Data Compression and Decompression with Image steganography using api

Aadesh Garg, Amisha Prajapati, Amisha linjhara, Ananya Shrivastava

Department of Computer Science and Engineering, Acropolis Institute Of Technology and Research, Indore, Madhya Pradesh, India

Abstract: Image Steganography is the process of hiding information which can be text, image or video inside a cover image. The secret information is hidden in a way that it not visible to the human eyes. Deep learning technology, which has emerged as a powerful tool in various applications including image steganography, has received increased attention recently. The main goal of this paper is to explore and discuss various deep learning methods available in image steganography field.

This paper presents a review kind of data compression techniques. Data compression is widely used by the community because through a compression we can save storage. Data compression can also speed up a transmission of data from one person to another. In performing a compression requires a method of data compression that can be used, the method can then be used to compress a data. Data that can be compressed not only text data but can be images and video.

Keyboard: Image steganography, GAN steganography, CNN steganography, information hiding, image data hiding

Data Compression, compression techniques, lossless compression, Huffman, Shannon Fano, Tunstall, RLE, LZW.

Date of Submission: 23-03-2023

I. INTRODUCTION

Technology has blitz scaled over the past years leading to a wide usage of multimedia for transferring data, especially Internet of Things (IoT). Usually, the transfer happens over insecure network channels. In particular, the internet has gained accelerated popularity for exchanging digital media and individuals, private companies, institutions, governments use these multimedia data transfer methods for exchanging data. Though there are numerous advantages attached with it, one prominent disadvantage is the privacy and security of the data. The availability of numerous readily available tools capable of exploiting the privacy, data integrity and security of the data being transmitted has made the possibility of malicious threats, eavesdropping and other subversive activ ities. The prominent solution is data encryption where the data is converted into a cipher text domain using encryption key. At the receiving end, the cipher text is converted into plain text using a decryption key.

With the rapid development of technology with the support of software and hardware that increasingly facilitate widespread information quickly through the internet around the world. Information obtained can be sent easily via the internet as a the medium of communication for information technology experts. However, not all information can be sent easily. There is a large size that can hinder data transmission quickly and save on existing storage in the computer. To overcome the problem of Compression is the reduction of a file size from a large size to a smaller file size. A compression will be done to facilitate the transmission of a file with a large size and contains many characters. The workings of a compression are by looking for patterns of repetition in the data and replace it with a certain sign. The type of compression has two methods, lossless compression and lossy compression

II. Summary of the Method

After reviewing all the frameworks available, the method ologies are primarily grouped into three categories, namely, traditional image steganography methods, CNN-based image steganography methods and GAN-based image steganogra phy methods. Traditional methods are frameworks which use methods that are not related to machine learning or deep learning algorithms. Many traditional methods are based on the LSB technique. CNN-based methods are based on deep convolutional neural networks for embedding and extracting the secret messages and GAN-based methods use some of the GAN variants. Figure 1 gives an overview of a steganography and steganalysis architecture.

Automatic Steganographic Distortion Learning framework with GAN (ASDL-GAN for short) was introduced by Tang et al. in [43]. In this architecture, the generator is used to learn the probabilities for each pixel from the input cover image and the authors have proposed to use a novel activation function Ternary Embedding Simulator (TES) for generating the stego images from the generated probabilities. The discriminator helps in differentiating between the real and fake images. XuNet based architecture is used for discriminator D.

The general working principle is that Eve eavesdrops between Alice and Bob to check if there are any secret message embedded in the communication channel between them. The authors in [40] have used neural networks to train all the three components. For the steganographic scenario, Alice is trained to create the steganographic image while Bob recovers the secret message from the stego images. Eve helps Bob by giving the probability of the given image being a stego image. A model with four parts – Alice, Bob, Dev and Eve has been proposed by Wang et al. in [59]. Since the model is an unsupervised generative model, the authors have named the model Self-supervised Steganographic GAN (SSteGAN). Like in any other communication security paradigm, Alice and Bob try to communicate secretly while Eve eavesdrops on the communication channel.

