Annals of "Dunarea de Jos" University of Galati
Fascicle I. Economics and Applied Informatics
Years XXXI – nº3/2025
ISSN-L 1584-0409 ISSN-Online 2344-441X
www.eia.feaa.ugal.ro

DOI https://doi.org/10.35219/eai15840409564

Applications of Deep Learning in Image Steganography

Adrian Doroiman[*]

ARTICLE INFO

Article history:
Received September 29, 2025
Accepted November 22, 2025
Available online December 2025

JEL Classification
C45, C38, D83

Keywords:
steganography, classification, supervised learning

ABSTRACT

The current paper presents the concept of steganography, with a focus on image steganography. It outlines a few relevant techniques for image steganography. Provides implementation details and a comparative evaluation of design and results. Proposes further research steps.

1. Introduction

Steganography is the art of hiding in plain sight, or, in other words, the art and science of concealed communication, where the existence of a secret message may not even be suspected (Alabdali & Al Tuwairqi, 2021). The goal of steganography is to transmit the secret message using an unprotected or common medium, known as a carrier. This is the fundamental distinction between steganography and cryptography, the latter's objective being to protect the message by encoding it in a secure way, but essentially not hiding its presence. On the other hand, steganography aims to keep communication itself undetected.

The applications of steganography are multiple, ranging from protection of intellectual property through digital watermarking to enabling freedom of communication (Alabdali & Al Tuwairqi, 2021). Regardless of the carrier used, there have been significant advances of the available techniques over the course of history, with human ingenuity proving that there is an infinite number of solutions to the communication problem (Alabdali & Al Tuwairqi, 2021), (Zielińska, Mazurczyk, & Szczypiorski, 2012). And because the number of users and uses of these techniques cannot easily be controlled, there have been, historically (Zielińska, Mazurczyk, & Szczypiorski, 2012), and there may still be, ethical consideration over the usage of steganography.

2. Literature review

2.1 Steganography

While the current paper is focused on image steganography, it is important to point out that the generic steganographic principles can be applied to a multitude of environments. Basically, any medium that has some inherent redundancy or other characteristics that can be subtly manipulated is a candidate for steganographic techniques. A few such examples are mentioned in the following paragraphs.

Text steganography involves hiding information within text files. Techniques can be simple (e.g. manipulating spaces, line endings or tabs) to more complex linguistic methods that modify words, the structure of sentences or use grammars to generate cover texts (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

Audio steganography is about embedding secret messages in digital audio files by modifying the audio signal in a way that cannot be detected by the human ear. Techniques include LSB (Least Significant Bit), phase coding – modifying the phase of some of the frequency components -, echo hiding, and spread spectrum techniques – where the message is split across a range of frequencies (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

Video steganography is popular because video files have a large capacity for storing information, coupled with the redundancy caused by having multiple frames for similar data. The data can be embedded through techniques used by image steganography, or by techniques specific to the format of the compressed video streams (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

Network steganography, or protocol steganography, is an advanced technique which relies on the complexity of network protocols allowing the embedding of data in a manner which is hard to detect because

[*] Bucharest University of Economic Studies, Bucharest, Romania. E-mail address doroimandan22@stud.ase.ro (A. Doroiman).

of the inherent potentially lossy network traffic. There is error correction that is included in the protocols themselves which can be exploited, while other examples are manipulating packet order or even the timing of the packets (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

2.2 Steganalysis

Steganalysis is the science of detecting and, if possible, extracting of secret information hidden through steganography (Ahmad, et al., 2025), (Suresh & Sivanandam, 2021). As steganographic tools and techniques evolve to achieve greater performance in terms of imperceptibility and robustness, the steganalytic techniques have to evolve to maintain their own efficacity (Kheddar, Hemis, Himeur, Megías, & Amira, 2024).

The main challenge of the steganalysis scenarios is the absence of knowledge regarding the specific algorithm used to conceal the information. From this perspective, it may seem that the steganographic problem – hiding information – is favored, as both the carrier and the algorithm are not disclosed. However, this limitation had led to the development of "blind" or "universal" steganalysis techniques, that are capable of detecting the presence of hidden data regardless of the particular embedding algorithm used (Monika, Kumar, Dwivedi, Bera, & Sharma, 2017).

Beyond the primary goal of detecting the presence of a hidden message, steganalysis also includes gaining complementary information: the size of the hidden message, the algorithm used and the message itself (Fridrich & Kodovský, 2012). These objectives are increasingly difficult and fall in the category of supervised and semi-supervised learning: whether a hidden message exists or not can be considered a binary classification problem, while detecting the used algorithm is a multi-class classification task.

The steganalysis techniques fall into two main categories, from the perspective on their underlying methodologies: traditional and machine learning techniques.

A) Traditional techniques

Signature-based detection operates by searching for predefined patterns that can be associated with known steganographic techniques (Ahmad, et al., 2025). This approach has the advantage of high accuracy and speed, but effectiveness can be limited as it will not detect an unknown algorithm.

