International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 09 | September - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Multi-Stage Hiding of Image-into-Audio Steganography using CNN

Mrs. T Madhavi Kumari [1], Dr. A. Vinaya Babu[2], V. V. Chandra Teja[3]

[1](ECE, JNTUH College of Engineering Hyderabad, India)
[2](Rtd. Prof. Of CSE, JNTU & Dean Academics, Stanley College of Engg. Tech, Hyderabad, India)
[3](ECE, JNTUH College of Engineering Hyderabad, India)

Abstract – Sometimes a file consists of more information than it appears. A file may appear normal to an untrained eye, but expert recipients can extract more information from it. The increasing use of audio communication technology has accelerated the transfer of audio data via the Internet, making it a preferred carrier for covert communication. In this paper we propose a deep neural network based frame work to perform audio steganography. There are mainly two phases present, embedding phase and extraction phase. In the embedding phase the image contents are hidden in the audio carrier and in the extraction phase the hidden image is extracted from the embedded audio. The suggested framework will hide the secret message in a progressive manner, making the concealing process easier and capable of outperforming competing algorithms.

Key Words: Audio steganography, Convolutional neural networks.

1.INTRODUCTION

Now a days, information security is critical in data communication domains. It also requires protection against unauthorised access during data transferring. Cryptography and steganography are the techniques used for information hiding and security. In this we mainly deal with the steganography technique. It is the art and science of embedding hidden messages in such a way that no one, apart from the sender and receiver, suspects the presence of message. It deals with hiding secret data in some cover media which may be an image, audio or a video. Steganography can also be used to conceal text, video, photos, or even audio data. It's the useful piece of information, limited only by the medium and the author's imagination.

The advantage of steganography compared to cryptography is that the intended hidden message is not drawn to itself as an object of inspection. Plainly visible encrypted information, no matter how impenetrable, provoke suspicion and may be incriminating in nations where encryption is outlawed. Whereas cryptography protects only the information present in a message, steganography is concerned with concealing both the fact that secret information is being transmitted and along with its contents.

Electronic communications in digital steganography may contain steganography coding within a transport layer, such as a document file, image file, software, or protocol. Because of their wide range and vast size, media files are the most suitable for steganography transmission. For example, a sender can begin with a basic image file and change the colour of every eighteenth pixel to corresponding to a letter of the alphabet. The change of image is so subtle that anyone who is not looking for it is unlikely to see it. The study of hiding data in sound is called as audio steganography. It safeguards against illegal reproduction when used digitally. Watermarking is a technique in which one piece of data (the message) is encrypted within another (the "carrier"). Its most common applications involves the media playback, mainly audio clips.

A vast range of steganography setups and procedures have been presented to attain perfect hiding performance. Deep network-based steganography algorithms outperformed hand-crafted embedding methods in terms of performance. A Deep neural network-related Image-To-Audio Steganography (D.I.T.A.S) framework is suggested to address the disadvantages of various other approaches. The main moto of this work is to hide image in audio cover file i.e performing audio steganography and we need to use the concept of CNN as well. A multistage network is designed in the framework: the network encodes the decreasing multilevel residuals of the image inside distinct audio subsequence and decoding of residuals from the modified carrier is done to produce the final revealed results that is secret image. Let us see how it is implemented in the next section.

2. PROPOSED METHODOLOGY

In this paper, we present a newer audio steganography approach for secure hiding of images by using convolutional neural networks. This section elaborates how our approach is implemented.

As mentioned earlier the project is divided mainly into two phases. One is the embedding phase and followed by extraction phase. It can be clearly understood by seeing the block diagram of proposed method shown in fig.(1) . Firstly we need to design a convolutional neural network in a multistage fashion .

