CHAPTER 1

INTRODUCTION

1.1 Background

Watermarking has been in existence for several centuries owing to its simplicity and low cost. With the advent of technology, regular watermarking has proved to be inadequate for protecting owners from malicious threats in areas like audio, video and images.

Digital Watermarking can verify the credibility of the content or recognize the identity of the digital content's owner. This is done by embedding data into digital multimedia content.

To improve accuracy in embedding and extracting the watermark, we can make use of neural networks. The structure of artificial neural network is based on the nervous system of the human brain.

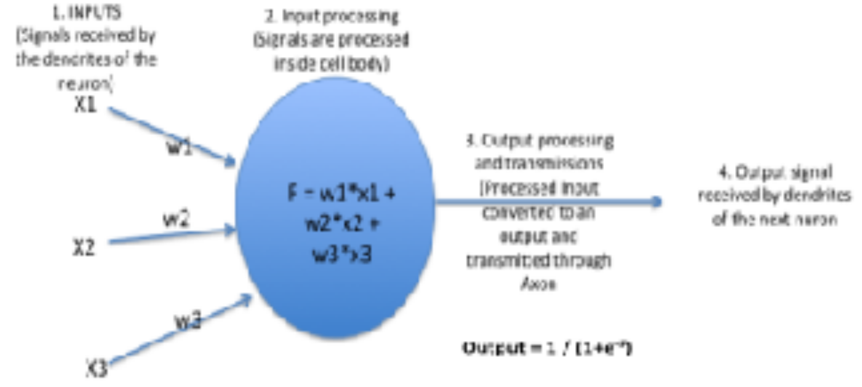
1. INPUTS (Signals received by the dendrites of the neuron)
2. Input processing (Signals are processed inside cell body)
X1
X2
X3
w1
w2
w3
F = w1*x1 + w2*x2 + w3*x3
3. Output processing and transmissions (Processed input converted to an output and transmitted through Axon
4. Output signal received by dendrites of the next neuron
Output = 1 / (1+e^-F)

Fig 1.1 Artificial Neural Network

With the higher usage of digital platforms, the security of entities like images, audios and videos has become a major challenge. Watermarking was chosen as it would be easier to keep these digital entities copyrighted because it is used to verify the identity and authenticity of that particular owner. Using Neural Networks over other means of prediction and analysis proved to be faster, more accurate and better to work with in terms of large image datasets.

1.2 Problem Statement

To extract watermarks from the images and check if they have been tampered with or not, using Neural Networks.

1.3 Specific Objective

The main objective of the project is to decipher if a particular image's watermark has been tampered with or not. Watermarking is solely for copyright and security purposes. In order for a receiver to know who the legitimate owner of a particular image is, or to decrypt the secret information embedded in the image, the image must not be tampered with or attacked. The images may be tampered with, using various types of attacks like adding noise, performing compression, rotating the watermark, and so on. This project aims at extracting watermarks successfully from an image, and analyze if the extracted watermark has been attacked. This extraction is done with Convoluted Neural Networks.

1.4 Limitations

The various types of attacks on watermarked images can be intentional or unintentional. The algorithm used in this paper is equipped to face several types of attacks.

The approach is a non-blind approach rather than that of a blind approach as the location of the watermark is given to the network in prior.

This approach is limited to only visible watermarks.

The training algorithm used for the CNN trains based on a single watermark at a time and does not train with various watermarks all together.

Embedding cost, including the resource area, delay and power, should be made to a minimum.

CHAPTER 2
LITERATURE SURVEY

The following section provides a review of the literature related to the watermarking of a digital image that makes use of the Artificial and Convoluted Neural Nets to extract the watermark from the images in the dataset. The models in this paper learn the position and type of watermark from these images and will then be able to predict the same watermark from the attacked or tampered images. The accuracy is calculated to see if the watermark can be correctly extracted even if the image has been tampered with. The following papers deal with Neural Network models and Image Watermarking along with pre-processing of the images using DWT and DCT.

