Video Steganography using Convolutional Neural Network and Temporal Residual Method

Fariha Aiman
M.Tech Student
Department of CSE
JNNCE, Shimoga

G. R. Manjula, PhD
Associate Professor
Department of CSE
JNNCE, Shimoga

ABSTRACT
Steganography involves data hiding in computer files. Steganographic coding, include a file including the document, image, program as well as guidelines, can be included in electronic communication in a transport layer. Because of their large size media files are suitable for the steganography transmission. This paper focuses on video steganography. Video steganography is nothing but hiding the complete secret video within the cover video. Firstly the residual of the secret video and cover video is obtained because hiding the residual video is much easier when compared to hiding original video. This model uses the deep convolutional neural network method. The model compares the particular model with other method. And all results show that this method is efficient.

Keywords
Video steganography, Data hiding, deep neural network, residual modeling

1. INTRODUCTION
The term steganography can accelerate to some earlier technique flourished in the 15th century. Steganography's aim is to conceal a hidden message in some transportation medium and interact underhandedly with a possible recipient who will know the decoding rule [1]. Steganography, which is necessarily distinct from cryptography, focuses on protecting confidential messages, enabling just the destination receiver to understand. In other words, the protecting channel could be noticeable to the public, yet somehow the receiving point can detect the existence and decipher the hidden information. To reality, some steganographic system can hide confidential data through simultaneously maximizing 2 requirements: reducing its shift well into the concealing form of media that can lead to the doubt of an enemy and minimizing residual among both decrypted confidential data as well as its proof [1]. Steganography analyzes have practical indication. For instance, there are also many criminal applications for steganography methods, for example, conceal orders which correlate illegal actions from pictures displayed on the social networking sites. Visual steganography techniques are the main scope of this work that hides a complete color image / video within another. The work has technical contributions in two-fold: First, the residual will be zero at most pixels between the two consecutive frames [1]. Highly scarce information makes it much easier to hide than to hide the initial frames. Inspired by this reality, it is suggested that inter-frame residuals be explicitly examined on each and every video frame instead of blindly implementing picture steganography model. The model comprises two areas in particular, one of which is specifically designed to hide the inter-frame disparity inside the cover video image and another just conceals the initial secret video [1]. A straightforward thresholding method looks at which branch should choose a hidden video frame. When the hidden secret video is revealed, two decoders are split, respectively showing difference or frame. Secondly, create the model based on profound convolutionary neural networks that is unprecedented in video steganography literature.

2. USES OF STEGANOGRAPHY
Steganography is useful for storing delicate information, such as hiding passwords of the system or keys in other documents. It is possible to apply steganography to audio, picture and video files. It can be used to hide military secrets in multiple areas such as banking, public industries, and also. Not every use of steganography is poor. One can insert watermarks into identity documents, by making duplicating the card more difficult for a counterfeiter [5]. Another positive is that it is possible to carry more safely completely legal, confidential data. In addition, businesses capitalize on the method to increase the security of everyday company [5]. Digimarc Corporation, for instance, a major provider of safe media solutions, offers governments around the globe with secure watermarking identification solutions.

3. FUTURE OF STEGANOGRAPHY
Steganography is still improving. New apps need to be developed with each discovery of a novel steganography format. These steganography advances have taken us to the techniques of inserting information into pictures, records and sound today. With today's computer steganography, it has become more complex to find and decode the hidden information. Steganalysts are presently operating difficult to locate concealed texts in images, documents, and noise. Steganalysis begins with files suspected of information. In order to help decrease the amount of documents, the steganalyst utilizes forensic statistical data. The researcher then contrasts information documents that are dubious with comparable information documents. The resemblance is focused on the very same digital camera or electronic recording system. The researcher examines image identification, audio identification, Quantitative identification (variations for image models or LSB) as well as evaluation of histograms, as well as organizational identification (viewing data assets /material content, change in volume, change in date / time, information – modifications, checksum). Upon detection of steganography and extraction of the data, it may still be encoded. Cryptanalysis methods can be implemented at this stage. Steganalyst ' fight against the concealed information has just begun. Much more needs to be performed to detect the hazardous information concealed behind the innocent images.

Steganography will continue to increase in the future as safety demands expand. The safety problems for "The Good" and "The Bad" are the same: "Information needs to be concealed." Simultaneously, both will operate to decode the fresh types of

steganography in order to acquire the information of the other. Newer forms of steganography and steganalysis will be created in any manner you look at it.