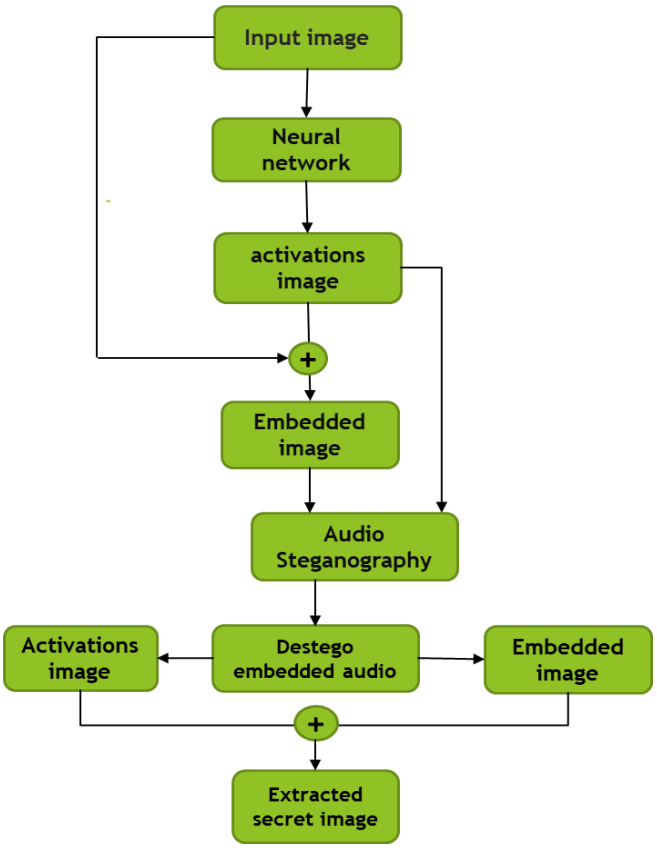

Figure.1 Proposed method Block Diagram.

2.1 Embedding Phase:

The first phase of this project is hiding the image into the audio carrier. For that an input secret image is chosen and is given as input to the convolutional neural network. The convolutional neural network architecture is shown in fig.(2) .

There are 4 stages in the network. Each stage consists of 3 convolutional 2d layers,1 batch normalization layer,1 relu layer,1 max-pooling layer. The input layer to the network is the image input layer and output layers are fc layer, softmax layer and class-output layer. All the convolutional layers and max-pooling layers have same padding and stride equal to one.In the first stage there are 8 filters of size 3x3 in the convolutional layers and 12 filters in the second stages,14 filters in the third stage and 16 filters in the fourth stage all are of same size 3x3.

The network is trained with stochastic gradient descent with momentum (SGDM). The image dataset used for training is VOC2012 dataset. It contains wide range of categories of images close to 20 that includes animals, birds, humans, trains, aeroplanes, tables, boats, dogs, cats, sheep etc.

In order to match the dimensionality constraint of audio and image, we perform FFT to the cover audio and upscale the magnitude by a scale of 1000000 and convert it into binary. LSB steganography is performed to hide the images in the cover audio.

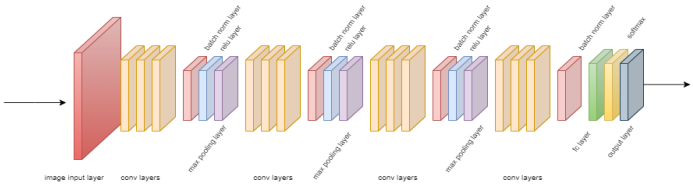

Figure.2 CNN Architecture

Step-1: First an input RGB image of size 120x230 is given as input to the convolutional neural network. The architecture of the network is shown in fig.(2).
Step-2: After giving input secret image to the network we collect the residual images which are nothing but the activations images from the convolutional layers of the each stage of the network.
Step-3: Now the obtained activation image is bitwise XOR-ed with the input secret image and we get embedded image which is the XOR image.
Step-4: Choose an cover audio of sufficient length which is greater than the size of images to be hidden.
Step-5: The original secret image and XOR image are now hidden in the cover audio using LSB steganography.

After performing hiding operation, again the embedded audio is retrieved back to its original format by performing above actions in reverse manner. It is done so that there will not be any major change in embedded audio and cover audio , even if there is some minute change in the audio it is unnoticeable by the human ear .Till here it is embedding process. Now we need to extract the images from the embedded audio in order to get back the secret image.

2.2 Extraction Phase:

The next phase of this work is to extract the secret image from the embedded audio file. The embedded audio file is processed in the following steps.

