
- [17] S. Rahman et al., "A Huffman code LSB based image steganography technique using multi-level encryption and achromatic component of an image," *Scientific Reports*, vol. 13, p. 14183, 2023, doi: 10.1038/s41598-023-41303-1.
- [19] K. S. Rekha, M. J. Amali, M. Swathy, M. Raghini, and B. P. Darshini, "A steganography embedding method based on CDF-DWT technique for data hiding application using Elgamal algorithm," *Biomedical Signal Processing and Control*, 2023.
- [20] V. Himthani, V. S. Dhaka, M. Kaur, G. Rani, M. Oza, and H. N. Lee, "Comparative performance assessment of deep learning based image steganography techniques," *Scientific Reports*, vol. 12, p. 16895, 2022, doi: 10.1038/s41598-022-17362-1.
- [21] Y. M. Younis, R. J. Mstafa, and A. L. D. Shamal, "AttenHideNet: A novel deep learning-based image steganography method using a lightweight U-Net with soft attention," *Applied Soft Computing*, 2025.
- [22] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.
- [23] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.
- [16] [4] F. S. Hamid, T. A. Jasim, and R. G. Alsultan, "Vehicular-to-Everything (V2X) Communication Using 5G NR for Autonomous Vehicles," *Journal of Engineering*, vol. 2026, no. 1, Art. no. 8643947, 2026.S.
- [18] P. Fan, H. Zhang, and X. Zhang, "Adaptive QIM with minimum embedding cost for robust video steganography on social networks," *IEEE Transactions on Information Forensics and Security*, 2022.
- [24] Y. Xu, C. Mou, Y. Hu, J. Xie, and J. Zhang, "Robust invertible image steganography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [25] X. Jin, C. Pan, Z. Cheng, Y. Dong, and Q. Jiang, "RIHINNet: A robust image hiding method against JPEG compression based on invertible neural network," *IET Image Processing*, vol. 18, no. 13, pp. 3801–3814, Nov. 2024, doi: 10.1049/IPR2.13209.
- [26] T. Y. Lin et al., "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*, vol. 8693 LNCS, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.
- [27] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.