IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

GAN-Based Steganography: Enhancing Data Concealment

Tushar Tikhe[1], Pranit Thorat[2], Vivek Rodge[3]
Department of Computer Engineering[1,2,3]
MES Wadia College of Engineering, Pune, India

Abstract: *Ensuring the secure transfer of information across networks is paramount to uphold confidentiality and privacy, both of which are critical in today's society. To safeguard information from unauthorized access, interception, or tampering during communication, various protective techniques are employed. Steganography is one such method, involving the concealment of information within a cover medium, such as an image, so that the presence of the hidden message remains undetected. This approach is particularly valuable when cryptographic measures alone are insufficient for protecting sensitive data. Even if the cover image is intercepted, the concealed message should only be accessible to the intended sender and recipient. Our proposed solution addresses security risks by embedding secret data within stego-images generated through Generative Adversarial Networks (GANs), and then accurately extracting it with a decoder. The advantage of using GANs lies in their ability to produce images that are visually realistic and appealing, thus minimizing the chances of the hidden message being discovered. Furthermore, this platform can be adapted to prevent digital piracy by embedding watermarks into digital content, and it can be further refined with expert input.*

Keywords: Steganography, GAN (Generative Adversarial Networks), Image, Encode

I. INTRODUCTION

Ensuring the security of sensitive information during transmission is paramount to prevent unauthorized access, interception, or detection. Recently, Generative Adversarial Networks (GANs) have shown significant potential in the field of steganography, which involves embedding hidden information within a cover medium. GANs are advanced deep learning algorithms capable of generating lifelike images, videos, or audio samples. By integrating a secret message into the noise vector used for image generation, GANs can produce cover media with concealed information that is not easily noticeable.

In the adversarial training process, both the encoder and decoder learn a vast number of parameters, essentially becoming the encryption and decryption keys themselves. This innovative approach offers several benefits over traditional steganography methods, including the creation of visually convincing cover media and enhanced protection against detection. Consequently, GAN-based steganography could revolutionize the security of confidential communications and data storage.

The architecture of this system includes three distinct networks: an encoder, a decoder, and a critic. The encoder embeds the secret message within the cover image, the decoder retrieves the hidden message from the image, and the critic provides feedback to refine the encoder's output, leading to more realistic image generation. The system's effectiveness is assessed based on three criteria: capacity, distortion, and secrecy. Additionally, the proposed GAN model boasts a straightforward and user-friendly interface.

Figure 1. A randomly selected cover image(left) and the corresponding Steganographic image generated by the application.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

II. METHODOLOGY

This section presents an intricate technical exploration of our steganography methodology, encompassing the structural framework, training regimen, and symbolic representation. The fundamental processes within steganography entail encoding and decoding, wherein encoding conceals a binary message within a cover image, yielding a steganographic image, while decoding retrieves the hidden message from the steganographic image.

Symbolism:

In our approach, designated as X, an innovative technique for image steganography is introduced, designed to accommodate diverse cover image dimensions and arbitrary binary data. The cover image is denoted as C, and the resultant steganographic image is represented as S, both constituting RGB colour images sharing dimensions WH. The binary message to be obscured within C is portrayed as M $\in$ {0,1}DW*H. Here, D serves as the upper threshold for payload capacity, while the actual payload, governed by the error rate p $\in$ [0, 1], is determined by (1-2p) D.

To generate S, C is sampled from the probability distribution of all natural images, Pc. Employing an encoder, ε (C, M), facilitates the creation of S, while a decoder, D(S), retrieves the secret message, M^. The optimization objective entails training ε and D to minimize the decoding error rate p and the disparity between the distributions of natural and steganographic images, dis(Pc, Ps). Furthermore, training a critic network, C(.), aids in estimating dis(Pc, Ps) and optimizing ε and D.

Let X $\in$ RDWH and Y $\in$ RD'WH denote two tensors sharing identical width and height dimensions but potentially differing depths, D and D', respectively. Concatenation along the depth axis is represented by Cat : (X, Y) $\rightarrow$ R(D+D')WH.

ConvD$\rightarrow$D' : X $\in$ RDWH $\rightarrow$ Φ $\in$ RD'WH signifies a convolutional block mapping input tensor X to feature map Φ. This block encompasses a convolutional layer with a kernel size of 3, a stride of 1, and 'same' padding, succeeded by a leaky ReLU activation function and batch normalization. In cases where this block serves as the terminal element in the network, activation and normalization operations are omitted.

Lastly, the adaptive mean spatial pooling operation, denoted as Mean: X $\in$ RDWH $\rightarrow$ RD, calculates the average of the W*H values within each feature map of tensor X.

