Information Security Using Image Steganography and Deep Learning

Naman Bhushan[1] and Upendra Kumar [2]

[1] Centre for Advanced Studies, Dr. A.P.J Abdul Kalam Technical University, Lucknow Uttar Pradesh 226031, India

[2] Department of Computer Science and Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh 226021, India

[2] Faculty of Engineering and Technology, Dr. A.P.J Abdul Kalam Technical University, Lucknow Uttar Pradesh 226031, India

Abstract

The method of hiding information in digital images in a clandestine manner has gained more significance for protecting information in modern-day communication systems. This research proposes a new method for hiding information in images, focusing on higher security, improved storage efficiency, and less disturbance to the accepted image quality measures, such as peak signal-to-noise ratio, structural similarity metrics. The outcomes reveal that the proposed system consistently surpasses traditional techniques in both performance and security. This research contributes a dependable and innovative solution to the domain of covert communication and serves as a strong foundation for advancing future work in secure image steganography.

Keywords: - Steganography, Hiding Information, Image Quality, Security, Digital Images

1.Introduction

As digital communication increasingly pervades our daily lives, the necessity of safeguarding confidential information in transit has expanded multifold. Encryption has been an old guard for data security for years, yet its very visible transformation of information into ciphertext unwittingly betrays the existence of confidential material and makes it a target for interception. In contrast, steganography presents a more discreet solution by embedding secret data within ordinary digital files—particularly images—where the presence of hidden content is visually undetectable.

Images are especially well-suited for steganographic use due to their widespread presence and substantial data-carrying capacity. But traditional approaches generally struggle to compromise on security, capacity embedding, and image fidelity. Such limitations make them weaker, particularly for high-security use cases. To address such concerns, advances in deep learning of computational intelligence have opened doors to more robust steganographic methods. Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN) techniques have proven the potential to securely and strongly embed and extract information, reducing detection risk while maintaining image quality.

This paper presents an improved steganographic approach that emphasizes confidentiality, capacity, and imperceptibility. Moreover, it also offers an organized survey of contemporary image steganography techniques, classified under classical techniques, CNN-based models, and GAN-based models. Through the evaluation of their mechanism, benefits, drawbacks, and practical usability, this work gives a rich platform for exploring future research and development in secure, hidden communication systems.

2. Background

Steganography based on images developed from earlier methods such as Least Significant Bit (LSB) substitution and Pixel Value Differencing (PVD), although being rudimentary and efficient, having discernible data capacity shortcomings and obvious observability. An imperative to alleviate these shortfalls led to employing mathematically established techniques. Sudoku-based methods bring in randomness in embedding data, whereas magic squares provide organized, well-balanced patterns where data can be hidden in a subtle manner. While each approach holds promise in isolation, their application together particularly in higher-order arrangements is underdeveloped and potentially more secure and adaptive steganographic systems.

Deep learning has transformed the field more recently. Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) enable improved data embedding with low visual distortion, making it harder to discover hidden content. These technologies have enabled steganography to be more feasible for use in practical real-world applications such as secure messaging and digital watermarking, especially when mathematical approaches are combined with AI-based models.

3. Objective

The present paper seeks to thoroughly examine state-of-the-art image steganography methods for their efficacy in secure data hiding. We systematically contrast conventional schemes with state-of-the-art deep learning-based strategies (CNNs and GANs) in terms of three critical parameters: embedding capacity, visual imperceptibility, and detection resistance. The study aims to establish optimal methods of fulfilling these contradictory demands with the overcoming of current limitations in practical use. Besides, we propose modifications of existing methods for their increased robustness and practical usability for secure digital communication and thus the development of more reliable steganographic solutions for modern security challenges.

