© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717350

StegaCBAM-Net: An Attention-Driven Deep Steganographic Encoder-Decoder

NAGALAKSHMI AVULA[1], SIVA NANDINI BOMMUPALLA[2], LAHARI GORIJAVOLU[3], DR. SELI MOHAPATRA[4]

[1,2,3] *Department of Computer Science and Engineering, R.V.R & J.C College of Engineering*

[4] *Associate Professor, Department of Computer Science and Engineering, R.V.R & J.C College of Engineering*

Abstract*— The method of image steganography allows one to communicate secretly via encoding confidential information into harmless-looking images without compromising the visual appearance of the latter. The current approaches to steganography like LSB embedding and transform domain embedding provide low computational complexity, yet they lack robustness, have limitations in terms of payload, and are prone to steganalysis [1], [2]. Steganography using deep learning networks like CNNs has led to improvements in terms of both capacity and concealment quality as they can learn the embedding techniques through training on datasets [3], [4]. Yet most of the existing techniques use shallow features extraction and pay insufficient attention to spatial/channel dependencies. The proposed paper introduces an Attention-Based Image Steganography Model (AEDISA) to increase imperceptibility, payload, and robustness in image steganography systems. This framework uses an end-to-end learning scheme consisting of an encoder-decoder model using three neural networks, including a preparation network, hiding network, and revelation network. In addition, the model uses multi-scale convolution (3×3, 4×4, and 5×5) and a novel spatial-channel attention mechanism based on the CBAM approach for efficient learning of features. The model is trained end-to-end using the ImageNet100 benchmark dataset. Our experiment shows promising results compared to the state-of-the-art CNN steganographic approaches by obtaining 79.74 dB PSNR, 0.9703 SSIM, and higher reconstruction accuracy.

***Index Terms*— Image Steganography, Deep Learning, CNN, Attention-Based Approach, CBAM.**

I. INTRODUCTION

The accelerated development of digital communications requires more secure information transmission. Although cryptography ensures the security of information, it cannot provide information hiding. Steganography is designed to solve this problem through embedding secret data into cover objects, thereby hiding the fact of communication [1]. For embedding media, images are suitable for high embed-ding capacities and redundancy. Classic algorithms such as Least Significant Bit replacement, Pixel Value Difference, and transform domain techniques have proven to be practically feasible. However, most of these methods cause noticeable distortions or have insufficient payload robustness [1], [6].

Recently, deep learning has greatly promoted the development of steganography. Baluja was the first to prove the effective-ness of full image embedding through a deep convolutional encoder-decoder architecture [3]. Later, researchers used the U-net structure, generative adversarial networks, and invertible neural networks to optimize reconstruction quality and security [4], [7], [8], [9].

However, there are still some limitations such as adaptive feature selection limitations, Poor channel and spatial dependencies modeling, Robustness issues in texture regions, Susceptibility to learned steganalyzers. In order to overcome these issues, this paper suggests a deep steganography frame-work based on multi-scale convolution with spatial-channel attention.

II. RELATED WORK

Steganography algorithms for images have been developed from conventional signal processing techniques to more sophisticated deep learning models. The available literature can be classified into classical steganography algorithms, CNN-based

steganography algorithms, and deep learning algorithms with attention mechanisms.

A. Traditional Image Steganography Approaches

Classical approaches in image steganography mainly involve spatial domain and transformation domain embedding approaches. These approaches include LSB-based substitution which is one of the early approaches due to ease of implementation and computation [1], [2],[31]. Other approaches like Pixel Value Differencing (PVD), DCT, and DWT have enhanced embedding performance and concealment capabilities [3],[29],[30]. However, these traditional approaches are associated with limitations such as having low payload capacity, lack of robustness against attacks, and being vulnerable to statistical steganalysis.

While conventional approaches have performed adequately in ensuring good quality images, their weakness lies in hand-crafted and rigid statistical modeling.

B. Deep Learning -Based Image Steganography

A major milestone in advancing the field was the application of deep learning for end-to-end training of both embedding and decoding processes through the use of encoder-decoder networks [28]. The seminal work of Baluja [4] showed how full secret images can be embedded into cover images while keeping high visual resemblance to their respective covers. The next step made by several researchers utilized deeper CNNs [27] in their approaches. For instance, Duan et al. [5] developed SteganoCNN in an effort to increase generalization ability and payload capacity.