Statistical steganalysis techniques are based on finding statistical anomalies in the data, that are indicative of a hidden message being present (Suresh & Sivanandam, 2021). The most relevant such techniques for the image domain are:

- Chi-Square Analysis: specific to LSB steganography, particularly in the cases where data is embedded sequentially. Based on the fact that LSB embedding tends to randomize the LSB plane, which can be found to be statistically different compared to normal images (Böhme, 2010).
- RS (Regular-Singular) Analysis: specific to LSB steganography. Classifies small groups of pixels as regular or singular in a way that allows capturing the difference compared to normal images (Alrusaini, 2025).
- SPA (Sample Pair Analysis): specific to LSB steganography. Analyzes how pairs of pixels and the differences between their values change relative to each other (Böhme, 2010).

Transform domain approaches operate in a transform domain, such as Discrete Cosine Transform (DCT) used in JPEG compression, or the Discrete Wavelet Transform (DWT) domain (Kheddar, Hemis, Himeur, Megías, & Amira, 2024).

Spread spectrum steganalysis techniques involve attempting to identify the presence of a characteristic signal hidden within the cover medium through correlation techniques (Chaganti, Ravi, Alazab, & Pham, 2021).

B) Machine learning techniques

Initial application of machines learning techniques to steganalysis led to the development of classifiers, such as Support Vector Machines (SVMs), Random Forests, or other ensemble methods (Alrusaini, 2025). These methods offered improved detection performance compared to standalone statistical tests (Alrusaini, 2025).
Deep Learning architectures brought a paradigm shift because of their ability to perform automatic feature extraction (Saleh, Potrus, & Al-Sumaidaee, 2023). The most popular solutions involved complex architectures including CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks) and GANs (Generative Adversarial Networks) (Kheddar, Hemis, Himeur, Megías, & Amira, 2024), (Apau, Asante, Twum, Hayfron-Acquah, & Peasah, 2024).

Notable Deep Learning models include Xu-Net – focused on capturing high-frequency artifacts introduced by the embedding (Alrusaini, 2025), SRNet (Steganalysis Residual Network) – featuring residual connections and support for both spatial and frequency-like domains (Alrusaini, 2025) and EfficientNet, a family of models that introduces a scaling method called "compound scaling" (Alrusaini, 2025).

2.3 Image Steganography

Digital images are the preferred medium for steganographic messages because of a few favorable factors: number, format and technique accessibility (Bisht, Singla, & Joshi, 2024), (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

The number factor is the prevalence of images on social media platforms, personal devices and digital space in general (Bisht, Singla, & Joshi, 2024). This provides vast opportunities for concealing hidden messages,

as it is impractical to search for proof of a hidden message without a proper way of filtering first through the raw data.

The format of images is also compatible with invisible alterations. While there is no unique format, there are popular formats (i.e. jpeg, gif, bmp, png, tiff) that are covered by a large spectrum of steganographic techniques. Most formats include a high degree of redundancy, as images often contain information that is not critical to human perception, like the least significant color bit. This redundancy provides ample space for hiding a message inside the image.

Technique accessibility means the relative ease of implementation for certain techniques, particularly spatial domain methods like LSB substitution (Jammi, Raju, Munishankaraiah, & Srinivas, 2010).

2.3.1 Objectives

Capacity

Capacity refers to the amount of information that can be embedded in a certain medium (Wang, Liu, & Zhang, 2025). Each embedding method has an upper limit to the message length that can be hidden, which is usually expressed as a percentage, or as bits per pixel. While the higher payload capacity is an objective, it is well understood that that it would come with the drawback of reduced imperceptibility and perhaps even security (Wang, Liu, & Zhang, 2025), (Sharma & Gupta, 2024).

Imperceptibility

Imperceptibility is the property of the modified object (also called stego object), that contains both the original object and the hidden data, of being undistinguishable by human senses from a normal object. This is done by minimizing the perceptible distortion caused by embedding the secret message (Sadeghi, Dadkhah, & Ghaemmaghami, 2022).

Undetectability

Undetectability is the property of the stego object to have the secret message embedded in such a way that its presence cannot be detected not just by human senses (see Imperceptibility), but also through steganalysis techniques (Wang, Liu, & Zhang, 2025). The goal of a steganographic method aiming to be considered secure is to ensure that the existence of a hidden message cannot be detected by any tools (Tan, Li, Yang, & Liu, 2025).

Data integrity

Data integrity focuses on preserving the accuracy and completeness of the hidden information through the process of encoding, embedding, extraction and decoding into the cover medium and from the stego medium (Sharma & Gupta, 2024). In order to achieve this goal, techniques such as error detection and correction codes can be used, in conjunction with verification methods such as hash functions (Sharma & Singh, 2012).

Robustness

Robustness refers to the ability of the embedded data to preserve its integrity during various transformations and alterations of the stego images, such as image compression, rotation, resizing, etc. (Wang, Liu, & Zhang, 2025). While robustness is not usually a primary goal, it can become important when the medium is expected to pass through noisy environments (Wang, Liu, & Zhang, 2025).