4. CONVOLUTIONAL NEURAL NETWORK

Convolutionary neural network (Conv Net or CNN) is very common deep learning method, a form of computer training that a model teaches straight from pictures, video, text, or sound to conduct classification functions. CNNs are especially helpful in discovering patterns for recognizing items, Face and images in pictures. It learns with picture information straight, using models to rank images and eliminate the need to extract detailed characteristics. Of course, neural networks are not new. A fast search will produce academic papers from the 1940s. The convolutionary flavor, however, is newer and has grown in popularity in latest years owing to a renewed focus on deep learning. Historically, Conv Nets were used to process image data in an academic field known as *computer vision*. ConvNets are noteworthy for how well they lend themselves to the task of recognizing image features, such as the faces of people, or the continuously updated external view of the environment in a self-driving car. ConvNets are also helpful for processing natural language (unstructured data) and for recognition of optical characters (OCR).

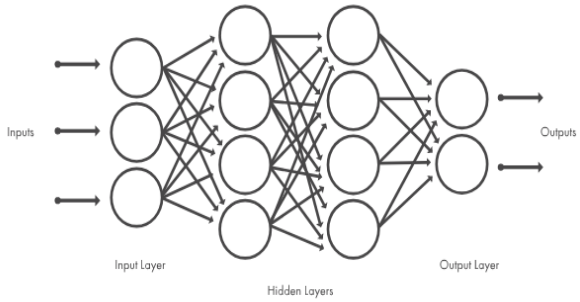

Figure 1 CNN layers

These levels conduct activities that change information with data-specific teaching characteristics. Three main layers are convolution layer, activation layer or ReLU layer, and pooling layer.

- **Convolutionary layer** positions entry images via a collection of customer-friendly filters, each activating some picture features.
- **Rectified linear unit (ReLU)** makes training quicker and efficient by converting adverse attributes to null and retaining favorable scores. It is sometimes related to as activating, because only enabled characteristics will be transmitted to another level.
- **Pooling** optimizes production by sampling variables down; decreasing the amount of variables that the network needs to know.

Over tens or hundreds of layers, these operations are repeated learning to distinguish distinct characteristics with each coating.

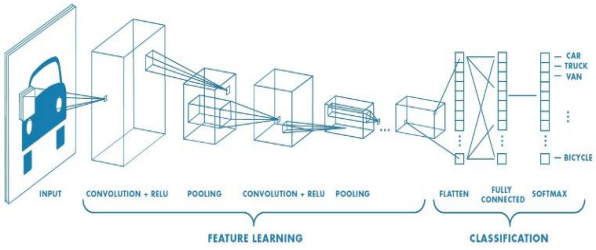

Figure 2 Network with many convolutional layers

Figure 2 demonstrates an instance of a multi-layered network. Filters are applied at distinct sizes to each training picture, and each converted image's output can be used for next layer as the input.

5. PROBLEM STATEMENT

This work focuses on video steganography. Usually, video consists of many frames and huge data and takes lot of time for embedding so to overcome this problem, convolutional video steganography using temporal residual method is proposed.

6. OBJECTIVES

- The project's goal is to provide a safe means of data communication using temporal residual technique.
- Its main aim of this job is video steganography, its job is concealing a full-size clip in an video.
- Convolutional video steganography using temporal residual method can be used to obtain security and privacy of the data.
- The main aim is to hide the video inside the cover video so that it remains unobserved and undetected or intact.
- And one of the objectives is to reduce the time for embedding.

7. PROPOSED METHOD

A novel solution with temporal residual modeling is used in this project for convolutionary video steganography. Briefly, it consists of two technical contributions: First, there are very few residuals between two consecutive images. Objectively, hiding this scarce remnant in another video image depicts an even easy task than hiding frame in another frame. Inspired by such a reality, rather than wrongly implementing steganography method to all images, it is suggested that frames be divided in to the two parts: reference frame and remaining frame. Every residual frame can be acquired through differentiation by a particular frame of reference. Accordingly, this system involves two sections at both the encryption and decryption phases, addressing both types of frame.

Validating this therapy empirically can considerably boost the perceptual quality of the container and increase the possibility of tricking the opponent. Second, this system is focused completely on deep cellular convolution networks that are a one of a type in video steganography. This profound video steganography system specifically comprises of 2 H-nets to conceal reference or residual and 2 R-nets to expose hidden clip. Without human annotations, complete system can be taught and variables of the network can be designed from start. Detailed tests are held out in tests to verify the powerful design of profound systems.