U-Net-based reversible image steganography [6] introduced skip connections to ensure higher quality of reconstruction. Adversarial learning techniques adopted GAN-based goals for better stego-image quality and robustness against steganalyzers [7], [10],[29].

Invertible neural network models helped enhance losslessness and embedding capacity [8], and private key-guided multi-image steganography ensured higher embedding security [9]. Nevertheless, CNN-based approaches continue to suffer from their reliance on sequential convolution operations, which fail to address long dependencies and salient areas.

C. Feature Learning through Attention for Steganography

Advancements in visual attention systems have led to their application in secure image embedding processes. The use of attention layers helps in better representation learning by focusing on relevant spatial and channel information.

This Woo et al. [11] introduced the Convolutional Block Attention Module (CBAM), which includes both channel and spatial attention improvements, with notable results in computer vision applications. Similarly, residual learning methods [12] and transformer-based attention systems [13] have further validated the benefits of adaptive feature modeling in complex images.

Building upon these advancements, recent steganography techniques have included attention layers to help with embed-ding localization and enhance secret recovery operations [24]. But the current implementations of attention-based steganography rely mainly on single attention layers, while multi-scale attention is yet to be explored in steganography research.

III. DATASET AND PREPROCESSING

The proposed model is evaluated using subsets of the ImageNet100 dataset, which consists of diverse natural images suitable for learning robust feature representations in deep steganography tasks.

The dataset is partitioned into four subsets for systematic experimentation: 15,000 images are used for training, 2,000 images for validation, 1,000 images for testing, and an additional 1,000 images for qualitative visualization.

All images are preprocessed to ensure consistency and computational efficiency. Each image is resized to a fixed resolution of 64 X 64 pixels to match the input requirements of the network. Furthermore, images are normalized using standard ImageNet statistics to stabilize training and improve convergence.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717350

To enhance generalization and reduce over fitting, data augmentation techniques are applied during training. These include random cropping and resizing operations, followed by normalization. Such augmentations enable the model to learn more invariant and robust feature representations across varying image distributions.

IV. METHODOLOGY

A. Proposed Attention-Enhanced Deep Image Steganography framework

The proposed Attention-based Deep Image Steganography System (AEDISA) adopts an end-to-end encoder-decoder framework that is capable of embedding a secret image into a cover image without degrading the visual quality while ensuring accuracy during secret extraction.

Let S be the secret image and C be the cover image. Then, the goal is to create a stego image C'similar to C while making it possible to retrieve the secret image S through the decoder $S\hat{}$ as shown below:

$$C' = E\ (S, C)$$

$$S\hat{} = D(C')$$

where $E\ (.)$ refers to encoder (embedding network), $D(.)$ refers to decoder (reveal network).

Design goals include two aspects:

- Minimizing distortion caused by cover image to stego image.
- Maximizing the accuracy of secret image extraction.

Proposed method structure includes: (1) Preparation Network; (2) Hiding Network; (3) Reveal Network; (4) Parallel Spatial-Channels Attention Mechanisms. Fig.1 shows the proposed model's architecture.