Security

Security implies, beyond Undetectability, two aspects: first, that the hidden message cannot be detected even with the knowledge of the embedding algorithm (Sadeghi, Dadkhah, & Ghaemmaghami, 2022). Secondly, that the hidden message, even if detected, cannot be decoded, as it would be protected by a secure encryption layer.

2.3.2 Techniques

Spatial domain techniques

Spatial domain techniques involve direct manipulation of the pixel values of the cover image when embedding the secret message (Al-Otaibi, et al., 2024).

The simplest method in this group is LSB substitution, where the least significant bits of the image are replaced by the bits of the secret message. While LSB is easy to implement and offers high capacity, it is vulnerable to statistical attacks (Alrusaini, 2025).

Other spatial domain techniques include PVD (Pixel Value Differencing), which embeds the secret by modifying differences between adjacent pixel values and EMD (Exploiting Modification Direction) (Al-Otaibi, et al., 2024).

Transform domain techniques

Transform domain techniques perform a conversion of the image from the spatial domain to a frequency domain using transformation like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), or Fast Fourier Transform (FFT) (Ahmad, et al., 2025). The secret message is embedded through the modification of the coefficients from the transform domain (Chaganti, Ravi, Alazab, & Pham, 2021).

Such methods are in principle more tolerant to processing operations like compression than the spatial domain techniques, therefore more robust (Kheddar, Hemis, Himeur, Megías, & Amira, 2024).

Spread spectrum techniques

In Spread spectrum techniques, the secret message is split across frequencies or spatial locations of the cover image (Al-Otaibi, et al., 2024). This makes the hidden data more resistant to noise, compression and attempts at removal, though at the cost of reduced capacity (Sharma & Singh, 2012).

Adaptive steganography

Adaptive steganographic techniques choose the location of data embedding based on the criteria of minimizing the statistical distortions and therefore analyze the characteristics of the cover image to identify complex or noisy regions which are fit for embedding (Sadeghi, Dadkhah, & Ghaemmaghami, 2022). Notable examples of such steganographic frameworks include Highly Undetectable Steganography (HUGO) and Universal Wavelet Relative Distortion (UNIWARD) (Alrusaini, 2025).

Patchwork techniques

Patchwork techniques are statistical approaches that embed the secret information by creating redundant patterns inside the cover image. They introduce small and widespread changes to increase undetectability (Al-Otaibi, et al., 2024).

Coverless steganography

Coverless steganography is a modern approach that does not modify the cover image. Instead, it relies on properties of the cover image (e.g. hash, size, or even statistical properties) (Tan, Li, Yang, & Liu, 2025). The cover image can be selected from a library, or can be generated on demand.

This approach has the advantage that steganalysis methods looking for data being embedded in the cover image will not be able to detect the presence of the hidden message.

Deep Learning-based steganography

Deep Learning models, particularly ones based on GANs and Transformers, can learn to embed data in efficient ways, that are resistant to steganalysis. Often a generator network is used for embedding and extracting data, while a discriminator network is attempting to distinguish between the cover and the stego images (Tan, Li, Yang, & Liu, 2025). The architecture of such models can be simple, based on sequential CNNs, or can be more complex.

An example of a modern architecture is U-Net. U-Net is a CNN architecture conceived for biomedical image segmentation (Ronneberger, Fischer, & Brox, 2015). It can be used as a Generator network in an adversarial context to embed the secret into the cover image.

3. Methodology

3.1 Proposed architectures

As the steganographic techniques can vary, in the current paper a sample of increasingly complex methods will be described. While all the objectives described in Chapter 2.2 above are considered important, some of them are consistently favored by certain methods or their particular implementation, and one of the goals of this paper is to point out both quantitatively and qualitatively such differences considered relevant.

The selected implementations are described below:

- A: LSB implementation
  - Sequential embedding of the message bits into the least significant bits of the cover image
  - Variable payload size, with approximately the same cover image size (1.5 MB)
- B: Deep Learning approach – basic
  - Convolutional Layers for down sampling and up sampling the data in the encoder and decode
  - Convolutional Layers in the Classifier model which will attempt to determine whether an image is a cover or a stego image
  - B1: 64 x 64 pixels cover image, 64 bits payload, 0,0651% embedding rate, summary details in Annex 1 – Model B1 summary
  - B2: 128 x 128 pixels cover image, 64 bits payload, batch normalization, 0,016% embedding rate
  - B3: 1024 x 1024 pixels cover image, 256 bits payload, training dataset used from (Agustsson & Timofte, 2017), additional batch normalization blocks, 0,001% embedding rate
- C: Deep Learning approach – U-Net
  - Additional skip connections CNN layers, specific to the U-Net architecture – summary details in Annex 2 – Model C summary
  - 256 x 256 pixels cover image, dataset used from (Agustsson & Timofte, 2017)
  - 256 bits payload
  - 0,016% embedding rate

3.2 Evaluation methods

For each of the proposed solutions, the following indicators were observed:

- Embedding rate
- Visual imperceptibility
- Detection score (through the appropriate method)
- Message retrieval accuracy

- Training loss (where applicable)

4. Results

LSB implementation

There are no visible artifacts on the 3 stego images, as shown in Figure 7 below. The embedding rate is approximately 0.60%. However, through the Chi Square Analysis method, the stego images are detected, except for the smallest payload scenario – with the embedding rate approximately 0.001%, where the detection probability is 0.0289.