III. MODELING AND ANALYSIS

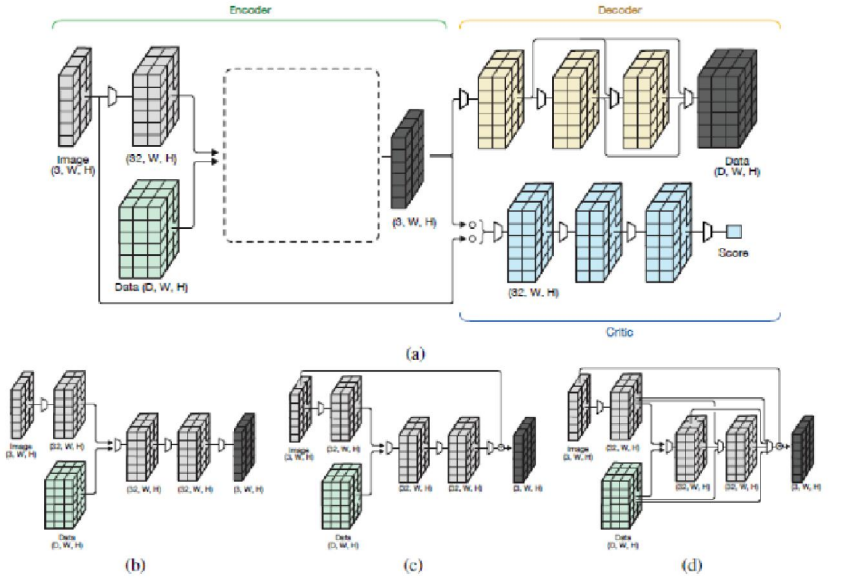

Figure 2. (a) The model architecture with the Encoder, Decoder, and Critic. The blank rectangle representing the Encoder can be any of the following: (b) Basic encoder, (c) Residual encoder and (d) Dense encoder. The trapezoids represent convolutional blocks, two or more arrows merging represent concatenation operations, and the curly bracket represents a batching operation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

In this scholarly investigation, we unveil the "GAN-Based Steganography application," a sophisticated generative adversarial network meticulously crafted to embed an arbitrary bit vector covertly within a cover image. Our proposed structural blueprint, depicted in Figure 2, delineates three core constituents: (1) an Encoder, tasked with assimilating a cover image alongside a data tensor or message input, subsequently engendering a steganographic output.

Our architecture for the "GAN-Based Steganography application" comprises an encoder module that takes a cover image C and a binary data tensor M $\in\{0,1\}DWH$ as input, where D represents the number of bits to be hidden within each pixel of the cover image. Initially, the cover image C undergoes convolutional block processing to derive tensor a:

a=Conv3→32(C)

Next, tensor a is concatenated with the message tensor M and further processed through another convolutional block to obtain tensor b:

b=Conv32+D→32(Cat(a,M))

ENCODER:

BASIC: In the basic approach, two convolutional blocks are sequentially applied to tensor b to produce the steganographic image:

$\epsilon b(C,M)$=Conv32→3(Conv32→32(b))

Residual: Incorporating residual connections aims to enhance the quality of the steganographic image by improving model stability and convergence. This involves adding the cover image C to the output of the primary encoder to generate a residual image:

$\epsilon r(C,M)=C+\epsilon b(C,M)$

Dense: The dense variant introduces additional connections between convolutional blocks, inspired by DenseNet architecture, to potentially enhance the embedding rate:

c=Conv64+D→32(Cat(a,b,M))

d=Conv96+D→3(Cat(a,b,c,M))

$\epsilon d(C,M)=C+d$

Each variant's output is a steganographic image $S=\epsilon b,r,d(C,M)S=\epsilon b,r,d(C,M)$ with the same resolution and depth as the cover image C.

DECODER:

The decoder network extracts the data tensor M from the steganographic image S:

a=Conv3→32(S)

b=Conv32→32(a)

c=Conv64→32(Cat(a,b))

$D(S)$=Conv96→D(Cat(a,b,c))

The decoder applies convolutional operations to obtain an estimate of the original data tensor M, denoted as ^M.

CRITICS:

To improve image quality and evaluate encoder performance, an adversarial Critic network is introduced. The Critic network consists of three convolutional blocks followed by a convolutional layer with a single output channel. Adaptive mean pooling is applied to the output of the convolutional layer to obtain scalar scores:

a=Conv32→32(Conv32→32(Conv3→32(S)))

$C(S)$=Mean(Conv32→1(a))

TRAINING:

Iterative optimization of the encoder-decoder network and the critic network is performed. The encoder-decoder network optimization involves joint optimization of three losses: cross-entropy loss (Ld), mean square error loss (Ls), and loss computed by the critic network (Lr). The training objective is to minimize Ld + Ls + Lr.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

For training the critic network, Wasserstein Loss (Lc) is minimized. In each iteration, the cover image C is paired with a data tensor M. The data tensor is composed of D*W*H bits generated randomly from a Bernoulli distribution. Standard data augmentation techniques are applied to C during preprocessing. The Adam optimizer is used with a learning rate of 1e-4, and gradient norm clipping is set to 0.25. Critic network weights are clipped to the range of [-0.1, 0.1]. The training process continues for 32 epochs.

