A Novel Deep Learning Architecture for Image Hiding

S. MUNI RATHNAM[1], G. SIVA KOTESWARA RAO[2]

[1,2]Dept. of Ece., Vemu Institute of Technology, Tirupati, INDIA

Abstract: Watermarking is a today's digital hiding technique within certain electronic content: for example, message, image, video, or audio recordings. Recent times, it was created as a modern copyright security tool. The pattern in zero watermarking technique isn't really inserted directly in the cover image, but has a logical relation with that cover image. In this article, we propose a powerful convolution neural Networks (CNN) and deep learning algorithm-based-watermarking technique in which the CNN produces robust inherent selected features and is merged with the XOR activity of host's watermark sequence. The outcomes of our proposed method present the courage of the watermark counter to many typical image processing techniques.

Keywords: Convolutional Neural Network (CNN), Deep Learning, Robust Features, Watermarking.

Received: June16, 2020. Revised: December 5, 2020. Accepted: December 16, 2020. Published: December 31 , 2020.

1. Introduction

Image processing and the world wide web have made life simpler for digital images to be duplicated, changed, replicated and reproduced at minimal price and with virtually expected shipping without even any quality reduction. Communication system has been so rapidly evolving and growing that it challenges data security and privacy [1]. Data verification, copyright security and duplication security therefore they play a vital role in overcoming the requirements of identifying new threats to the preservation of digital data.

Digital watermarking is defined as the integrating of data or information through digital technology. Digital image watermarking is broken down as spatial and frequency domain watermarking based on encoding location [2]. Rather than directly altering image pixels, spatial domain techniques change image pixels of images and frequency domain techniques include manipulating transform coefficients.

A quick right-click and save helps anyone to download the asset/content while uploading or publishing the digital content to public servers or portals. Their absolute choice is how, when and when they use it afterwards. Companies that spend time and money for generating original content without adequate security risk their assets being exploited by others, potentially affecting market opportunities and revenue incomes [3]. Watermarks can discourage or absolutely prevent this theft of content.

In several applications, Digital Watermarks are potentially helpful. Some of them are: Broadcast Monitoring, Advertising agencies want to know that all of the air time they buy from broadcasters is earned by them. A mistake that is vulnerable and expensive is a non-technical approach in which human interpretation is used to watch the broadcast and verify the originality by seeing or listening. Therefore, an auto-identification device should be in place that can store the broadcast identification codes. There are many methods, such as cryptography, that store the code number in the data frame, but it is unlikely that the data will survive any kind of changes, including format changes. Certainly, watermarking is a useful data control technique.

Ownership allegations, to justify his ownership, a legal owner may reclaim the watermark from digital content. With electronic copyright notices, there are restrictions, since they are easily removable [4]. It is not possible to copy the copyright notice written on a piece of paper along with the digital content. Copy protection and identity verification, these are used to avoid the creation of fraudulent copies of material by individuals. The transaction monitoring of the content is very similar to this issue. An owner may insert a watermark into digital content that identifies the copy buyer.

2. Related Works

In the article [1], author stated that Digital image authentication, as it is simple to corrupt with any image, is an extremely important consideration for the tech transformation. Several acceptable watermarking techniques have been developed to ease the problem, depending on the desired applications. However, achieving a watermarking method that is both stable and safe is difficult. This article describes descriptions of common device structures for watermarking and lists several basic specifications that are used for many different applications in the design of watermarking techniques. In order to find the state-of-the-art approaches and their drawbacks, the latest developments in digital image watermarking techniques are also examined.

In article [2], author demonstrates Digital technology is becoming more prevalent with the growing advancement of information technology, so the need for copyright protection has risen. To prevent digital data from being corrupted, digital watermarking will be used. Author presents an algorithm using LWT (Lifting Wavelet Transform) to insert fractal images into the wavelet domain. From Quadtree decomposition, fractal images are captured and processed into binary images. It uses this binary image as a watermark. In order to secure information from threats, the cover image is transformed into RGB color space and binary watermarks are embedded into mid-frequency cover image bands. The watermark is embedded in the RGB image's blue portion. The PSNR values which are obtained in this research are not up-to satisfied levels.