Step-1: We performed LSB steganography for hiding the images, so during extraction of the image the lsb bits are collected from the embedded audio.
Step-2: The LSB bits are collected and reshaped according to the size of images consecutively one after another.
Step-3: Now the extracted XOR image and activation image are again bitwise XOR-ed to produce to the final revealed result ie the secret image.

This is implementation procedure of our proposed work and results are shown in the upcoming section.

3.RESULTS AND ANALYSIS

For the process of evaluation, in this project 4 categories of images are used. The 4 categories are: Animals, Birds, Humans, Others.In the other category there are images of trains, planes, vehicles, scenery pictures, etc.The cover audio is mono channel uncompressed .wav file sampled at 44.1Khz frequency with 16 bits per sample. Only .wav files with the specified configuration are used in this project.

Figure 3.1(a)

Figure 3.1(b)

Figure 3.1(c)

Figure 3.2(d)

Figure 3.1 Input images:(a)Animal,(b)Bird,(c)Human and (d)other.

Figure 3.2(a)

Figure 3.2(b)

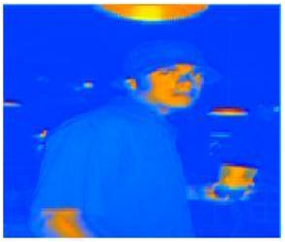

Figure 3.2(c)

Figure 3.2(d)

Figure3.2 Stage-1activation images:(a)animal,(b)bird,(c)human and (d)other

Figure 3.3(a)

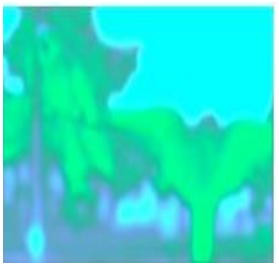

Figure3.3(b)

Figure 3.3(c)

Figure 3.3(d)

Figure 3.3 Stage-2 Activation images:(a)animal,(b)bird,(c)human,(d)other

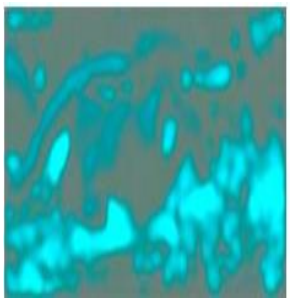

Figure 3.4(a)

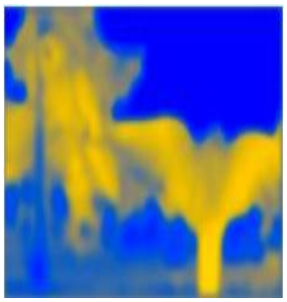

Figure 3.4(b)

Figure 3.4(c)

Figure 3.4(d)

Figure 3.4 Stage-3 Activation images:(a)animal,(b)bird,(c)human,(d)other

The whole project is done on matlab software. The network is trained with initial learn rate 1e-4 and dropped by a factor 2 for every 5 epochs. The entire network is trained for 150 epochs with a mini batch size of 40.

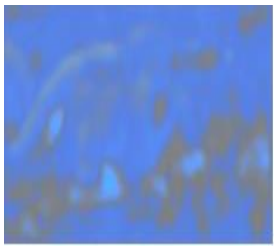

Figure 3.5 (a)

Figure 3.5 (b)

Figure 3.5 (c)

Figure 3.5 (d)

Figure 3.5 Stage-4 activation images:(a)animal,(b)bird,(c)human,(d)other

The activation images obtained from different stages are now performed XOR operation with input image. For that any stage activation image can be chosen and that is XOR-ed with the input image. In figure 3.6 stage-1 activation image and input image XOR result is shown.