Figure 7. LSB implementation visual analysis

Deep Learning approach – basic

The training loss for the B1 models, over 10 epochs and a dataset of 3670 images is shown in Figure 8 below. Detection probability is 0.92, and message reconstruction accuracy is 100%.

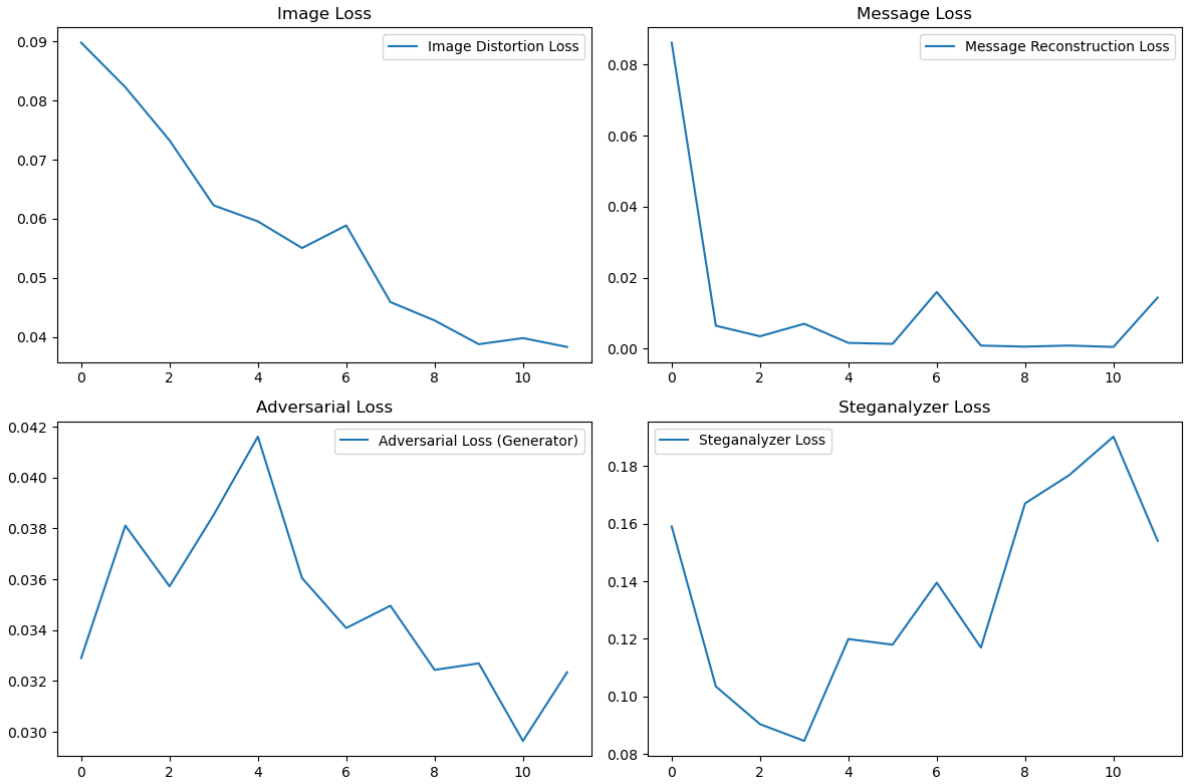

Figure 8. Training loss for B1 – 64x64

The training loss for the B2 models over 1 epoch and a dataset of 800 images is shown in Figure 9 below.