4. Research Gap

Current steganographic techniques are faced with ominous security and resilience constraints despite mounting cybersecurity threats. While strong cryptography provides strong data protection, its combination with steganography is not strongly matured, and most operations rely on fundamental LSB embedding that is strongly vulnerable to modern steganalysis. The area does not have hybrid solutions that best integrate cryptographic security with imperceptible data hiding, plagued by low payload capacity without affecting imperceptibility and susceptibility to AI-based steganalysis tools and low adaptability across various media types and network conditions. Deep learning techniques in this area are still in their infancy, with the majority of research utilizing small datasets while overlooking real-world factors such as image compression and format changes. There exists an imperative demand for an adaptive platform that combines cryptographic techniques with clever embedding mechanisms for dynamic adaptation based on security demands without compromising the high capacity of data and robust detection resistance. Filling in these loopholes would dramatically promote secure communication mechanisms for secure transfer of sensitive data where both concealment and encryption are crucial, albeit crafting such robust steganography techniques supplemented with cryptography remains an ongoing open issue in cyber security research.

5. Related Work

Current research has indicated that conventional image steganography methods usually do not achieve a balance between payload capacity, visual quality, and detection resistance. Subramanian et al. [1] surveyed some of the current methods and pointed out that most traditional methods, including LSB and PVD, are becoming more susceptible to current steganalysis techniques. Early steganalysis techniques, including those presented by Fridrich and Kodovsky [2], established the foundation for pattern detection through statistical irregularities. Holub et al. [3] built upon this with universal distortion functions that allow for assessing embedding quality in various domains. Incorporating deep learning into steganalysis has tremendously enhanced detection power. Xu et al. [4] and Ye et al. [5] constructed convolutional neural network (CNN) structures which are structurally optimal for detecting concealed data in spatial domains. Zhang et al. [6] improved these models further through employing depth-wise separable convolutions and multi-level pooling, with greater efficiency in detecting stego-content. Yedroudj et al. [7], another notable contribution, proposed YedroudjNet—a fast but efficient CNN model specifically designed for spatial steganalysis. Fridrich's group then continued this research with Boroumand et al.

[8] writing a deep residual network that performs better than earlier models by learning more complicated image features. Bas et al. [9] also made a contribution to the field by running the BOSS challenge, providing a standardized test dataset and evaluation framework that encouraged the creation of more resilient models. Goodfellow et al. [10] changed the game with Generative Adversarial Networks (GANs) that not only provide a generative framework for creating realistic synthetic images but also serve as the basis for embedding methods that are imperceptible and more difficult to detect. These early studies demonstrate progress from manually designed feature-based detection to deep learning-based approaches and highlight the continued necessity of hybrid models that maintain image integrity while ensuring maximum embedding capacity and security.

6. Methodology

The methodology for this study is structured into several key stages designed to develop, embed, and evaluate steganographic techniques using deep learning models. The goal is to ensure secure, high-capacity, and imperceptible information hiding in images.

6.1. Data Collection

In this research, the collection of data for steganography is performed using open-source datasets like BOSSBase, ImageNet, and CelebA that contain a diverse range of images with various resolutions and types of content. The images are used as cover media to embed confidential data, which can be text, grayscale images, or color images. The images are preprocessed by resizing, normalization, and conversion of formats to be compatible with the deep learning models. Every cover-secret pair is preprocessed for embedding with CNN, GAN, and CycleGAN architectures. All the data are stored securely, anonymized, and arranged to facilitate model training, validation, and evaluation procedures.

6.2. Data Pre-Processing

All cover images are subject to standardized preprocessing to allow consistent assessment of steganographic methods. Images are down sampled and converted to lossless PNG, then resized to 128×128 pixels by Lanczos interpolation, weighing computation efficiency against preservation of features. Pixel values for CNN/GAN-based methods are normalized to [-1,1] range for compatibility with usual deep learning routines. Original aspect ratios are preserved in the dataset during resizing to avoid distortion artifacts. We form well-balanced partitions (70% train, 15% validation, 15% test) maintaining equal distribution of various embedding techniques (LSB, PVD, DCT, neural networks). All stego-cover pairs are recorded with full metadata such as algorithm parameters, payload capacity (0.1-0.4 bpp), and detection resist measurements for enabling comprehensive comparison of the techniques.

