SIMULATION OF STEGANALYSIS USING CNN

Manushree Joshi[1], Mohammad Kaif Khan[2], Shivani Singh[3], Yash Sahni[4]

[1 2 3 4] *Undergraduate Students, Computer Science and Engineering Department, Institute of Technology and Management GIDA Gorakhpur, Uttar Pradesh, India*

ABSTRACT

Steganography is a technique of data hiding that embeds the secret message inside a digital media for providing a method of invisible communication. Different kinds of digital file formats can be used for steganography of which digital images are the most popular as it is present in a massive number in the internet.There exists a large variety of steganographic techniques, for hiding secret information in image. Several efforts are made to establish ways of detecting whether or not an image contains a steganographic element. Steganalysis is the technique of detecting the presence of steganography that can serve as an effective way to judge the security performance of steganographic techniques. In this paper, we provide an overview of digital image Steganalysis technique for detecting steganographic method and identify the area to look out for the hidden information. These techniques are discussed and analyzed in terms of their ability to detect secret message in an image file. We have also reviewed some researches on steganography.

The main aim of this paper is to review the previous work done and available ways, present trends and discuss the challenges that are currently available in the studies. Along with these, the datasets that are commonly used and publicly available , the evaluation metrics considered are also discussed. Finally, a comparison on the performance among the methods and a possible discussion identifying the gaps in the present studies, pros and cons of the methods are elaborated.

Keywords: - *Image Steganography, Steganalysis, Least Significant Bits (LSB), Convolutional Neural Network (CNN), Generative Adversarial Network (GAN).*

1. INTRODUCTION

Technology has blitz scaled over the past years which is leading to a wide usage of multimedia for transferring data, especially Internet of Things (IoT). Usually, these transfer takes place over insecure network channels. In particular, the internet came across accelerated popularity for exchanging digital media and individuals, private companies, institutions, governments use these multimedia data transfer methods for exchanging data. Though there are number of advantages attached with it, one prominent drawback is the privacy and security of the data. There are number of tools available already which are capable of exploiting the privacy, data integrity and security of the data being transmitted which has made the possibility of malicious threats, eavesdropping and other subversive activities. A new research topic, steganography, has gained acceptance in this context to hide the data that is not perceptible to human eyes.

Fig-1: General working of steganography.

Techniques of information hiding have been available for a long time but recently their importance has been accelerated. The main reason behind this is the increase in the data traffic through the internet and social media networks. Steganography, which is used to hide the information in plain sight, allows the use of wide variety of the secret information forms like image, text, audio, video and files. Cryptography is the popular method used in the field of information hiding, but, steganography has gained popularity in recent times.

Steganography can be defined as the procedure of hiding a secret small multimedia data inside another but much larger multimedia data such as image, text, file or video. Image steganography is a technique to hide an image inside any other image. In image steganography, the cover image is manipulated in such a way that the hidden data is not visible thus making it non suspicious as in the case of cryptography. Inversely, Steganalysis is used to detect the presence of any secret message covered in the image and to extract the hidden data. Steganalysis is a process of classifying if the image is either a stego image or a normal image. Apart from classifying the image, further investigation is carried out to detect the location and the content of the secret image inside the cover image.

With the availability of massive amounts of data, deep learning (DL) has become the trend and is extensively used for many applications. Deep learning is a useful tool in various applications like image classification, automatic speech recognition, image recognition, natural language processing, recommendation systems, processing of medical images. Though research on steganography is quite recent, it has benefited from DL methods including Convolutional Neural Networks (CNNs) Generative Adversarial Networks (GANs) based methods and their deployment in both steganography and Steganalysis.