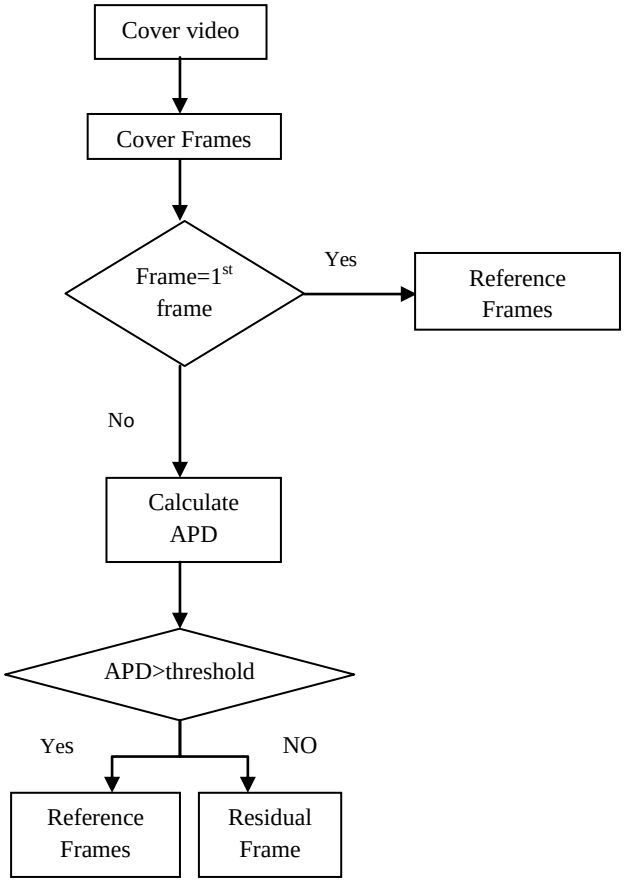

Figure 3 Classification of Reference and Residual frames

First, it is necessary to select the cover video, then select the secret video to be hidden, then display the cover video with its residual video, then display the secret video with the residual video, then display the stego video, then finally extract the video. A first image is generally held as the frame of reference. Then for labeling of next frames APD value is calculated if the value of the APD is higher than the limit value then the frame will be marked as a reference frame otherwise the frame will be regarded as a remaining frame.

The model consists of five computing phases:

Step-1: Labeling of Reference / Residual Frame: This job utilizes a simple threshold approach to label a picture as a reference or residual form. The very first image of a clip is definitely labeled as a reference. With regard to first frame, the preceding frames in the very same video linearly measure the average pixel-wise discrepancy (APD) .Once any frame's APD score exceeds a certain specified limit, it would be established as a fresh reference and then utilized to measure all later images. Operation goes on till the labeling of every frame [1].

Step-2: Concealing confidential data: A divide-and-conquer system is the main differentiator of this technique for others. Note that two hiding networks, called Reference H-network or Residual H-network, are designed. Each frame is supplied by their label into the respective H-net. It should be explained that no parameter is shared by these two H-nets. They are optimized separately for encoding only particular frame types. This work describes the fresh frame as container, which looks comparable to that of cover, and somewhere hides a secret. Select the U-network system for both H-networks in reality [1].

Step-3: Secret Revealing (decoding): An entry is simply the container, and the result (called mystery decoded) is another picture required to be precisely the truth in the ideal situation. Or else model will be taught to reduce the inconsistency between the truth and its decoded form. 2 r-networks will be launched similar as H-nets to expose the frame or residual secret. In other words, the decoder does not know which R-net can be the finest carrier. Delay this choice to another stage. Its box image can be delivered to both the r-networks and 2 separate decrypted pictures will be acquired. It is explained that, despite the same network architecture, two R-networks need not express parameters [1].

Step-4: Classification of Frames or residual frames: The suggested spatial residual method poses fresh difficulties for the classic system. In Step-3, two copies of decoded confidential texts from R-network or the Residual R-network is obtained. It goes without saying that only one of the secret messages is accurate. In reality, all possible messages can be exhaustively enumerated: The real-reference and false remaining, real residual or false reference, total 4 legitimate instances. Construct this as a four way ranking problem. Reference-or-Residual (R-o-R) Nets is designed for judging a decoded input message. Like the R networks, let R-o-R Network has a 5-layer mainframe, each combined with the BN levels and Leaky ReLU. The main distinction for R-nets would be that the network top is a linear, completely linked element accompanied by a certain coating of softmax. Ultimately, softmax gets back a 4-d deterministic parameter, categorizing the decrypted data given an input picture [1].