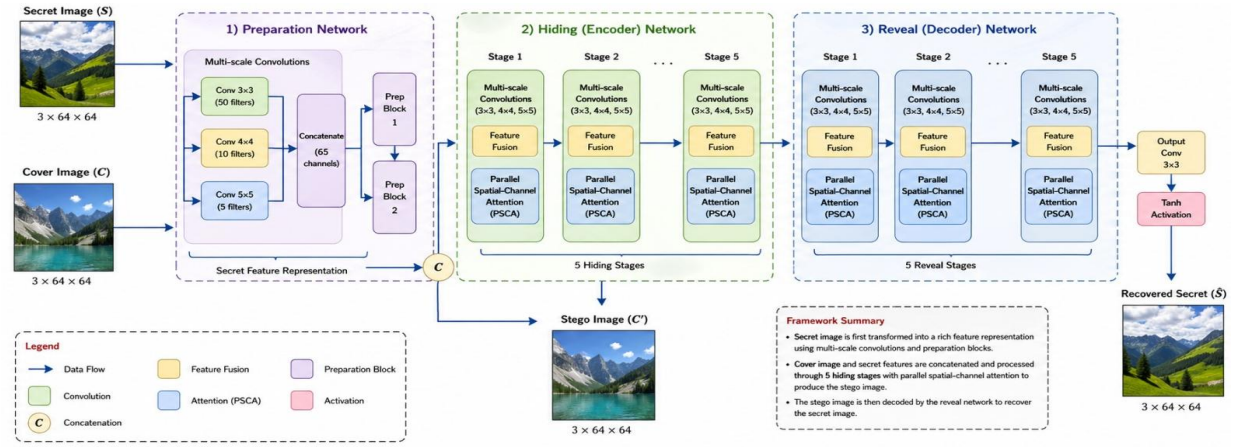

Fig. 1: Proposed architecture of StegaCBAM-Net.

B. Preparation Network

Prior to embedding, the secret image is subjected to hierarchical feature extraction using a preparation network. For the purpose of extracting features from various receptive fields, three separate convolution layers are employed:

- 50 filters of size 3x3
- 10 filters of size 4x4
- 5 filters of size 5x5

Two stacked preparation blocks are used for converting the pixel domain secret message to higher dimensional covert representation. Multiscale convolution helps in the extraction of both fine textures and edge structure at multiple scales, increasing the effectiveness of embedding.

C. Hiding Network

The hiding network has covert elements within the cover media. The fused feature map then goes through five hiding stages hierarchically. The design involves multi-scale convolution with feature fusion and attention-based refinement. Fig.2 shows a hiding stage.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717350

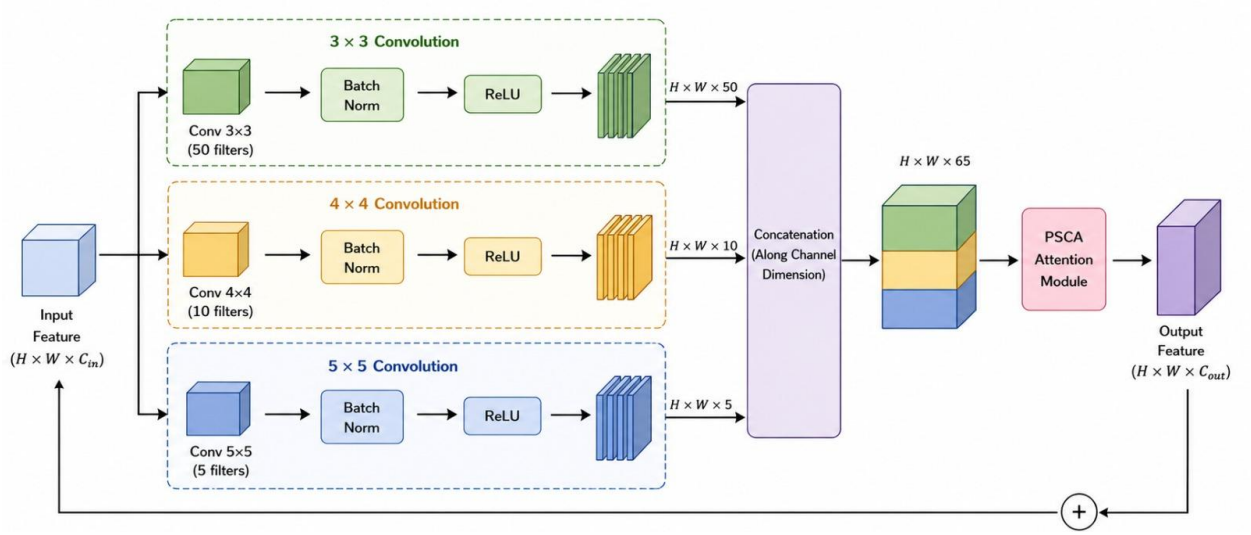

Fig. 2: Multi-scale hiding block structure used in the proposed framework.