6.3. Model Selection

For steganalysis, we selected a specialized CNN architecture with attention to detecting hidden patterns within images. The network has 5 blocks of convolutional layers (Conv2D + ReLU + BatchNorm) with growing filters (32-256), max-pooling layers. Two dense layers (512, 256 units) with dropout (0.5) are responsible for feature extraction prior to feeding it to the classification head. We use the Adam optimizer (lr=0.001) with binary cross-entropy loss, since our task involves discrimination between stego and clean images. Performance is measured in terms of detection accuracy, precision, recall, and F1-score, with special interest in reducing false negatives in security-sensitive steganalysis tasks. The architecture aims to balance computational efficacy with detection sensitivity for different embedding schemes.

7. System Architecture

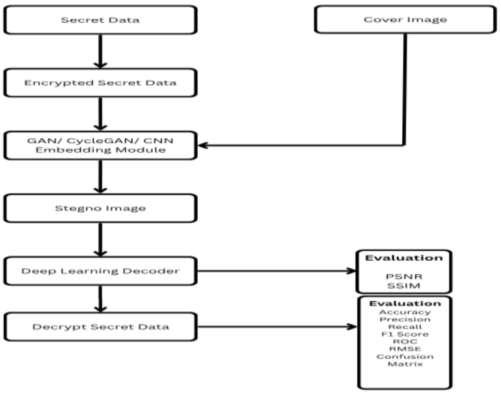

Figure 1. Block Diagram of the Proposed Deep Learning-Based Steganography System

8. Results And Discussions

Training & Testing Ratio	Model Used	Dataset Used	Accuracy	Precision	Recall	F-1 Score	Root Mean Square Error
80:20	Convolution Neural Network	Bossbase	0.9717	0.9996	0.8727	0.9318	0.0219
70:30			0.9817	0.9995	0.9168	0.9564	0.0160
60:40			0.9097	0.9084	0.9095	0.9090	0.0631
80:20		DIV2K	0.9606	0.9612	0.9599	0.9605	0.0487
70:30			0.9957	0.9940	0.9919	0.9929	0.0075
60:40			0.9957	0.9939	0.9916	0.9927	0.0074
Average			0.9691	0.9762	0.9479	0.9572	0.0274

Table 1. Performance Metrics of CNN on Bossbase and DIV2K Datasets with Varying Train-Test Splits

Training & Testing Ratio	Model Used	Dataset Used	Peak Signal to Noise Ratio (PSNR)	Structural Similarity Index Measure (SSIM)
80:20	Generative Adversarial Network	BOSSBASE	34.89	0.9957
70:30			36.23	0.9961
60:40			35.02	0.9967
80:20		CELEBA	25.02	0.9336
70:30			30.85	0.9857
60:40			30.13	0.9847
Average			32.02	0.9820

Table 2. Evaluation of GAN Performance on BOSSBASE and CELEBA Datasets Using PSNR and SSIM Across Different Train-Test Splits

Training & Testing Ratio	Model Used	Dataset Used	Peak Signal to Noise Ratio (PSNR)	Structural Similarity Index Measure (SSIM)
80:20	Cycle-Consistent Generative Adversarial Network	Horse 2 Zebra	75.25	0.8497
70:30			71.50	0.0105
60:40			58.65	0.0083
80:20		CELEBHQ	79.15	0.7550
70:30			65.94	0.8810
60:40			45.83	0.9015
Average			66.05	0.5685

Table 3. Performance Evaluation of Cycle-Consistent GAN on Horse2Zebra and CELEBHQ Datasets Using PSNR and SSIM at Varying Train-Test Splits

9. Limitations

One of the significant limitations of this research is the reliance on publicly available image datasets, which might not fully capture the variability and richness of real-world situations, e.g., resolution changes, noise, or transmission compression artifacts. Furthermore, the deep learning models used—specifically GAN and CycleGAN—are computationally intensive and susceptible to issues like unstable convergence and mode collapse. The accuracy of the framework is highly dependent on synchronization between the embedding and extraction models, which can be prone to inconsistencies in real-world deployment. Furthermore, although integration of cryptographic methods enhances security, it adds computational overhead and complexity, which constrains real-time usage.