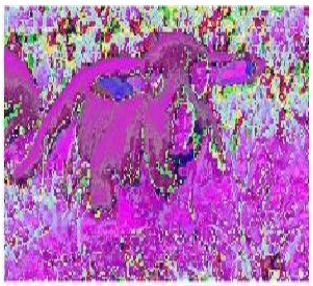

Figure 3.6(a)

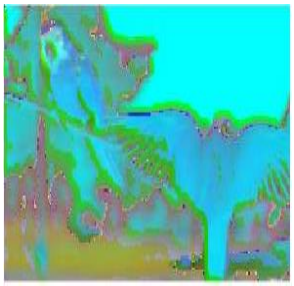

Figure 3.6(b)

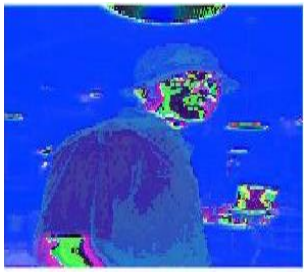

Figure 3.6(c)

Figure 3.6(d)

Figure 3.6 XOR images of 1st stage activation and input images:(a)animal,(b)bird,(c)human,(d)other.

Now the xor image and the activation image are embedded in the cover audio as mentioned in section 2. The extraction phase results are shown in figure 3.7 and 3.8.

Figure 3.7 (a)

Figure 3.7(b)

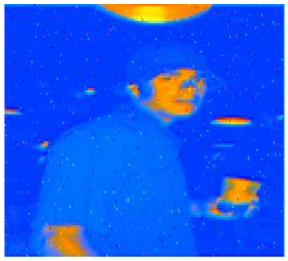

Figure 3.7(c)

Figure 3.7(d)

Figure 3.7 Extracted 1st stage activation images:(a)animal,(b)bird,(c)human,(d)other

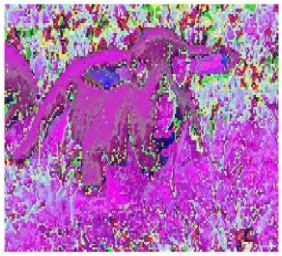

Figure 3.8 (a)

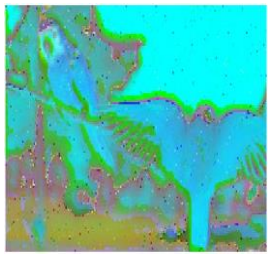

Figure3.8 (b)

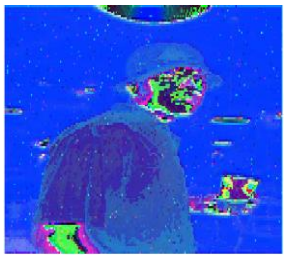

Figure 3.8 (c)

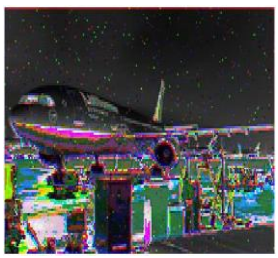

Figure 3.8 (d)

Figure 3.8 Extracted xor images:(a)animal,(b)bird,(c)human,(d)other.

The final revealed output image result is shown in figure 3.9.

Figure 3.9 (a)

Figure 3.9 (b)

Figure 3.9 (c)

Figure 3.9 (d)

Figure 3.9 Final Revealed images:(a)animal,(b)bird,(c)human,(d)other

To analyze the results obtained mainly 3 statistical parameters are chosen they are (PSNR,MSE,SSIM) to measure revealed output image. Table 1 shows the PSNR and MSE values of input and revealed images.

Image Category	MSE	PSNR
ANIMALS	1.0117	28.08
BIRDS	1.0055	28.11
HUMANS	1.0505	27.92
OTHERS	1.0552	27.90

Table.1 PSNR and MSE values of input and Revealed Image

According to the concept, the greater the PSNR, the better the degraded image has been reconstructed to resemble the original image and the more effective the reconstructive method. Table 2 shows the SSIM value. SSIM is used for computing the visual difference between two given images. Here we compute SSIM between input image and revealed image. SSIM formula is given below:

$$\text{SSIM}\,(I_1, I_2) = \frac{(2\mu_{I_1}\mu_{I_2}+\alpha)(2\sigma_{I_1I_2}+\beta)}{(\mu_{I_1}^2+\mu_{I_2}^2+\alpha)(\sigma_{I_1}^2+\sigma_{I_2}^2+\beta)} \qquad (1)$$

Here I_1 and I_2 are input and revealed images respectively. μ_{I_1} *and* μ_{I_2} are averages of I_1 and I_2 respectively. $\sigma_{I_1I_2}$ is the covariance between I_1 and I_2. $\sigma_{I_1}^2$ *and* $\sigma_{I_2}^2$ are the variances of I_1 and I_2 respectively. SSIM value should be close to 1 between input and revealed images.