The author in [4], given a brief description about watermarking applications and some of the techniques. With the growth of the internet, the regular production of digital media such as music, photographs and videos has become available for the public. In order to ensure and promote data verification, encryption and copyright enforcement to digital media, digital watermarking technology is being introduced. In order to stop unauthorized data duplication, it is considered the most critical technology in today's world. For audio, video, text or pictures, digital watermarking may be added.

3. Implementation

Like other popular watermarking schemes, the proposed watermarking method is consisted of the master share generation stage and the image verification stage. The process of master share generation is shown by Fig. 1, although the stage of image verification is shown in Fig. 2. The formerly trained CNN is working as an integral part of our structure in both processes. First, throughout this portion, we will discuss about the CNN architecture used in the proposed model and some hyper-parameters used in the CNN training process. Next, all phases consisting of the proposed model are listed.

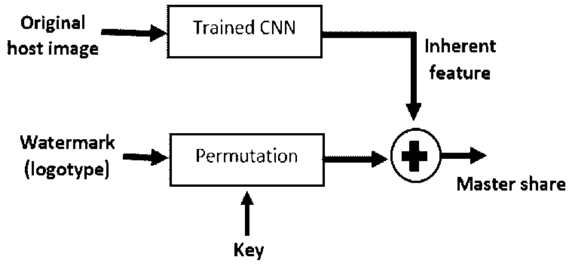


Fig 1: Master Share Generation

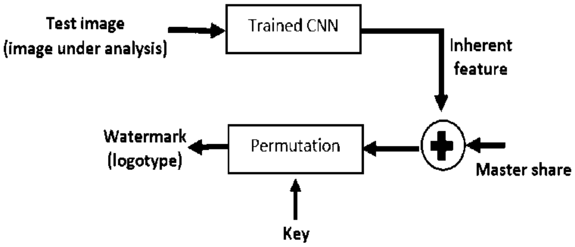


Fig 2: Stage of Image Verification

The CNN architecture of our model collectively has 27 number of CNN Layers, in which it has 12 convolutional layers, 3 max-pool layers, 4 batch normalization layers, 3 ReLU layers, 2 fully connected layers and a single image input layer, SoftMax layer and classification layer. The CNN architecture used in our model is shown in Fig. 3, which contains of 27 number of CNN layers. The inherent image characteristics/features are derived from these convolutional layers.

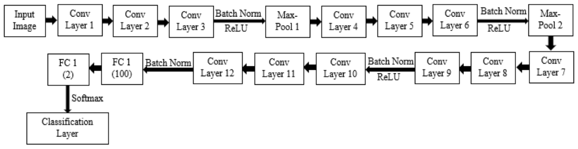


Fig 3: CNN Architecture

The CNN setup/configuration and description about hyper-parameters are shown in Table I and Table 2 respectively.

Layers	Configuration
Image Input Layer	300X300X3
Conv Layer- 1,2,3	3X3, 8, Stride-1
Batch Normalization	-
ReLU	-
Max-pool Layer	2X2, Stride-2
Conv Layer- 4,5,6	3X12, Stride-1
Batch Normalization	-
ReLU	-
Max-pool Layer	2X2, Stride-2
Conv Layer- 7,8,9	3X14, Stride-1
Batch Normalization	-
ReLU	-
Max-pool Layer	2X2, Stride-2
Conv Layer- 10,11,12	3X16, Stride-1

Batch Normalization	-
Fully Connected-1	100
Fully Connected-2	2
SoftMax Layer	-
Classification Layer	-

Table I: Layers Configuration

Options	Properties/value
Optimizer	Stochastic gradient descent algorithm
Learn Rate Schedule	Piecewise
Learn Rate Drop Factor	0.2
Learn Rate Drop Period	5
Maximum Epochs	30
Mini Batch Size	30

Table 2: CNN Hyper-Parameters

The best hyper-parameters for the proposed CNN structure, which is given by Table II, were chosen after many experiments. The training was carried out using a Graphics Processing Unit (GPU), NVIDIA and a MATLAB Deep Neural Networks Toolbox in version of R2018a.

Let's discuss about the Master Share Generation and Image Verification Stage. In Master Share Generation, after the CNN is trained, the image 's inherent features are collected from the CNN fully connected Layer 1 containing 100 valid data output. This output data is transformed by using threshold value 0 into binary data, with a positive value equal to 1, otherwise 0. The pattern of the watermark is a binary matrix, which is shown in Fig. 4. Using the user's hidden key to the random binary image, the watermark sequence is down - sampled.