D. Parallel Spatial-Channel Attention Module

An important aspect of the present study includes the de-sign of Parallel Spatial-Channel Attention (PSCA) technique, which is employed across the entire process of encoding and decoding stages. Existing CNN-based steganography approaches apply equal treatment to all feature maps, leading to the possible existence of ineffective embedding regions. This designed module is intended to enhance the relevant portions.

a) Channel Attention:

Channel attention learns important embedding channels:

$Mc(F)=\sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$

Refined output:

$$Fc=Mc(F)\odot F$$

where σ is sigmoid activation

$\odot$ is denotes element-wise multiplication

b) Spatial Attention:

Spatial attention emphasizes significant embedding regions:

$Ms(F)=\sigma(f^{7\times7}([AvgPool(F);MaxPool(F)]))$

Refined spatial features:

$$Fs=Ms(F)\odot F$$

c) Parallel Attention fusion:

Here both attentions are computed in parallel:

$$Fatt=Fc+Fs$$

This improves simultaneous modeling of spatial textures and channel dependencies. The attention module is shown as fig.3.

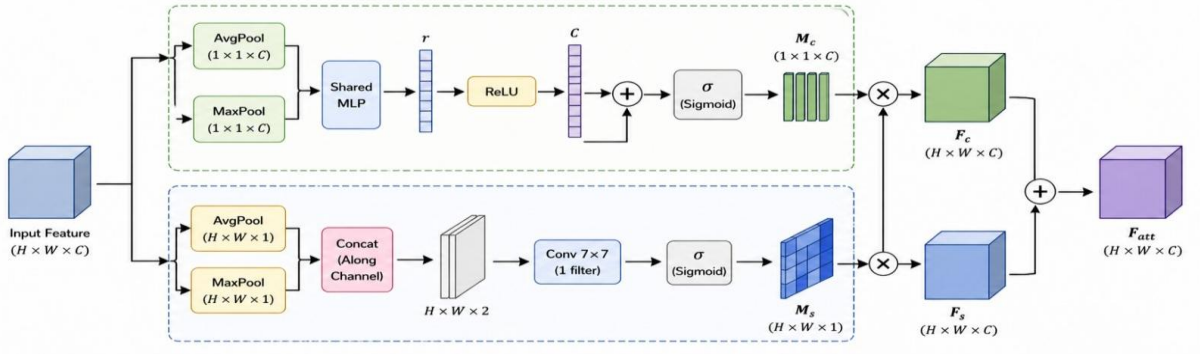

Fig. 3: Parallel Spatial-Channel Attention Module

E. Reveal Network

The reveal network recovers the original hidden image from the stego image created by the generator network. It follows an architecture similar to that of the encoder, which includes five reconstruction blocks based on convolution, Multi-scale feature extraction,

Attention refinement.
Recovered secret image:

$$S\hat{}=D(C')$$

Final output uses Tanh activation:

$$S\hat{}=\tanh(Wx+b)$$

to preserve normalized pixel values.

F. Training Procedure

The proposed framework is trained using an end-to-end optimization strategy to jointly learn the embedding and reconstruction processes. The model is trained for 500 epochs with a batch size of 64 to ensure sufficient convergence. To stabilize the initial training phase and avoid abrupt gradient updates, a warm-up strategy is employed for the first 75 epochs. The Adam optimizer is utilized for parameter updates due to its adaptive learning rate mechanism and effectiveness in training deep neural networks. All computations are per-formed using GPU acceleration to improve training efficiency and reduce execution time. The detailed optimization procedure followed during training is summarized in Algorithm 1.

Algorithm 1:
Input: Secret image S, Cover image C
Output: Trained encoder E and decoder D

1) Initialize encoder network E
2) Initialize decoder network D
3) for epoch = 1 to N do
4) for each batch (S, C) do
5) Generate stego image:
   $C' = E(S, C)$
6) Recover secret image:
   $\hat{S} = D(C')$
7) Compute cover loss:
   $Lc = \|C - C'\|^2$
8) Compute secret reconstruction loss:
   $Ls = \|S - \hat{S}\|^2$
9) Compute total loss:
   $Ltotal = Lc + \beta Ls$
10) Backpropagate gradients
11) Update parameters of E and D
12) end for
13) end for
14) Return trained models E and D

G. Computational Complexity

The time complexity of the proposed framework mainly consists of the convolutional process. Assuming that the size of the feature map is n, kernel size is k, and the number of channels is c, the time complexity of one layer in convolutional neural network could be calculated as $O(nk^2c)$.