The implementation of an artificial neural network for reversible watermarking was discussed in "Improving Prediction Based Digital Image Reversible Watermarking By Neural Networks" by Mahsa Afsharizadeh and Hossei Ebrahimpor-komleh at the Second International Congress On Technology, Communication And Knowledge (ICTCK 2015) November, 11-12, 2015" [1]

Digital image reversible watermarking hides data in a digital image in such a way that it is possible to recover the original image after the watermark extraction. The prediction is very important in the prediction based reversible watermarking. The more accurate prediction yields smaller prediction errors and therefore more efficient embedding procedure. In [1], the authors proposed a high capacity prediction based reversible watermarking through the use of artificial neural networks. The structure of artificial neural network is based on the human brain nervous system and has learning ability so it is a powerful tool to predict. They had used two artificial neural networks for predicting. Artificial neural networks provide higher accuracy than coordinate logical operations. Test results of [1] indicated that artificial neural networks predict more accurate compared to coordinate logic operation. Also their results showed the achieved to higher capacity and quality than many other methods.

They have improved the work done by Naskar and Chakraborty [9] in order to increase the hiding capacity and visual quality of the watermarked image. For this purpose, they had tried to make a more accurate prediction operation. Naskar and Chakraborty [9] used coordinate logic operation for predicting the pixel values. They had used artificial neural networks for more accurate prediction. They proved that the use of artificial neural networks led to more accurate prediction compared to coordinate logic operation. [1] dealt with an overview about coordinate logic and artificial neural networks and how they made predictions individually.

In the prediction based reversible watermarking method using coordinate logic operation, the relationship between neighboring pixels around a pixel was used to estimate the pixel value. The prediction error was equal to the difference between the predicted value and the actual value. Finally, the watermark bits were embedded in the values of the prediction errors. The following function was used for prediction:

$$P = f(N1, N2, N3, N4) = CAND(COR(N1, N2, N3), COR(N1, N2, N4), COR(N1, N3, N4), COR(N2, N3, N4))$$

After discussing the prediction methodology involved with the neighboring pixels, artificial neural nets were introduced. Artificial neural network is composed of a number of the processing elements called neurons. The neurons are connected to each other by their weights. While using the neural networks, the weights are adjusted to improve the performance of the neural networks. The structure of artificial neural network is based on the nervous system of the human brain. An artificial neural network is a network of neurons and weights which operate under the learning rules. These rules determine how an initial set of weights should be adjusted to achieve better performance. All neural networks are based on three cases:

1) The pattern of connections between neurons called network architecture.

2) The method of determining the weights of the connections, which is called the learning algorithm.

3) Activation functions.

The prediction and extraction were discussed extensively, after which their results was discussed. They had improved the work done by Naskar and Chakraborty [9]. Naskar and Chakraborty used coordinate logic operations to predict a pixel value. The watermark bits were embedded in the difference values between the predicted and the actual value of pixels. The prediction operation in the reversible watermarking methods was very important because an accurate prediction led to smaller prediction errors, which was why ANN was used. The watermark bits were embedded in the prediction errors using difference expansion technique so that smaller prediction errors led to better embedding operation and the embedded image quality was less reduced. In the work of Naskar and Chakraborty, the first set of predicted pixels was predicted by the base pixels then the second and third set of predicted pixels were predicted using the base and the first set of the predicted pixels. They used two neural networks. Both of these neural networks had four inputs and one output. The inputs for the neural network designed to predict the first set of the predicted pixels, were the pixel values of the four base pixels in four directions around the central pixel. The output was the value of the first set pixel located in the centre. The second neural network was used to predict the second and third set of predicted pixels and its inputs were the two base pixels and two predicted values of the first set predicted pixels located in four directions around the central pixel. The output was the value of the central pixel belongs to second or third set of pixels. The maximum embedding capacity for the proposed method and the Naskar and Chakraborty's method was compared by the different threshold values for the images Lena, Baboon, Airplane, Boat and Barbara. Their method achieved higher embedding capacities in all the different threshold values compared to Naskar and Chakraborty's method. They also found that PSNR improvement values of their method and the other methods for the embedding capacity values 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7.

It was also shown that their method achieved higher PSNR values in comparison with the other considered methods. The quality to capacity comparison between the method implemented by [1] and the method proposed by Naskar and Chakraborty[9] for the images Lena, Baboon, Airplane and Boat, were well represented. It was shown in all of these figures that their method [1] achieves higher PSNR values compared to the other methods. A prediction operation by artificial neural networks for reversible watermarking, was used. A neural network was made more accurate predictions compared to coordinate logic operation. They used back propagation method for learning the neural network. Sigmoid activation function was used for the network. Their results showed the effectiveness of using neural networks for predicting the pixel values in [1] when compared to the coordinate logic operation and some of the other methods.