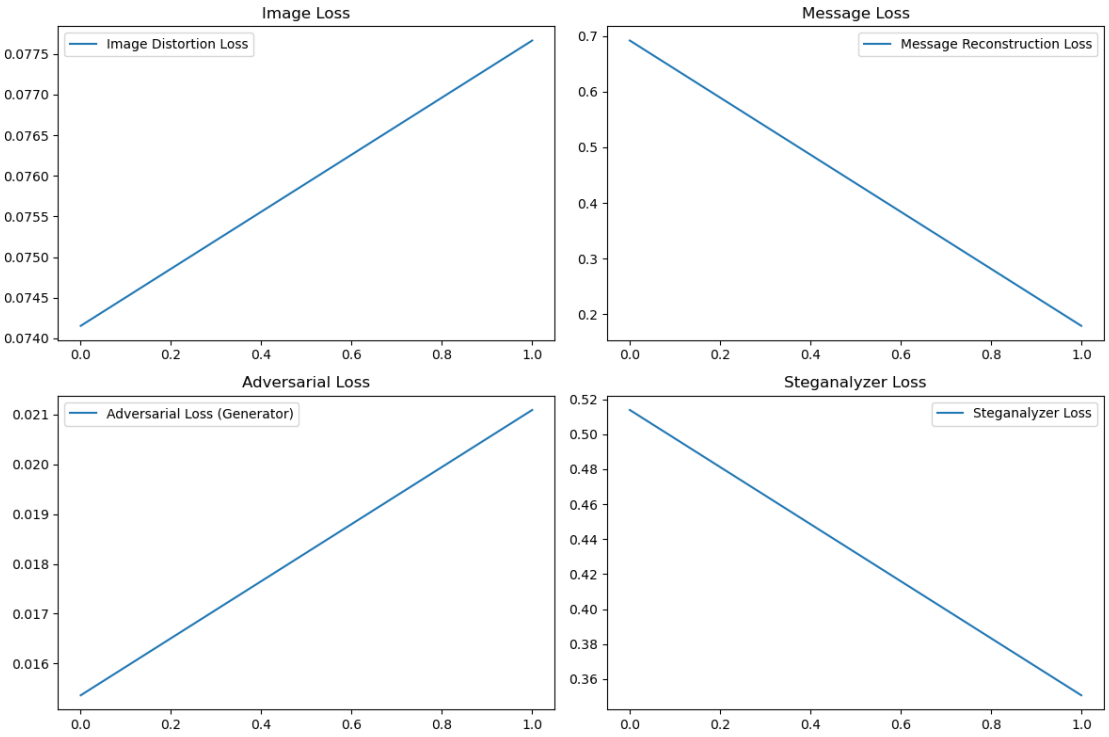

Figure 9. Training loss for B2 – 128x128

The training loss for the B3 models, over 1 epoch and a dataset of 800 images, is shown below. While message reconstruction accuracy is good and detection score is low – 0.0233 –, the stego image is not visually similar to the cover, as shown in Figure 11. Therefore, the need for a better architecture is apparent.

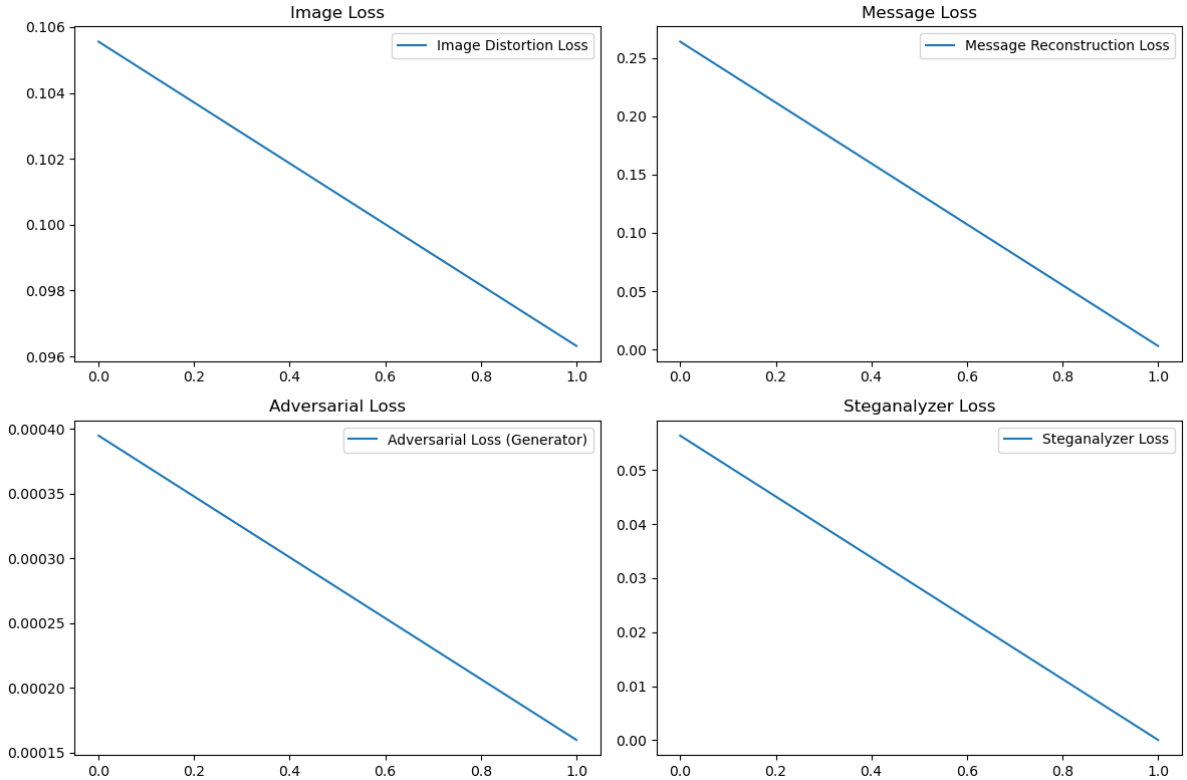

Figure 10. Training loss for B3 – 1024x1024.