The data compression can dramatically decrease the amount of storage a file takes up. For instance, in a 2:1 compression ratio, a 20 megabyte (MB) file takes up 10 MB of space. In the result of compression, administrators spend less money and less time on storage.Data compression optimizes backup storage performance and has recently shown up in primary storage data reduction. The compression will be an important method of data reduction as data continues to grow exponentially.Usually any type of file can be compressed, but it's important to follow best practices when choosing which ones to compress. In general, some files may already come compressed, so compressing those files would not have a significant impact

The Run-length encoding (RLE) is a very simple form of data compression in which runs of data are stored as a single data value and count, rather than as the original run. This technique is the most useful on data that contains many such runs: for example, simple graphic images such as icons and line drawings. Take thisexample; consider a screen containing plain black text on a solid white background. This will be many long runs of white pixels in the blank space, and many short runs of black pixels within the text. For example, take a hypothetical single scan line, with B representing a black pixel and W representingwhite:

WWWWWWWWWWWWBWWWWWWWWWWWWBBBWWWWWWWWWWWWWWWWWWWWWWWWWB

Here we apply a simple run-length code to the above hypothetical scan line, we get the following:

12WB12W3B24WB

III. Compresssion Versus Data Deduplication

The Compression is often compared to data deduplication, but the two techniques operate differently. Deduplication is a type of compression that looks for redundant chunks of data across a storageor file system and then replaces each duplicate chunk with a pointer to the original. The Data compression algorithms reduce the size of the bit strings in a data stream that is far smaller in scope and generally remembers no more than the last megabyte or less of data.The File-level deduplication eliminates redundant files and replaces them with stubs pointing to the original file. The Block-level deduplication identifies duplicate data at the subfile level. This system saves unique instances of each block, uses a hash algorithm to process them and generates a unique identifier to store them in an index. The Deduplication typically looks for larger chunks of duplicate data than compression, and systems can deduplicate using a fixed or variable-sized chunk. The Deduplication is most effective in environments that have a high degree of redundant data, such as virtual desktop infrastructure or storage backup systems. The Data compression tends to be more effective than deduplication in reducing the size of unique information, such as images, audio,videos, databases and executable files. Many storage systems support both compression and deduplication.

IV. Dependencies

The major requirement for the resources for designing and developing the proposed smart map is as follows.

- HTML
- CSS
- Java
- MySQL

HTML: It stands for Hyper Text Markup Language. It is the standard markup language for creating web pages. It describes the structure of a web page. HTML consists of a series of elements. HTML elements tell the browser how to display the content.

CSS: It stands for Cascading Style Sheets. It describes how HTML elements are to be displayed on the screen, on paper, or in other media. It can control the layout of multiple web pages all at once and saves a lot of work. External stylesheets are stored CSSfile

IV.Conclusion: The advantages of compression are a reduction in storage hardware, data transmission time and communication bandwidth -- and the resulting cost savings. The compressed file also requires less time for transfer, and it consumes less network bandwidth than an uncompressed file. The important disadvantage of data compression is the performance impact resulting from the use of CPU and memory resources to compress the data and perform decompression. The Lossless compression techniques, as their name implies, involve no loss of information. Even, If data have been losslessly compressed, the original data can be recovered exactly from the compressed data after a compress/expand cycle.

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. Other than LSB, PVD, DCT and EMD are commonly used. The hiding capacity of the traditional methods are limited as over burdening the cover image by exploiting more pixels for hiding the secret message may led to distortions. Also, the autoencoder-decoder structure with VGG as base, U-Net and Xu-Net are the most prevailing architectures used for CNN-based image steganography methods. More recently, GAN architecture has gained significant attention for their ability to deal with image reconstruction tasks. Image steganography can be considered one such image reconstruction task where the cover image and the secret information is taken as input to reconstruct a steganographic image which is close to the cover image in resemblance

V. REFERENCES

[1] Wikipedia. (2020). Steganography. [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,'' in Proc. Int. Conf. Comput. Sci. Cham, Switzerland: Springer, 2019, pp. 31–43

[3] Lung-Jen Lee, Wang-Dauh Tseng, Rung-Bin Lin, and Cheng-Ho Chang, “ Pattern Run-Length for Test Data compression”, IEEE Transaction on Computer-Aded Design of Integrated Circuits And System, Vol.31, No.4,April,2012.

[4] Mohammad Arif, R.S.Anand, “Run Length Encoding for Speech Data Comprassion”, IEEE International Conference on Computional Intelligence and Computing Research, 2012.