IV. RESULTS AND DISCUSSION

We conduct experiments to train and evaluate our model using the Div2k and COCO datasets. For each of the three model versions, we assess their performance across six distinct data depths, where D∈{1, 2,..., 6}, representing the target bits per pixel. The data tensor's geometry, derived from randomly generated data, is DWH. Following the default train/test split recommended by the developers of the Div2K and COCO datasets, our trials are conducted. Table 1 presents the typical RS-BPP, PSNR, and SSIM metrics on the test set. Our models are trained on GeForce GTX 1080 GPUs, with each epoch lasting approximately 10 minutes for Div2K and 2 hours for COCO.

Table 1: The relative payload and image quality metrics for each dataset and model variant.

Dataset	D	Accuracy			RS-BPP			PSNR			SSIM		
		Basic	Resid.	Dense	Basic	Resid	Dense	Basic	Resid.	Dense	Basic	Resid.	Dense
Div2K	1	0.95	0.99	1.00	0.91	0.99	0.99	24.52	41.68	41.60	0.70	0.96	0.95
	2	0.91	0.98	0.99	1.65	1.92	1.96	24.62	38.25	39.62	0.67	0.90	0.92
	3	0.82	0.92	0.94	1.92	2.52	2.63	25.03	36.67	36.52	0.69	0.85	0.85
	4	0.75	0.82	0.82	1.98	2.52	2.53	24.45	37.86	37.49	0.69	0.88	0.88
	5	0.69	0.74	0.75	1.86	2.39	2.50	24.90	39.45	38.65	0.70	0.90	0.90
	6	0.67	0.69	0.70	2.04	2.32	2.44	24.72	39.53	38.94	0.70	0.91	0.90
COCO	1	0.98	0.99	0.99	0.96	0.99	0.99	31.21	41.71	42.09	0.87	0.98	0.98
	2	0.97	0.99	0.99	1.88	1.97	1.97	31.56	39.00	39.08	0.86	0.96	0.95
	3	0.94	0.97	0.98	2.67	2.85	2.87	30.16	37.38	36.93	0.83	0.93	0.92
	4	0.87	0.95	0.95	2.99	3.60	3.61	31.12	36.98	36.94	0.83	0.92	0.92
	5	0.84	0.90	0.92	3.43	3.99	4.24	29.73	36.69	36.61	0.80	0.90	0.91
	6	0.78	0.84	0.87	3.34	4.07	4.40	31.42	36.75	36.33	0.84	0.89	0.88

We observe that across all iterations of our model, performance is consistently better on the COCO dataset compared to the Div2K dataset, likely due to differences in content types between the two datasets. Images from the Div2K dataset typically depict open landscapes, whereas those from the COCO dataset tend to be more cluttered, featuring multiple objects, thereby providing our model with more diverse surfaces and textures for embedding data.

Furthermore, our analysis reveals that the residual variant, despite having comparable image quality but a smaller relative payload, ranks second to our dense variant in terms of both relative payload and image quality. Conversely, the simple variant consistently exhibits the poorest performance, with relative payloads and image quality scores 15–25% lower than those of the dense variant.

It's worth noting that while dense models offer a higher relative payload, there may be a decline in the average peak signal-to-noise ratio, indicating a reduction in similarity between the cover image and steganographic images.

Detection of Steganographic Images:

To evaluate the effectiveness of our steganography technique in evading detection by steganalysis software, we conduct tests on the undetectability of steganographic images generated by our Dense models using two open-source steganalysis techniques: Statistical Steganalysis and Neural Steganalysis.

We randomly select 1,000 cover photos from the test set for statistical steganalysis, generating corresponding steganographic images using our Dense architecture with a data depth of 6. Subsequently, we utilize the StegExpose tool, which incorporates several widely used steganalysis approaches, to analyze the data. Our findings indicate that our model exhibits potential to evade detection by conventional steganalysis tools, meeting the fundamental requirements for a viable steganography process.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

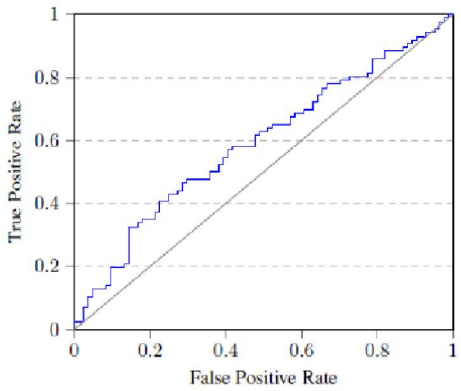

Figure 2: ROC curve produced by StegExpose Library for set of 1000 steganographic image generated by Dense architecture.