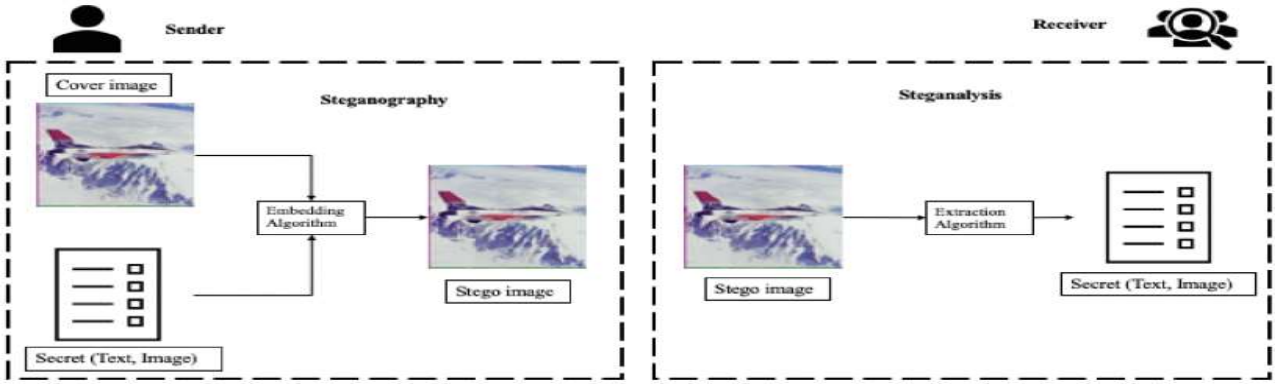

Fig-2: General working principle of steganography and Steganalysis.

2. OBJECTIVE OF THE STUDY

1. To recognize suspected data streams, determine whether or not they have hidden messages encoded into them, and if applicable, recover the hidden data.
2. To identify suspected packages, determine whether or not they have a payload encoded into them, and recover that payload.

3. LITERATURE REVIEW

After reviewing all the frameworks available, we have grouped the methodologies primarily into three categories, namely, traditional image steganography methods, CNN-based image steganography methods and GAN-based image steganography methods.

3.1 Traditional-Based Steganography Method

Conventionally, Least Significant Bits (LSB) substitution method is employed to perform image steganography. Images are usually of higher pixel quality, out of which few pixels are used. LSB methods works under the assumption that modifying a few pixel values would not show any visible changes. The secret information is converted into a binary form. The cover image is scanned to determine the least significant bits in the noisy area. The binary bits from the secret image are then substituted in the LSBs of the cover image. As overloading the cover image may lead to visible changes, the presence of the secret information may be leaked, hence, the substitution method has to be performed cautiously.

3.2 CNN-Based Steganography Methods

Image steganography using CNN models is mainly inspired from the encoder-decoder architecture. Two inputs – cover image and the secret image are fed as the input to the encoder to generate the stego image and the stego image is given as input to the decoder to output the embedded secret image. The basic principle is the same except different methods have tried different architectures. The way the input cover image and the secret image are concatenated are also different in different approaches while the variations in the convolutional layer, pooling layer are expected. The number of filters used, strides, filter size, activation function used and loss function vary from method to method. An important point to note here is that the size of the cover image and the secret image has to be same, so every pixel of the secret image is distributed in the cover image.

Convolutional Neural Networks have shown to learn structures that correspond to logical features. These features increase their level of abstraction as we go deeper into the network. Firstly, the ConvNet will have a good idea about the patterns of natural images, and will be able to make decisions on which areas are redundant, and more pixels can be hidden there. By saving space on redundant areas, the amount of hidden information can be increased. The exact way in which the network will hide the information cannot be known to anybody who doesn't have the weights because the architecture and the weights can be randomized,

To concatenate the cover image and the secret image, a Separable Convolution with Residual Block (SCR) is used. The embedded image is given as the input to the encoder for constructing the stego image which is fed to the decoder to output the decoded secret image.To obtain this, ELU (Exponential Linear Unit) and Batch normalization are used. A new cost function, called the variance loss is proposed to reduce the effect of noise in the generated container image An encoder-decoder architecture was proposed by Rahim *et al.* in. This method differs from the others in the way the inputs are given. The encoder part consists of two parallel architectures each for the cover and the secret image. Features from the cover image and the secret images are extracted through the convolutional layer and concatenated. The stego image is constructed with the help of these concatenated features.