10. Future Scope

This work opens up a number of productive lines of future research towards enhancing steganalysis. One could explore hybrid models that combine CNNs with transformers to more effectively capture local and global image features. Extending the testing to higher-resolution images (e.g., 512×512 pixels) and other datasets would make the scenarios more realistic. Research into frequency-domain analysis methods could be a useful complement to spatial feature extraction. The framework can be extended to identify newer steganographic techniques such as coverless and adaptive steganography. Constructing strong defenses against adversarial steganography is still a pressing challenge. Moreover, the development of standardized benchmarks for comparing various steganalysis techniques would be useful for the research community. Lastly, investigating real-time detection systems and hardware acceleration could make practical deployment scenarios possible.

Conclusion

This work demonstrates the power of deep neural networks in modern steganalysis, where consistent detection performance is realized over a broad variety of steganographic embedding mechanisms. The proposed convolutional design demonstrates impressive capacity to find subtle statistical imperfections characteristic of hidden data payloads and to do so while maintaining utmost resource parsimony in computation. The results highlight the inherent trade-offs in detection sensitivity, false positive rates, and processing efficiency of realistic steganalysis deployments. As steganographic techniques grow increasingly sophisticated, this research highlights the need for creating more resilient, adaptive detection systems that can meet new challenges. These contributions further the essential debate on covert communication detection, offering meaningful insights for cybersecurity use where clandestine data transmission detection remains a priority.

Reference

1. NANDHINI SUBRAMANIAN, OMAR ELHARROUSS, SOMAYA AL-MAADEED AND AHMED BOURIDANE Image Steganography: A Review of the Recent Advances https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9335027&isnumber=9312710

2. Fridrich, J., & Kodovsky, J. (2012). Rich models for steganalysis of digital images. IEEE Trans. Inf. Forensics Security, 7(3), 868–882. https://doi.org/10.1109/TIFS.2012.2190402

3. Holub, V., Fridrich, J., & Denemark, T. (2014). Universal distortion function for steganography in an arbitrary domain. EURASIP J. Inf. Security, 2014, 1–13. https://doi.org/10.1186/1687-417X-2014-1

4 . Xu, G., Wu, H. Z., & Shi, Y. Q. (2016). Structural design of convolutional neural networks for steganalysis. IEEE Signal Process. Lett., 23(5), 708–712. https://doi.org/10.1109/LSP.2016.2548622

5. Ye, J., Ni, J., & Yi, Y. (2017). Deep learning hierarchical representations for image steganalysis. IEEE Trans. Inf. Forensics Security, 12(11), 2545–2557. https://doi.org/10.1109/TIFS.2017.2736699

6. Zhang, R., Zhu, F., Liu, J., & Liu, G. (2020). Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. IEEE Trans. Inf. Forensics Security, 15, 1138–1150. https://doi.org/10.1109/TIFS.2019.2947664

7. Yedroudj, M., Comby, F., & Chaumont, M. (2018). Yedroudj-Net: An efficient CNN for spatial steganalysis. ICASSP 2018, 2092–2096. https://doi.org/10.1109/ICASSP.2018.8461438

8. Boroumand, M., Chen, M., & Fridrich, J. (2018). Deep residual network for steganalysis of digital images. IEEE Trans. Inf. Forensics Security, 14(5), 1181–1193. https://doi.org/10.1109/TIFS.2018.2814842

9. Bas, P., Filler, T., & Pevný, T. (2011). "Break our steganographic system": The ins and outs of organizing BOSS. In Int. Workshop on Information Hiding, 59–70.

10. Goodfellow, I., et al. (2014). Generative adversarial nets. NeurIPS 27.

11. Bossbase dataset :- https://www.kaggle.com/datasets/lijiyu/bossbase

12. DIV2K Dataset :- https://www.kaggle.com/datasets/sharansmenon/div2k

13. CelebA Dataset :- https://www.kaggle.com/datasets/jessicali9530/celeba-dataset

14. Horse 2 Zebra Dataset :- https://www.kaggle.com/datasets/balraj98/horse2zebra-dataset

15. CelebHQ Dataset :- https://www.kaggle.com/datasets/lamsimon/celebahq