To accommodate colour photos, we make minor adjustments and evaluate our model's ability to evade deep learning-based steganalysis techniques in Neural Steganalysis. We experiment with various relative payloads and training set sizes while training this model to identify steganographic images generated by the "GAN-Based Steganography application". Without precise knowledge of the model parameters, it becomes challenging for an external party to develop a model capable of recognizing steganographic images produced by the "GAN-Based Steganography application".

The capacity to encrypt up to 2.0 bpp (bits per pixel) at a fixed detection error rate of 20% positions the "GAN-Based Steganography application" ahead of WOW, S-UNIWARD, and HILL methodologies.

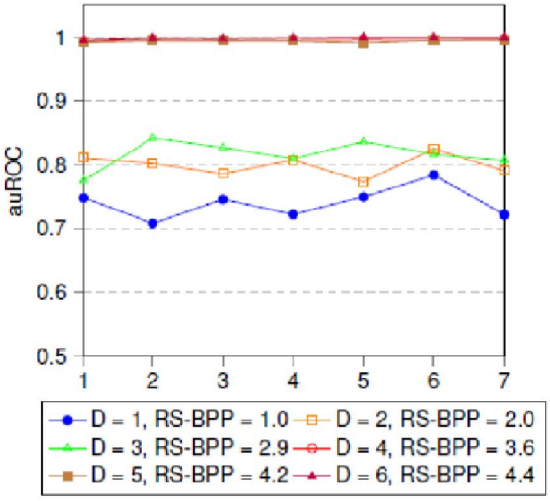

Figure 3: This plot shows performance of steganography detector on held-out test set. The X-axis indicates the number if different GAN instances used and Y axis indicates the area under ROC curve

V. CONCLUSION

Making slight adjustments to accommodate color photos, we assess our model's capability to evade deep learning-based steganalysis techniques in Neural Steganalysis. We conduct experiments with various relative payloads and training set sizes, training this model to identify steganographic images generated by the GAN-Based Steganography application". Lack of precise knowledge about the model parameters poses a challenge for external parties attempting to develop a model capable of recognizing steganographic images produced by the GAN-Based Steganography application".

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

The capacity to encrypt data up to 2.0 bits per pixel (bpp) at a fixed detection error rate of 20% positions the GAN-Based Steganography application" ahead of WOW, S-UNIWARD, and HILL methodologies.

ACKNOWLEDGEMENTS

It is with great pleasure that we present the project report on 'GAN-Based Steganography'.
First and foremost, we would like to express our sincere appreciation to our guide, Prof. Mr. S.B. Shinde. His unwavering guidance and invaluable advice played a pivotal role in the successful completion of this project. His insightful suggestions were instrumental in ensuring the quality and accuracy of this report.
We extend our heartfelt gratitude to Prof.Nuzhat F. Shaikh, Head of the Computer Engineering Department at MES Wadia College of Engineering, for her kind cooperation and encouragement throughout the completion of this report.
We also wish to express our thanks to our Principal, Dr. Manisha P. Dale, and all the faculty members for their wholehearted support and cooperation during the preparation of this report. Additionally, we are grateful to our laboratory assistants for their valuable assistance in the laboratory.
Last but certainly not least, we acknowledge that the foundation of our success and confidence rests on the blessings of our dear parents and the unwavering support of our beloved friends.

REFERENCES

[1]. P. P. Bandekar and G. C. Suguna, "LSB Based Text and Image Steganography Using AES Algorithm," in Proceedings of the 3rd International Conference on Communication and Electronics Systems, ICCES 2018, Oct. 2018, pp. 782–788. doi: 10.1109/CESYS.2018.8724069.

[2]. K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni, "SteganoGAN: High-Capacity Image Steganography with GANs," Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.03892

[3]. Panjab University. University Institute of Engineering and Technology and Institute of Electrical and Electronics Engineers, 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS) : (December 21-22, 2015), University Institute of Engineering and Technology, Panjab University, Chandigarh, 160014, India.

[4]. S. Agarwal and S. Venkatraman, "Deep Residual Neural Networks for Image in Audio Steganography (Workshop Paper)," in Proceedings - 2020 IEEE 6th International Conference on Multimedia Big Data, BigMM 2020, Sep. 2020, pp. 430–434. doi: 10.1109/BigMM50055.2020.00071