Step-5: Reconstruction of the residual image: This stage is not mandatory if the Step 4 assigns a signal as a reference. Moreover, it is not visually comprehensible per se for a residual frame. To obtain the hidden video frame, Decrypted residuals should be added to the correct reference frame. Since process video frames are in temporary order, the recent reference frame can be recorded for residual reconstruction. H-nets and R-nets are taught collectively in suggested scheme before the RoR network. For studying H-nets / R-nets, the overall loss function is comprised individually losses described in each network. Remember to return decoded reference or residuals from the H-net output container frame and R-net. After a typical image segmentation treatment, H-net defines a failure on H network as a summary of the pixel wise distinction between the container cover as well as a failure on R-net to compare decrypted references and residuals with the old prints. Use normal cross entropy reduction to implement to learn RoR network [1].

8. ADVANTAGES OF CNN

Due to a data processing phase known as convolution, conventional neural networks are so named. A thorough mathematical treatise is beyond the scope of this short article on what is convolution. It is adequate to emphasize that convolution is useful in extracting information characteristics, and more so in aggregating and filtering those characteristics, thereby reducing the amount of characteristics (i.e. pooling). This built-in feature decrease is what makes such networks capable of analyzing information rich in features.

9. RESULTS AND ANALYSIS

Results and analysis include the various screenshot of the proposed work.

Figure 4 Snapshot of cover video

Figure 5 Snapshot of residual of cover video

Figure 6 Snapshot of secret video

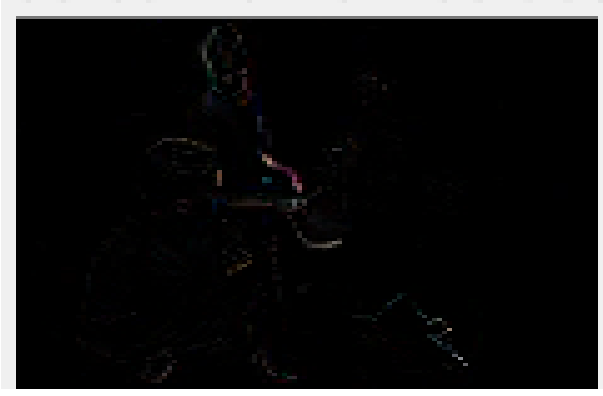

Figure 7 Snapshot of residual of secret video

Figure 8 Snapshot of stego video

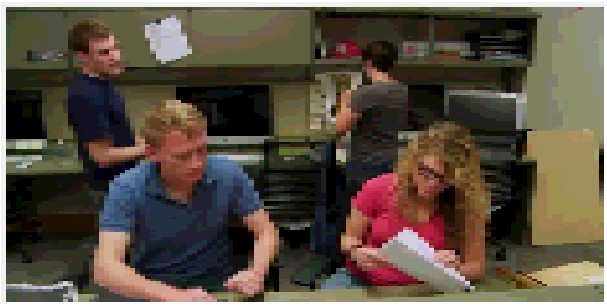

Figure 9 Snapshot of extracted video

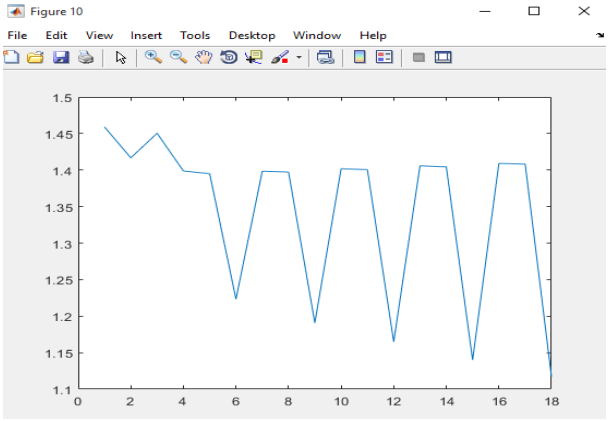

Figure 10 Snapshot for graph of MSE

Table 1: MSE as well as PSNR value of various cover videos and secret videos

Cover video	Secret video	MSE	PSNR
Cover video 1	Secret video 1	0.4167	51.93
Cover video 1	Secret video 2	0.3333	52.90
Cover video 1	Secret video 3	0.4167	51.93
Cover video 1	Secret video 4	0.3333	52.90

Cover video 2	Secret video 1	0.3333	52.90
Cover video 2	Secret video 2	0.3333	52.90
Cover video 2	Secret video 3	0.3333	52.90
Cover video 2	Secret video 4	0.3333	52.90
Cover video 3	Secret video 1	0.3333	52.90
Cover video 3	Secret video 2	0.3889	52.23
Cover video 3	Secret video 3	0.3333	52.90
Cover video 3	Secret video 4	0.5278	50.90
Cover video 4	Secret video 1	0.3333	52.90
Cover video 4	Secret video 2	0.3333	52.90
Cover video 4	Secret video 3	0.6389	50.07

Table 2: Snapshot for different cover videos and secret videos

Cover video	**Residual of cover video**	**Secret video**	**Residual of secret video**	**Stego video**	**Extracted video**

The PSNR (peak signal to noise ratio) value demonstrates the stego video quality. Higher PSNR shows the video's superior performance. MSE (mean square error) the MSE number shows the failure rate; the reduced MSE price shows the reduced error.