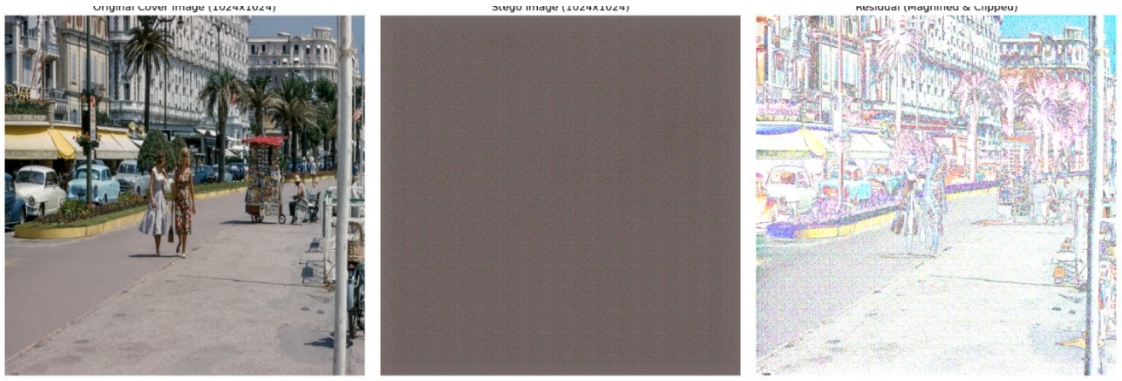

Figure 11. Sample images for B3: cover, stego and residual.

Deep Learning approach – U-Net

Two models were built, with a variation of hyperparameters. The first model was able to produce visually indistinguishable stego images, but with a high error rate on message recovery, as shown in Figure 12 below. Detection scores 0.5674, message accuracy 53.2%.

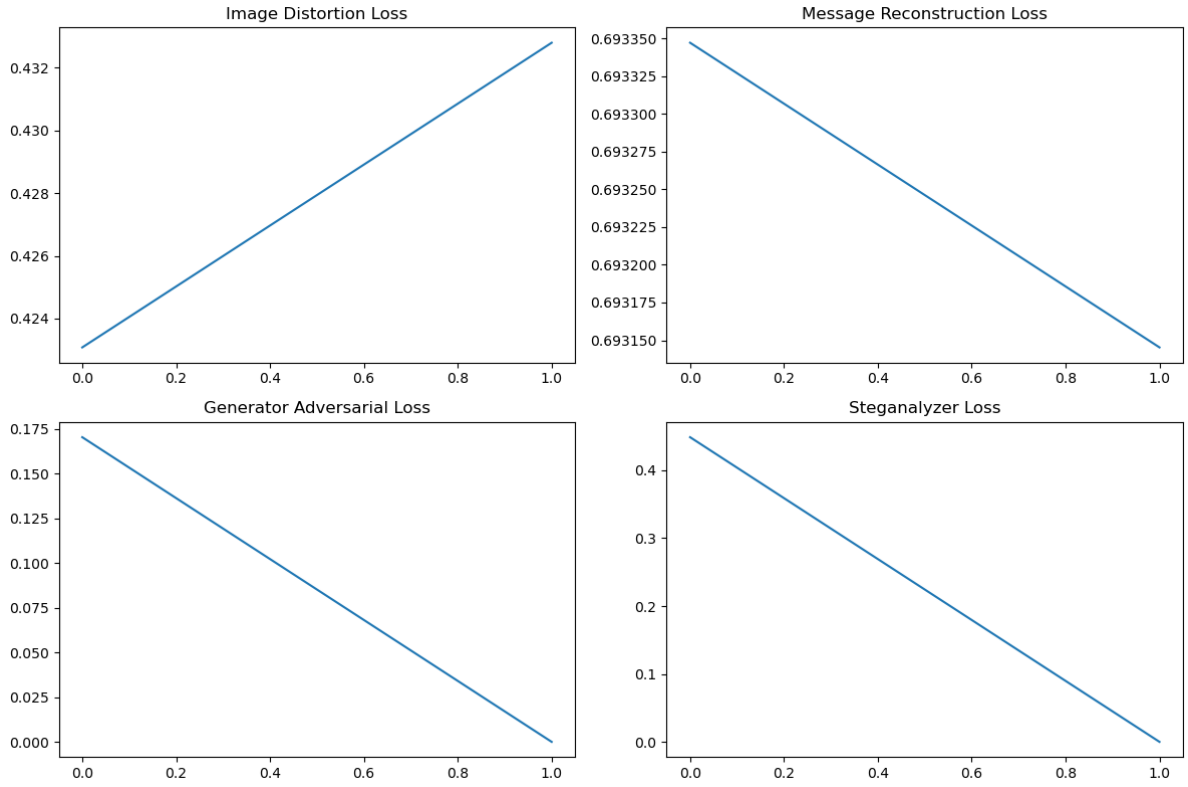

Figure 12. Training loss for C1 – 256x256

Figure 13. Sample images for C1: cover, stego and residual.

The second model (C2) was trained with a higher weight assigned to the message loss. The training loss over 10 epochs of 800 images can be observed below. Message accuracy is 100%, while detection score is 0.00.