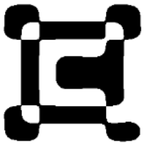


Fig 4: Binary Pattern

The master share is created in between binary series of the inherent features of image and concatenated binary watermark pattern by XOR operation.

$$Master\ Share\ =\ AF\ \otimes \hat{Q}$$

From fully connected layer, AF of binary series/sequence is extracted, $\hat{Q}$ is the sequence of permutated binary watermark. $\otimes$ is the representation of XOR operation. In a safe way, the master share is managed by the user.

In the stage of image verification, the copyright of some images was checked at this point using the image features retrieved by the trained CNN and the collected master share. It is possible to easily obtain the inherent features of the image under research, insert the image into the trained CNN and retrieve 100 valid output data from the fully connected layer 1. Then, in the Master share generation stage, these data are compressed using the same process. We extracted the permutated watermark sequence using the binary function and master share, which can be expressed as:

$$\tilde{Q} = \widetilde{AF} \otimes Master\ Share$$

Obtained features of fully connected layer 1 is $\widetilde{AF}$. $\tilde{Q}$ is the collected permutated watermark sequence. The watermark pattern is retrieved from $\tilde{Q}$ using the same secret/user's key.

4. Results and Discussions

Several blurred images that are not used in the training process are created to assess the robustness of the proposed watermarking scheme. I JPEG compression with distinct consistency metric, mean filter, Gaussian smoothing filter, noise in the image and distortions are included in the watermark robustness assessment.

We used PSNR as performance metrics. The peak signal-to - noise ratio in decibels between two images is calculated by the PSNR block. This ratio is used between the main and a compressed image as a performance indicator. The greater the PSNR, the higher the accuracy of the image that is compressed or recovered. In MATLAB, we use 'psnr' command to get the psnr value of given input. PSNR can be mathematically represented as:

$$20\log_{10}(Max(Max(f)/\sqrt{MSE}))$$

The matrix data of our original image is represented by f. The maximal matrix value in our original image "proven to be fine" is MAX(f).

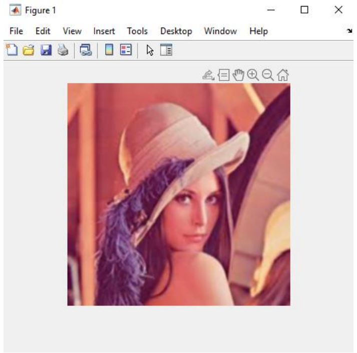


Fig 5: Input/Cover Image

Let’s discuss about the evaluation of our proposed model results, the above fig 5 is the cover/input image which is obtained by selecting the option of gaussian filter with factor 1. The cover image we have used here is a Matlab inbuild image of Lena. We can call this cover image is also as Attacked Image.

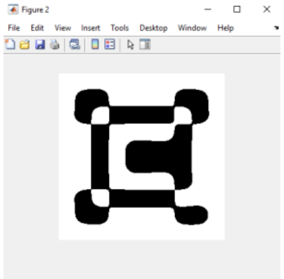
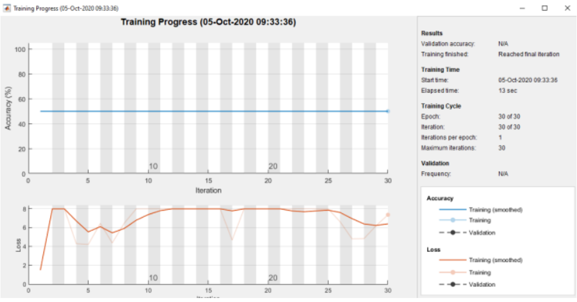


Fig 6: Pattern to hide

Image in fig 6 is the binary pattern image which we want to hide in the source image by CNN algorithm



Results
Validation accuracy: N/A
Training finished: Reached final iteration

Training Time
Start time: 05-Oct-2020 09:33:36
Elapsed time: 13 sec

Training Cycle
Epoch: 30 of 30
Iteration: 30 of 30
Iterations per epoch: 1
Maximum iterations: 30

Validation
Frequency: N/A

Fig 7: Training Progress

In the process of training the deep learning networks, it is also helpful to track the training progress. By plotting different metrics during training, you will understand how the training is performing. For instance, you will decide when and how rapidly the accuracy of the network is increasing.