Table1 demonstrates the comparison of MSE as well as PSNR values for different cover video and secret video.

Table 3: Table demonstrates the comparison for PSNR value of different methods

Cover video	Secret video	PSNR(Kaur)	PSNR(Kothari)	PSNR(CNN)
Riga.mp4	Flag.mp4	43.05	40.13	51.14
Flag.mp4	Riga.mp4	35.91	35.61	52.90
Flowers.mp4	Flag.mp4	40.2	39.38	52.90

Table 3 compares the PSNR value for different methods. PSNR values of different methods are calculated [13]. Greater PSNR shows the video's superior performance. Kothari method has lowest PSNR value, kaur method also has less PSNR value compared to CNN method [13]. Proposed method has highest PSNR value, so it can be found that the suggested technique is more effective opposed to other techniques.

10. CONCLUSION AND FUTURE WORK

Proposed method uses convolution neural network and temporal residual method to conceal the confidential video into the cover video. Cover video of format mp4 has been used in the proposed method. Secret video should also be in the format of mp4. Inter frame difference helps in forming residual video and hiding the residual video is easier than hiding original video. To make full use of the broad estate of inter frame variations, the proposed method uses the temporal residual modeling method; this method deals individually with the reference frame and residual frame. . Usually, video consists of many frames and huge data and takes lot of time for embedding so to overcome this problem, convolutional video steganography using temporal residual method is proposed. Table 1 show the MSE and PSNR values. Table 3 shows the PSNR values for methods, values indicate that the PSNR compared to other methods, the method proposed is better. So it can be found that the suggested technique is effective.

Future research will include observing more sophisticated profound designs, such as C3D (3D ConvNets), that can be useful for stronger processing error.

11. REFERENCES

[1] Xinyu Weng1,2, Yongzhi Li1,2, Lu Chi1, Yadong Mu1,2Convolutional Video Steganography with Temporal Residual Modeling 1Institute of Computer Science & Technology 2Big Data Scientific Research Center Peking University, China.

[2] International Conference on Computer, Communications and Electronics (Comptelix) Manipal University Jaipur, Malaviya National Institute of Technology Jaipur & IRISWORLD, July 01-02, 2017.

[3] Multi-layers Video Steganography: A Novel Technique for Image Hiding "Information and Computer Science Dept., King Fahd University of Petroleum & Minerals, Saudi Arabia; 2,3Computer Engineering Dept., King Fahd University of Petroleum & Minerals, Saudi Arabia".

[4] International Journal of Computer Applications (0975 – 8887) Volume 95– No.20, June 2014 "Secure Data Hiding Technique Using Video Steganography and Watermarking".

[5] International Journal of Computer Applications (0975 – 8887) Volume 77 – No.17, September 2013 "A Novel Data Embedding Technique for Hiding Text in Video File using Steganography".

[6] International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 2851-2860 "Secure Data Transmission Based On Combined Effect Of Cryptography And Steganography Using Visible Light Spectrum".

[7] Khan, Z., et al., Threshold based Steganography: A Novel Technique for Improved Payload and SNR. International Arab Journal of Information Technology [Online]. 2016, 13(4).

[8] Sudeepa, KB., et al., A New Approach for Video Steganography Based on Randomization and Parallelization., Procedia Computer Science, 2016. 78: p. 483-490.

[9] Hasso, Abdul-Rhman S., Steganography in Video Files. International Journal of Computer Science Issues (IJCSI), 2016, 13(1): p. 32-35.

[10] Mstafa, R. J., et al., A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes. Multimedia Tools and Applications. 2015. p. 1-23.

[11] Balaji, R. and G. Naveen. Secure data transmission using video Steganography. Electro/Information Technology (EIT), on IEEE International Conference, 2011. p. 1-5.

[12] Khan, M., et al., Dual-level security based cyclic18 steganographic method and its application for secure transmission of keyframes during wireless capsule endoscopy ,Journal of medical systems, 2016. 40(5): p. 116.G.C.KesslerandC.Hosmer. An overview of steganography. Advances in Computers, 83:51–107, 2011.

[13] Comparison of Video Steganography Methods for Watermark Embedding David Griberman1, Pavel Rusakov Department of Applied Computer Science, Riga Technical University, Latvia 2016/19.