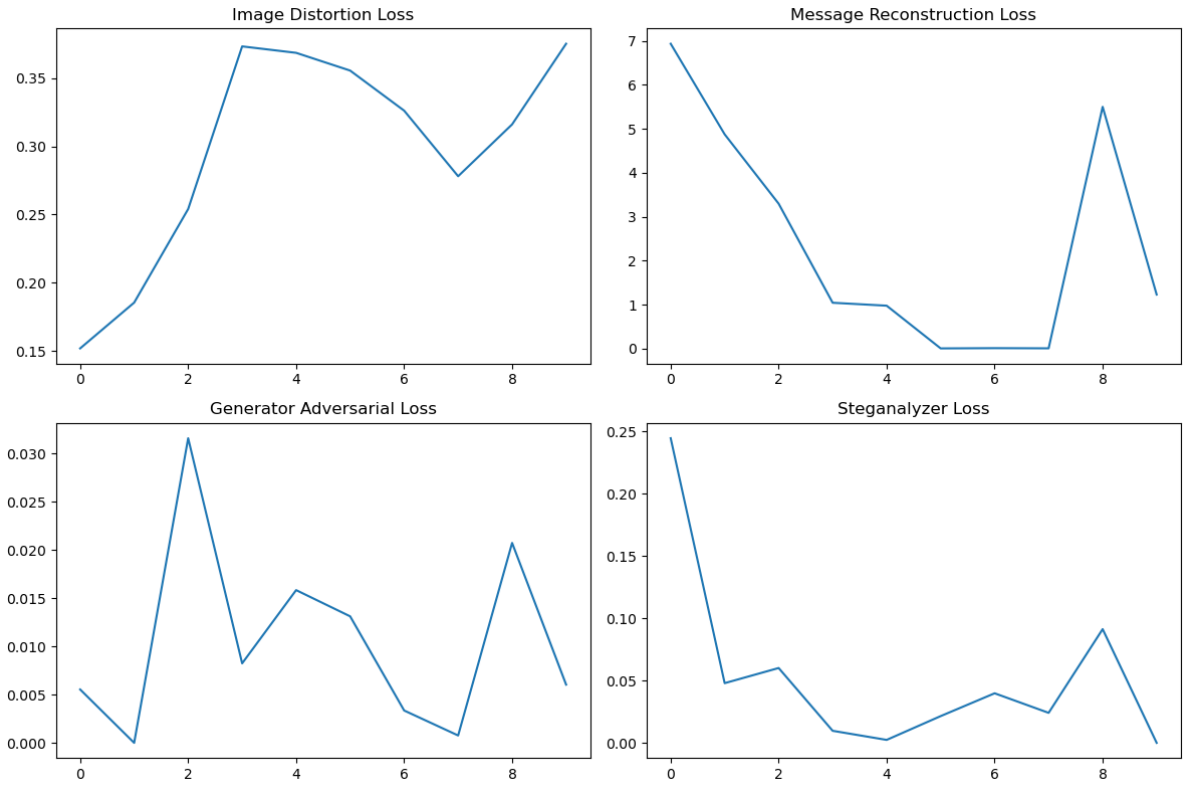

Figure 14. Training loss for C2 – 256x256

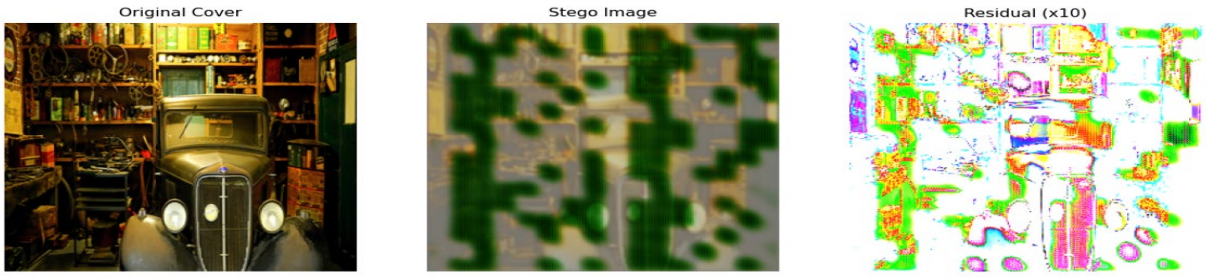

Figure 15. Sample images for C2: cover, stego and residual.

5. Conclusions

It is apparent that in order to meet all the steganographic objectives, a fine tuning has to be performed on the hyperparameters so that the objectives are properly balanced. It can be observed that models C1 and C2 are imbalanced, with the former prioritizing undetectability, and the latter message integrity. The U-Net models (C) are clearly superior to the standard convolutional models (B1, B2, B3) and can provide visually undetectable images, to a level comparable to the LSB model A, but a much better detectability score.

It is likely that a much higher number of epochs (50-200), in conjunction with a larger dataset, could lead to convergence of the model's loss towards 0. An auxiliary approach can be to train just the encoder and decoder to a satisfactory level, and only afterwards to enable the adversarial convergence with the steganalyzer model.