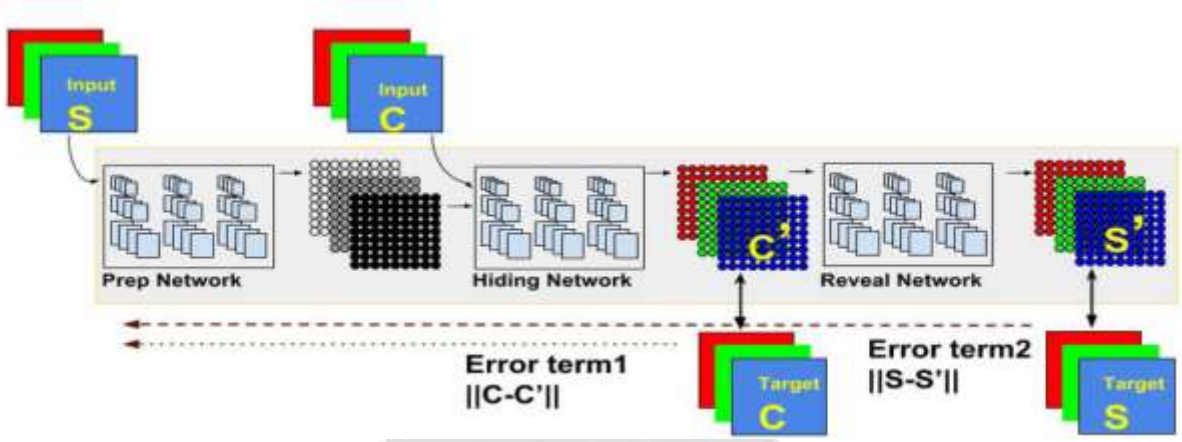

Fig-3: General working of CNN model.

The model is composed of three parts: The Preparation Network, Hiding Network (Encoder) and the Reveal Network. The goal of this model is to be able to encode information about the secret image S into the cover image C, generating C_prime that closely resembles C, while still being able to decode information from C_prime to generate the decoded secret image S_prime. This decoded image S_prime should resemble the secret image S as closely as possible.

The Preparation Network is responsible for preparing data from the secret image that is to be concatenated with the cover image and later fed to the Hiding Network. The Hiding Network working on it further transforms that input into the encoded cover image C_prime. It is noted here that the loss function for the Reveal Network is different from the loss function for the Preparation and Hiding Networks.

3.3 GAN-Based Steganography Methods

General Adversarial Networks are a type of deep CNN. It was introduced by Goodfellow *et al.* in 2014.In GAN, the generator and discriminator networks are the two networks that compete against each other to generate a perfect image in GAN architecture. The generator model is given the data which generates the output that is a close approximation of the given input image. The discriminator networks discriminates the generated output and classify the images generated as either fake or real. These two networks are trained in such a way that the generator model tries to imitate the input data as close as possible with minimum noise. The discriminator model is trained to effectively find out the fake images. Many variations on GAN have been proposed since it is introduced, making it more powerful and suitable for synthetic image generative tasks.

GANs have a good performance in the image generation field when compared to the traditional and CNN methods. Image steganography can be considered as one such image generation task where two inputs – the cover image and the secret image are given to generate one output – stego image. The existing methods used for image steganography using a GAN architecture can be grouped into five categories - a three network based GAN model, cycle-GAN based architectures, sender-receiver architecture using GAN, coverless model where the cover image is generated randomly instead of being given as input.
Generally, a GAN model consists of two main components: the generator and the discriminator. A new network named, the steganalyzer is introduced in some of the methods in Image steganography. The main functions of these three components of GAN model are as follows

1. A generator model, G, to generate stego images from the cover image and the random message.
2. A discriminator model, D, to classify the generated images from the generator as either real or fake.
3. A steganalyzer, S, to check is the input image has a confidential secret data or not.

4. DATASETS USED

There exist one data set, BOSSBase, that was specifically created to deal with the problems of steganography. To further evaluate the performances of the algorithms some existing datasets, which are used for other purposes including object recognition and face recognition, are re-modeled to fit for the purpose of our experiments.

4.1 BOSSBase

Break Our Steganographic System (BOSS) is the first scientific challenge conducted. It took the image steganography from being a research topic to a practical application. The main aim of the competition was to develop a better Steganalysis method that can break the steganographic images created by the HUGO (Highly Undetectable stego) algorithm. The data set consists of a training set and testing set along with the HUGO algorithm that can be used to create the steganography images. The training data set consists of 10,000 grayscale cover images with dimensions 512×512 . The testing set consists of 1000 grayscale images with dimensions 512×512 . There is an option to download the datasets with steganography images solely for the purpose of Steganalysis.