Category	SSIM
Animals	0.94
Birds	0.93
Humans	0.96
Others	0.98

Table.2 SSIM values between input and revealed image.

4.CONCLUSIONS

This paper has presented a deep learning technique for Multi-Stage Residual Hiding for Image-into-Audio Steganography. For this model, we have used CNN (Convolution Neural Network) Network for the process of finding the activations of the given secret image that further which are used for hiding into the audio. By hiding the image contents at multiple level, the proposed method not only controls payload capacity more flexibly, but it also makes the hiding process easier. As we embed different stage input images , even if some part of the carrier is lost , the secret image can be restored to an certain extent.

5.FUTURESCOPE

In future, we can improve results of Steganography. And also we can extend this concept of hiding the information in different forms that may include an image, text, audio or video as a secret message and can be embedded with any type of data and revealed at only receiver side. In this project only .wav files are used as cover audio, in future various other formats should also be compatible and should give satisfactory results. This hiding process can be used in different fields for hiding secret information mainly in military applications.

REFERENCES

[1] Shumeet Baluja, "Hiding images in plain sight: Deep steganography," Advances in Neural Information Processing Systems (NIPS), pp. 2069–2079, 2017.

[2] Eric Cole and Ronald D. Krutz, "Hiding in plain sight: Steganography and the art of covert communication," 2003.

[3] Ron G. Van Schyndel, Andrew Z. Tirkel, and Charles F. Osborne, "A digital watermark," in IEEE International Conference on Image Processing (ICIP), 1994.

[4] Raymond B. Wolfgang and Edward J. Delp, "A watermark for digital images," IEEE International Conference on Image Processing (ICIP), 1996.

[5] M. Asad, J. Gilani, and A. Khalid, "An enhanced least significant bit modification technique for audio steganography," in International Conference on Computer Networks and Information Technology, 2011, pp. 143–147.

[6] Tom Pevn, Tom Filler, and Patrick Bas, "Using high dimensional image models to perform highly undetectable steganography," vol. 6387, pp. 161–177, 2010.

[7] Vojtch Holub and Jessica Fridrich, "Designing steganography distortion using directional filters," in IEEE Workshop on Information Forensic and Security, 2012.

[8] Vojtch Holub, Jessica Fridrich, and Tom Denemark, "Universal distortion function for steganography in an arbitrary domain," Eurasip Journal on Information Security, pp. 1–13, 2014.

[9] Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei- Fei, "Hidden: Hiding data with deep networks," Proceedings of the European Conference on Computer Vision (ECCV), pp. 657–672, 2018.

[10] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Xu Bing, and Yoshua Bengio, "Generative adversarial nets," in International Conference on Neural Information Processing Systems (NIPS), 2014.

[11] Denis Volkhonskiy, Ivan Nazarov, and Evgeny Burnaev, "Steganographic generative adversarial networks," 2017.

[12] Haichao Shi, Jing Dong, Wei Wang, Yinlong Qian, and Xiaoyu Zhang, "Ssgan: Secure steganography based on generative adversarial networks," 2017.

[13] Jamie Hayes and George Danezis, "Generating steganographic images via adversarial training," in Advances in Neural Information Processing Systems 30, pp. 1954–1963. Curran Associates, Inc., 2017.

[14] Atique Ur Rehman, Rafia Rahim, M Shahroz Nadeem, and Sibt Ul Hussain, "End-to-end trained cnn encodedecoder networks for image steganography," 2017.

[15] Pin Wu, Yang Yang, and Xiaoqiang Li, "Imageinto- image steganography using deep convolutional network," in Advances in Multimedia Information Processing - PCM 2018 - 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings,Part II. 2018, vol. 11165 of Lecture Notes in Computer Science, pp. 792–802, Springer.