The employment of multi-scale convolutional kernels increases the computation overhead for parallel processing; however, it provides better feature learning from various receptive fields. Moreover, the employment of the parallel spatial–channel attention module imposes a slight overhead on the computation overhead. Notwithstanding the increased computation overhead, the attention-based module effectively facilitates the representation of features, which is beneficial for embedding learning and feature reconstruction

VI. RESULTS AND DISCUSSION

The performance of the proposed Attention-Enhanced Deep Image Steganography approach is analyzed using traditional image quality measures such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

The PSNR value of the proposed approach is found to be 79.74 dB. The high PSNR values signify that the distortion caused by the embedding process is low and that the reconstruction is of high accuracy. Also, the SSIM value is 0.9703 respectively.

The comparative analysis between the performances of the conventional steganography technique and the proposed approach is shown in Table 1 based on previous work in [11], [12],[13],[14]. Conventional techniques like LSB are less imperceptible because of their simplicity, while techniques that use deep learning networks like CNN and U-net perform better. However, the proposed technique, i.e., attention-based CNN, performs even better, in terms of PSNR and SSIM.

Table 1: Comparative Study

Paper	Dataset	Method used	SSIM	PSNR
[31] Arya & Soni	Lena and Baboon	GAN	-	56.95

[6] Duan et al	ImageNet	GAN+UNET	0.964	40.66
[26] Al-Afandy et al.	RGB image	LSB + Image Cropping	-	62.53
[27] Chambial & Sood	Tiny ImageNet	CNN	0.965	76.214
Proposed Model	ImageNet 100	CNN+ Attention	0.9703	79.74

The enhancement in performance is due to the use of multi-scale convolutional feature extraction and spatial-channel attention operations. The multi-scale convolutional feature extraction helps the model to concentrate on the significant parts of the image and ignore the unnecessary parts of the cover image, thus increasing the embedding efficiency and minimizing the amount of distortion. This will help in concealing the secret data into less noticeable areas of the cover imagen conclusion, the model proposed above can be considered more efficient compared to traditional and current deep-learning based steganography approaches because it provides better imperceptibility and reconstruction accuracy along with feature representation capacity.

VII. CONCLUSION

This study demonstrated the implementation of a novel approach called Attention-Enhanced Deep Image Steganography which is expected to ensure good invisibility and reconstruction qualities. In this work, the authors have applied a combination of a deep CNN and multi-scale feature learning together with a spatial-channel attention mechanism to effectively transmit the information contained in secret images through the use of the corresponding cover images. For testing the performance of the suggested scheme, the authors used common objective image quality assessment criteria including the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). According to the obtained results, the PSNR and the SSIM values were 79.74 and 0.9703 respectively.

The comparative analysis presented in Table 1 reveals that conventional approaches for steganography, like those using LSB embedding, have lower imperceptibility and poor robustness properties. However, by employing CNN and U-Net architectures, the performance has been improved, but the proposed approach provides even better results due to the attention mechanism that helps focus on important features and increases the imperceptibility and robustness measured via PSNR and SSIM.

Using multi-scale convolution helps the network capture detailed and general features of the images, while applying an attention mechanism increases feature selection by eliminating non-useful information. Therefore, the proposed network can hide secret information in less perceptible parts of the images.

Future research might also address issues such transformer-based attention mechanisms, increasing defenses against adversarial attacks, and scaling up for use with higher resolution images in steganographic applications. Other areas of improvement could be optimizing the algorithm for real-time and resource-limited conditions, applying it to videos, and dealing with multiple images in steganography. Combining encryption and steganography can ensure better security in communication systems.

Future work will pursue Bayesian hyperparameter optimization, incorporation of airline-specific and route-specific features, post-hoc interpretability analysis using SHAP values, evaluation on public real-world airfare datasets, and extension to multi-task learning jointly predicting fare, seat availability, and cancellation probability.

REFERENCES

[1] O. Etzioni et al., "To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price," Proc. ACM SIGKDD, pp. 119-128, 2003.

[2] N. Provos and P. Honeyman, “Hide and seek: An introduction to steganography,” *IEEE Security & Privacy*, vol. 1, no. 3, pp. 32–44, May–Jun. 2003, doi: 10.1109/MSECP.2003.1203220.