4.2 CelebA

Large-scale CelebFaces Attributes data set, also known as CelebA data set, is a vast data set. It has more than 200k images that can be used for face recognition, face detection, face localization and other face-related operations. These data set consists of images from various sources, locations, background and poses and is best suitable for steganography. The probability of using a photo/face image as the cover for hiding secret images is very high.There are 40 different annotations available like with/without glasses, emotions, hair styles, other accessories like hat besides the images.

4.3 ImageNet

ImageNet is also a very large dataset containing images from the WordNet hierarchy. Each node here contains more than 500 to 1000 images. ImageNet contains only the links or thumbnails to the original image and does not have any copyrights to the image The data set of ImageNet consists of images of varying size.The number of images, classes they belong to, background and the image size can be selected from the wide range available, based on the requirements.

4.4 MNIST Handwritten Digits

Modified National Institute of Standards and Technology database (MNIST) is another data set that can be used for various image processing applications and computer vision. MNIST handwritten data set consists of a training and testing set with images of handwritten digits 0 to 9. Images in this data set are normalized, black and white with dimensions 28×28 pixels. The training set consists of 60,000 images and testing set consists of 10,000 images.

4.5 COCO

Common Objects in Context (COCO) data set was mainly developed for object detection, segmentation and image captioning purposes. It is also a huge data set with images from 80 object categories. Each class contains at least 5 images. This data set comes along with the class annotation and the segmentation annotation and there is no predefined training and testing split. The data set split can be carried out based on the research topic and the user convenience.

5. CHALLENGES

The following are some noted challenges for consideration in image steganography problems.

5.1 Data Availability

Though image steganography belongs to unsupervised learning and its main goal is image reconstruction, there is no proper benchmark data set available for it except BOSSBase. Most of the methods deal with hiding RGB images inside RGB cover images. Finding a suitable dataset can be challenging. ImageNet is the most commonly used dataset with a major drawback being the image size. The images are very small in size of 64×64.

5.2 Real-time steganography

Steganography models are trained on a huge amounts of datasets, like in, 45000 training images are used. However, when it comes to the real-time steganography, it gets difficult. The implementation of the trained model for performing the steganography and steganalysis requires transferring the stego image through an untrusted channel to the receiving end. The capability of the trained model in dealing with real time live images which may contain noises, skewing, blurring is not proved. The implementation of the model for real-time steganography is still questionable.

6. CONCLUSION

Image steganography is the method used in transmitting secret information by hiding it in plain sight inside a cover image. Deep learning methods are widely used in every field and has been used in the research of steganography. Review of all the related works led to categorizing them into three groups vastly. Most of the traditional based steganography methods use the LSB substitution and some of its variants. The hiding capacity of the traditional methods are limited as over burdening the cover image by exploiting more pixels for hiding the secret message.

Steganalysis is an effective mechanism to analyses the embedding performance of steganographic techniques. This paper presents a review of the main Steganalysis techniques for detecting secret data.There is no benchmark image datasets to perform the image steganography while most of them use the ImageNet, CelebA or BOSSBase. Each of the methods have their own evaluation methods and metrics and hence there is no common platform for comparisons. Traditional methods are less secure as it is only a matter of detection of presence of the secret message.

We presented the strengths and weaknesses of various stego-systems in terms of Steganalysis. A steganalyst is frequently interested in more than whether or not a secret message is present. The ultimate goal is to extract the secret message. However, in the absence of the knowledge of the stego technique and the stego image, the process of detection and analysis can be time consuming or completely infeasible.In summary, this paper has elaborated on the techniques used in the recent times for image steganography, the current trends. Along with it, details on the datasets and evaluation metrics are detailed. Challenges faced is also evaluated in this paper. It can be concluded that deep learning has tremendous potential in the image steganography field taking into consideration that all the challenges are filled.

7. REFERENCES

[1]. *Steganography*, 2020, [online] Available: https://en.wikipedia.org/wiki/Steganography
[2]. H. Shi, X.-Y. Zhang, S. Wang, G. Fu and J. Tang, "Synchronized detection and recovery of steganographic messages with adversarial learning", *Proc. Int. Conf. Comput. Sci*, pp. 31-43, 2019.
[3]. N. F. Hordri, S. S. Yuhaniz and S. M. Shamsuddin, "Deep learning and its applications: A review", *Proc. Conf. Postgraduate Annu. Res. Informat. Seminar*, pp. 1-6, 2016.
[4]. N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen", *Computer*, vol. 31, no. 2, pp. 26-34, Feb. 1998.