Another approach to using back propagation neural nets for image watermarking is discussed in “An Approach of an Image Watermarking Scheme using Neural Network” By Asmaa Qasim Shareef, Ph.D Roaa Essam Fadel- University of Baghdad University of Baghdad at International Journal of Computer Applications (0975 – 8887) Volume 92 – No.1, April 2014. [2]

Image is considered as a communication channel to transmit messages in the watermarking schemes. An approach of an image watermarking scheme using neural network was implemented in this paper [2]. The main requirements of the watermark were the robustness and the imperceptibility. Watermark is a method where the invisible mark is placed on top of the information as a protection and identification of the ownership of the image. Information that requires watermark could be such as banknote, ID card and any other valuable documents. This is performed to protect digital images against illegal reproduction and illegal modifications. The most widely used technique for watermarking images is to add a pattern on top of an existing image. Usually this pattern could be a logo or something similar, which distorts the underlying image.

In [2], watermark came from the weights of an identify image that were loaded from a learned feed-forward neural network and the neural network was learned by using the back-propagation learning algorithm. The noised image was damaged by the salt and pepper noise. In order to identify the cover of extracted watermark, feed-forward neural network was used in the watermarking identification. The results of this scheme showed the robustness of proposed algorithms that had preferable performance for both identification and watermarking of a noised image.

They performed Gaussian Filter for the image noised with salt and pepper. The process of learning the neural network about the noised image had taken a few steps: By applying the Gaussian law to remove the noise: a) They applied window slide operation on the image matrix to divide it in to N*N window. b) They had to find the value of the fuzzy function for all the neighbor's pixels of the centre pixel of the window. c) Next, they found the new value of the centre pixel and all its neighbors' pixels inside the window. d) Later, they rearranged all the new values of the pixel in the window and then took the medium value and replaced it with the centre value of the window.

The output image was produced to the back- propagation learning algorithm in order to learn the multi-layered feed-forward neural network after applying the window slide to split the image into wind of 3*3 and enter each row of this window to learn it and save output in a new array and the weights in file. [2] performed Multi-layered Feed-Forward Neural Network. The number of inputs in the input, hidden and the output layer was initialized to 3. Both the weights connecting the input layer to the hidden layer, W1 and the weights connecting the hidden layer to the output layer W2, which were arranged as a matrix were initialized randomly. X_j was chosen as an input and Y_j as the output.

Back-propagation Learning Algorithm was performed such that: activation function was propagated from input to hidden layer and from hidden to the output layer. The error was calculated based on the actual output and the target output. The error in the hidden layer was calculated and the weights are adjusted accordingly between different layers. This process was repeated as many times to attain the best weights possible.

Their procedure to perform the watermark was to: Save the weights that were resulted from the algorithm into a file; Enter the noisy image into the neural network algorithm with the weights of the resulted filtered image. Then the resulted image from the neural network was displayed without noise. They inputed the cover image and entered it to the neural network algorithm as an input and saved the weights into another file. They later entered the watermark image to the neural network with weights of the covered image and then it produced the output hiding image. To extract the cover image from the watermark image [2] had to enter the produced hiding image to the neural network with the weights of the covered image.

[2] described the use of back-propagation learning algorithm that was applied to the feed-forward multi-layered neural network, with an additional advantage of hiding the trained network weights within the original cover image. They had implemented an approach to embed a nearly invisible watermark into an image and to extract the cover image as well, by only giving the cover image as input with no external weights being added. On the other hand, the watermark was unique to individual image, and would have been destroyed completely in case of any alteration, which was a property against hacking.

The robustness of the watermarking with CNN, in case of attacks, is outlined in "Robust and high capacity watermarking for image based on DWT-SVD and CNN," by W. Zheng *et al.*, at the *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Wuhan, 2018, pp. 1233-1237. [3]

With the rapid development of computer Internet, digital watermarking technology is considered as an important branch of information hiding technology research, quickly become a hot field of multimedia information security. As digital images are vulnerable to some common attacks during transmission, it is very necessary to design a watermarking algorithm which can resist all kinds of common attacks. Nowadays, most of the watermarking algorithms published rely much on the locations of the pixels for watermark embedding, which results in less robustness.

In [3], a robust watermarking algorithm based on convolution neural network (CNN) was proposed. They introduced discrete wavelet transform (DWT) technology and singular value decomposition (SVD) technology, to achieve the embedding process of watermark. The network was established in the spatial domain based on the pixels' relationships of watermark, host image and watermarked image. After that, the pixels of the watermarked image were slightly modified with the network. In addition, some attacks were taken to the watermarked image. Simulation shows that their algorithm had good performance and that the watermark extracted could be clearly identified.