[3] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[4] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 2069–2079.

[5] X. Duan, N. Liu, M. Gou, W. Wang, and C. Qin, "SteganoCNN: Image steganography with generalization ability based on convolutional neural network," *Entropy*, vol. 22, no. 10, Art. no. 1140, 2020, doi: 10.3390/e22101140.

[6] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang, and C. Qin, "Reversible image steganography scheme based on a U-Net structure," *IEEE Access*, vol. 7, pp. 9314–9323, 2019, doi: 10.1109/ACCESS.2018.2890443.

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

[8] S. Lu, R. Wang, T. Zhong, and P. Rosin, "Large-capacity image steganography based on invertible neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 10816–10825.

[9] H. Kweon, J. Park, S. Woo, and D. Cho, "Deep multi-image steganography with private keys," *Electronics*, vol. 10, no. 16, Art. no. 1906, 2021.

[10] Q. Li, X. Wang, X. Wang, B. Ma, C. Wang, and Y. Xian, "A novel grayscale image steganography scheme based on chaos encryption and generative adversarial networks," *IEEE Access*, vol. 8, pp. 168166–168176, 2020.

[11] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[13] A. Vaswani *et al*., "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 248–255.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.

[16] A. Paszke *et al*., "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 8024–8035.

[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[18] Y. Zou, G. Zhang, and L. Liu, "Research on image steganography analysis based on deep learning," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 266–275, 2019.

[19] W. You, H. Zhang, and X. Zhao, "A Siamese CNN for image steganalysis," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 291–306, 2021.

[20] Z. Xiang, J. Sang, Q. Zhang, B. Cai, X. Xia, and W. Wu, "A new convolutional neural network-based steganalysis method for content-adaptive image steganography," *IEEE Access*, vol. 8, pp. 47013–47020, 2020.

[21] H. Kato, K. Osuge, S. Haruta, and I. Sasase, "A preprocessing by using multiple steganography for intentional image downsampling on CNN-based steganalysis," *IEEE Access*, vol. 8, pp. 195578–195593, 2020.

[22] X. Duan, D. Guo, N. Liu, B. Li, M. Gou, and C. Qin, "A new high-capacity image steganography method combined with image elliptic curve cryptography and deep neural network," *IEEE Access*, vol. 8, pp. 25777–25788, 2020.

[23] C. Qin, P. Ji, C.-C. Chang, X. Zhang, and J. Dong, "Deep learning in reversible steganography and watermarking: A survey," *Information Fusion*, vol. 76, pp. 187–204, 2021.

[24] A. Rehman, R. Rahim, and S. Kadry, "Secure image steganography using deep convolutional neural networks with attention mechanisms," *IEEE Access*, vol. 10, pp. 45812–45828, 2022.

[25] B. Bashir and A. Selwal, "Towards deep learning-based image steganalysis: Practices and open research issues," *IEEE Access*, vol. 10, pp. 11451–11478, 2022.

[26] K. A. Al-Afandy, O. S. Faragallah, A. Elmhalawy, E. S. M. El-Rabaie, and G. M. ElBanby, "High security data hiding using image cropping and LSB least significant bit steganography," in *Proc. 4th IEEE Int. Colloq. Inf. Sci. Technol. (CiSt)*, 2016, pp. 400–404.

[27] S. Chambial and D. Sood, "Image steganography using CNN," *International Research Journal of Engineering and Technology (IRJET)*, vol. 9, no. 2, pp. 755–762, Feb. 2022.

[28] N. K. Chahar, A. Dhaka, A. Nandal, and V. Kumar, "Deep learning-empowered image steganography: Architectural innovations and benchmarking," *International Journal / Online First Article*, 2025.

[29] K. R. Malik et al., "A hybrid steganography framework using DCT and GAN for secure data communication," *Scientific Reports*, vol. 15, 2025.

[30] B. K. Sahoo et al., "SteganoSNN: SNN-based steganography with encryption," *arXiv preprint arXiv:2511.06573*, 2025.

[31] A. Arya and S. Soni, "Performance evaluation of secret image steganography techniques using least significant bit (LSB) method," *International Journal of Computer Science Trends and Technology*, vol. 6, no. 2, pp. 160–165, 2018.