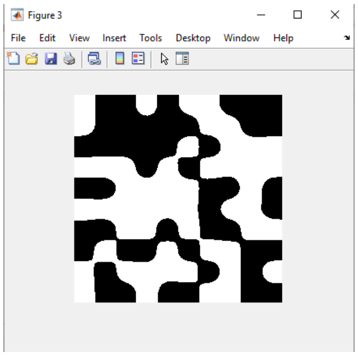


Fig 8: Embedded/Encrypted Image

The above fig 8 is the encrypted image. It is encrypted by the user’s secret key which is used as the same key for both encryption and decryption. Fig 9 is the final decrypted image which will obtained by the secret key. After the subsequent attacks, Table 3 displays some attacked images and recovered watermark patterns.

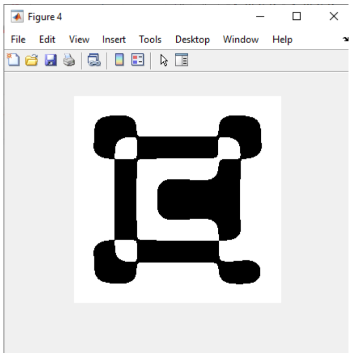


Fig 9: Decrypted/Recovered Image

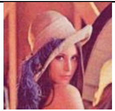
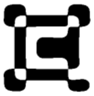
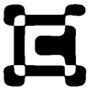
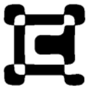
Attack	Attacked Image	Recovered Image
Mean Filter (3X3)		
Gaussian Noise (0.001)		
Impulsive Noise (0.03)		

Table 3: Some attacked images and recovered watermark patterns.

S. No	Existing Method	Proposed Method
1	41.72	50.57
2	41.72	51.31
3	41.72	51.05

Table 4: Comparison of PSNR with Existing and Proposed Method

The above table 4 gives the comparison of PSNR between existing and proposed methods. It states that the existing method performs nearly 40% in PSNR value where proposed model performs nearly 50%. From these tables, we assume that greater robustness can be offered by our proposed method.

5. Conclusion

In this article, we suggested a watermarking method based on CNN in which the inherent host image characteristics/features are derived from the first fully connected layer of the qualified CNN. In order to produce master share, which is stored safely, the extracted features are binarized and logically connected with the owner's watermark pattern by XOR operation. Watermarking is effectively used in the delivery of healthcare services or electronic medical records, trying to take advantage of this watermarking technique's distortion-free embedding capacity. The key disadvantage of traditional watermarking is the high computational complexity of the testing process, in which robust inherent characteristics must be derived from each medical image input. This model gives the better performance when compared to traditional models.

References

[1] Mahbuba Begum and Mohammad Shorif Uddin, “Digital Image Watermarking Techniques: A Review”, MDPI, 17 February 2020.

[2] Ashish Kamble and Sushama S. Agrawal, “Wavelet Based Digital Image Watermarking Algorithm Using Fractal Images”, IEEE Conference Record # 45616; IEEE Xplore ISBN: 978-1-7281-0167-5.

[3] Mediavalet, “Why Watermarks are Important and How to Use Them”, © 2020 Mediavalet Inc.

https://www.mediavalet.com/blog/watermarks-are-important

[4] Aaqib Rashid, “Digital Watermarking Applications and Techniques: A Brief Review”, International Journal of Computer Applications Technology and Research Volume 5–Issue 3, 147-150, 2016, ISSN:2319–8656.

[5] V. Y. Wang, J. F. Doherty and R. E. Van Dyck, “A Wavelet-based Watermarking Algorithm for Ownership Verification on Digital Images,” IEEE Trans. on Image Processing, vol. 11, no. 2, pp. 77-88, August 2002.

[6] LI, Cheng-Hao & WANG, Shuenn-Shyang. (2000). “digital watermarking using fractal image coding,” IEICE Trans. E83-A.

[7] Soheila Kiani, Mohsen Ebrahimi Moghaddam, “A Multi-purpose Digital Image Watermarking Using Fractal Block Coding,” Journal of Systems and Software, Volume 84, Issue 9,2011, Pages 1550-1562,