References

1. *Agustsson, E., & Timofte, R. (2017, July). NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.*
2. *Ahmad, M. A., Al-Qhtani, E., Samak, A. H., Ibrahim, A., Elloumi, M., & Ahmed, A. (2025, January). Deep Learning-Based Steganalysis for Detection and Classification of Possible Hidden Content in Images. Fusion: Practice and Applications (FPA), 17, 377–393. doi:10.54216/FPA.170228*
3. *Alabdali, N. J., & Al Tuwairqi, S. M. (2021). An Overview of Steganography through History. International Journal of Scientific Engineering and Science (IJSES), 4, 1–6. Retrieved from http://ijses.com/wp-content/uploads/2021/03/117-IJSES-V4N12.pdf*
4. *Al-Otaibi, S., Alanazi, A., Alarifi, A., Alshammari, M., AlGhamdi, M. A., & AlBadran, B. (2024, May). A Novel Multi-Layered Security Framework: Integrating Enhanced RSA and LSB Steganography for Robust Data Protection. Journal of Sensor and Actuator Networks, 13, 341. doi:10.3390/jsan13050341*
5. *Alrusaini, O. A. (2025). Deep learning for steganalysis: evaluating model robustness against image transformations. Frontiers in Artificial Intelligence, 8, 1532895. doi:10.3389/frai.2025.1532895*

6. *Apau, R., Asante, M., Twum, F., Hayfron-Acquah, J. B., & Peasah, K. O. (2024). Image steganography techniques for resisting statistical steganalysis attacks: A systematic literature review. PLoS One, 19(9), e0308807. doi:10.1371/journal.pone.0308807*
7. *Bisht, A., Singla, A., & Joshi, K. (2024, July). A Review on Image Steganography Techniques. International Research Journal on Advanced Engineering Hub (IRJAEH), 2, 1986-1996. doi:10.47392/IRJAEH.2024.0271*
8. *Böhme, R. (2010, July). Principles of Modern Steganography and Steganalysis. Information Hiding. IH 2010. Lecture Notes in Computer Science, vol 6387. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-14313-7_2*
9. *Chaganti, R., Ravi, V., Alazab, M., & Pham, T. D. (2021, October). Stegomalware: A Systematic Survey of Malware Hiding and Detection in Images, Machine Learning Models and Research Challenges. TechRxiv. doi:10.36227/techrxiv.16755457.v1*
10. *Fridrich, J., & Kodovský, J. (2012). Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, 7, 868–882. doi:10.1109/TIFS.2012.2190402*
11. *Jammi, A., Raju, Y., Munishankaraiah, S., & Srinivas, K. (2010). Steganography: An Overview. International Journal of Engineering Science and Technology, 2 (10), 5985-5992.*
12. *Kheddar, H., Hemis, M., Himeur, Y., Megías, D., & Amira, A. (2024). Deep Learning for Steganalysis of Diverse Data Types: A review of methods, taxonomy, challenges and future directions. Neurocomputing, 586, 127528. doi:10.1016/j.neucom.2024.127528*
13. *Monika, Kumar, P., Dwivedi, Y. P., Bera, S., & Sharma, M. (2017). Review on Universal Steganalysis Techniques Based on the Feature Extraction in Transform Domain. International Journal of Engineering Research and Development, 13, 07–11.*
14. *Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 9351, pp. 234–241. Springer International Publishing. doi:10.1007/978-3-319-24574-4_28*
15. *Sadeghi, M., Dadkhah, S., & Ghaemmaghami, S. (2022). Adversarial cost-based steganography by mimicking the modification probability distribution of cover image. Journal of Soft-computing and Data-mining, 3, 1–10. Retrieved from http://jsdp.rcisp.ac.ir/article-1-1302-en.html*
16. *Saleh, A. A., Potrus, M. S., & Al-Sumaidaee, S. A. (2023). Image Steganalysis Based on Deep Convolutional Neural Network (DCNN): A Survey. Journal of Applied Computer Science & Mathematics, 17, 5–13. doi:10.4316/JACSM.202301001*
17. *Sharma, M., & Singh, S. (2012, July). Data Hiding with Integrity using LSB & Modulus 4 Bit Technique. International Journal of Computer Engineering and Research, 2, 189–192.*
18. *Sharma, P., & Gupta, R. (2024, August). A Novel Method of Image Steganography for Enhanced Security and Data Integrity. International Research Journal of Modernization in Engineering Technology and Science, 6.*
19. *Suresh, M., & Sivanandam, S. N. (2021). Approaches and Methods for Steganalysis - A Survey. International Journal of Network Security & Its Applications (IJNSA), 13, 01–14. doi:10.5121/ijnsa.2021.13401*
20. *Tan, J., Li, X., Yang, B., & Liu, F. (2025). Image Steganography Method Based on Adaptive Embedding and Generative Adversarial Network. Computer Modeling in Engineering & Sciences (CMES), 80, 1787–1805. doi:10.32604/cmes.2024.057590*
21. *Wang, Z., Liu, Y., & Zhang, M. (2025, March). AI-Powered Steganography: Advances in Image, Linguistic, and 3D Mesh Data Hiding - A Survey. arXiv preprint arXiv:2503.12345.*
22. *Zielińska, E., Mazurczyk, W., & Szczypiorski, K. (2012). The Advent of Steganography in Computing Environments. Tech. rep., Warsaw University of Technology. Retrieved from https://arxiv.org/abs/1201.4847*