Digital watermarking technology was to embed a number of logo information (watermark) directly into the multimedia content. After embedding, the use value of original contents remained unaffected and logo information was not perceived by the perception system, but only through a dedicated detector or reading device, where watermark can be extracted. Unlike encryption, digital watermarking technology can’t prevent the occurrences of piracy, but can determine whether the object is protected, monitor the spread of protected data, authenticate and illegally copy, resolve copyright disputes problems and provide evidence for the court.

Convolution neural network (CNN) is a kind of deep neural network, which has become one of the hotspots in many fields of science. In [3], they try to combine digital watermarking technology with convolution neural network to improve the robustness of watermarking on the basis of existing watermarking algorithms.

SVD is one of the techniques for applications in image processing. Specifically, SVD has been used for image compression, image hiding and digital watermarking. SVD is defined:

$$A = U * S * V^T$$

The image was decomposed into three matrices: two orthogonal matrices U, V and a diagonal matrix S. U, V are called left and right singular vectors. Coefficients diagonal matrix S are called the Singular Values (SVs) of the matrix A. In digital watermarking, SVD has some advantages: Reducing the size of signal embedded in the image. And the SVs of the watermarked image are less influenced by attacks. It can be used as a robust feature in digital watermarking.

A proposition on a blind digital image watermarking technique by combining DWT with SVD to improve the robustness and the capacity was given. In detail, SVs of watermarks were replaced with the suitable SVs of HH sub-bands of the original images. Additionally, their method generated keys that ensure the security for the watermarks in the embedding and the extraction process.

One loop of learning process consisted of the following three stages: Watermark embedding, attack simulation, and weight update.

1. Firstly, using singular value decomposition, the watermark was decomposed into singular values and singular vectors; Secondly, the singular vectors were barbarized and merged to obtain singular sequences; Finally, the singular sequences were scrambled to obtain the watermark sequence.

2. For embedding algorithm, A DWT transform of the host image was carried out to obtain four strip LL, LH, HL, HH. The singular values of the watermark were then embedded into the HH strip. The LL strip was decomposed by multilevel wavelet transform, and the watermark sequence was embedded into the low frequency component and high frequency component of LL strip.
3. In CNN, the watermarking was divided into blocks sized 16x16. Based on the relation between the divided watermarking and host image, the convolution neural network was constructed. Using different key K, different watermarking image were generated as training data for network to adjust the weights. They did the same division operating to host image as watermarking, and made it as convolution neural networks input, and then adjusted the watermarking images pixel value according to networks output.

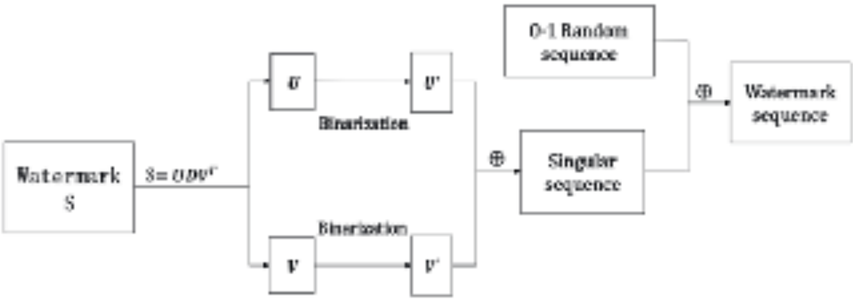

Fig 2.1 The Process of Generating Watermark Sequence

The attacks applied were Average Filtering, Motion-Blur, Sharpen and so on. Then, the Peak Signal to Noise Ratio (PSNR), Normalized Correlation (NC) and the correlation coefficient (CC) of the extracted watermark with respect to the original watermark were calculated.

The security aspects required in digital image watermarking regarding different domains is elaborated in "Robust Medical Image Watermarking Using Frequency Domain and Least Significant Bits Algorithms," by E. Elbasi and V. Kaya at the *2018 International Conference on Computing Sciences and Engineering (ICCSE)*, Kuwait City, 2018, pp. 1-5. [4]

Watermarking and steganography are getting importance recently because of copyright protection and authentication. While watermarking, stamps, logo, noise or images are embedded to multimedia elements such as image, video, audio, animation, software and text. Several works have been done in watermarking for different purposes. In this research work they had used watermarking techniques to embed patient information into the medical magnetic resonance (MR) images. There were two methods that had been used; Frequency Domain (Digital Wavelet Transform-DWT, Digital Cosine Transform-DCT and Digital Fourier Transform-DFT) and Spatial Domain (Least Significant Bits-LSB). Their experimental results showed that embedding in frequency domains resisted against one group of attacks, and embedding in spatial domain resisted against another group of attacks. Peak Signal Noise Ratio (PSNR) and Similarity Ratio (SR) values were the two metrics used for testing. The two values attained, had given very promising results for information hiding in medical MR images.

Fragile implied that the watermark should not resist tampering, or would resist only up to a certain, predetermined extent. In digital watermarking, visible watermark can be seen by eyes, other hand invisible watermarked image can't be visible. Invisible watermarked image is generally utilized for security. Watermarking algorithm should be secure and resist against common attacks such as filtering, compression, cropping, rotation etc. In this paper they embedded binary stamp on medical images (brain, breast and neck MR) using frequency domains and Least Significant Bits (LSB) algorithms.

They first elaborated on Discrete Wavelet Transform, which divided images into four sub bands. These sub bands are LL, LH, HL and HH. The magnitudes of DWT coefficients were larger in the lowest bands (LL) at each level of decomposition. Embedding the watermark in larger amount of sub bands (HL, LH, HH) gave productive robustness of watermark. However, the image quality can decrease which can be measured by PSNR. The discrete wavelet transformation was similar to the discrete Fourier transforms (or much more to the windowed Fourier transform) with a completely different merit function. The main difference was that Discrete Fourier Transformation decomposed the signal into sines and cosines, i.e. the functions localized in Fourier space; in contrary the digital wavelet transformation used functions that were localized in both the real and Fourier space. Then they discuss Discrete Cosine and Fourier

Transforms. DCT divided image into different frequency bands. The frequency components were ordered in a sequential order such as low frequency, mid frequency, and high frequency components. If most of the high frequency coefficients were zero, then they represented a smooth block. Discrete Cosine Transform attempted to de-correlate the multimedia element. After de-correlation of image data each DCT coefficient was encoded independently without losing compression efficiency. In DFT, first they extracted the components of the image to be watermarked, computing its full frame DFT, and then they took the magnitudes of the coefficients. They had a sequence of T samples F (t), indexed by t = 0..T-1, and the Discrete Fourier Transform was defined as F (k), where k=0..T-1. Finally, they discussed Least Significant Bit Algorithm. The most common method of watermark embedding is to embed the watermark into the least significant-bits (LSB) of the cover object. LSB is a simple method and it suffers from many drawbacks. This method resisted against some attacks such as cropping, any addition of undesirable noise or lossy compression. If hacker knows the algorithm, then he can easily find out embedded watermark such as image, stamp, text etc.

The experimental results in [4] show that if PSNR values were higher, the compressed or reconstructed image was of better quality. If PSNR value was infinity, it meant that two images were identically same and it was for this reason that the PSNR value result was infinity. In regard to the attack that was made on the images, highest PSNR values were in JPEG, Resize and Gaussian attacks. Rotate and histogram attacks had the worst PSNR values.

In conclusion to their work, watermarking in medical images had been analyzed by using DWT, DCT, DFT and LSB algorithms. The purpose of this study was for calculating PSNR and SR values on medical images via different algorithms and comparison of the results. In [4], logo images were inserted into the medical images (MR) by utilizing DCT, DWT, LSB and DFT methods and also, different attacks were connected on the subsequent images. Three different MR images were used. DFT method was found as the best PSNR value. DCT, DFT and LSB values were almost same. When comparing DCT, DFT and DWT method's SR values, DFT SR values were worst SR values. As a result of their algorithm, embedding in MR images using frequency domain resists against one group of attacks, and embedding using spatial domain resisted against another group of attacks.

Making use of Probabilistic Neural Networks for a similar watermarking purpose is covered in "Robust digital image watermarking using DWT, DCT and probabilistic neural network," by G. Kulkarni and S. Kuri, at the *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, Mysuru, 2017, pp. 1-5. [5]

[5] enhanced the performance of watermarking using Probabilistic-Neural-Network including three-level DWT and DCT transformations. In the embedding process, embedding of the watermark was done by training the PNN network. The new trained embedded values were used at the extraction side and inverse three-level DWT and DCT were applied. The inverse transformation gave the extracted watermark. Robustness test was carried out with various attacks on the different image and calculation of performance metrics.

The two main techniques of digital watermarking were:

1. Spatial domain watermarking.
2. Frequency domain watermarking.

Spatial domain watermarking techniques alter the pixel values Frequency-domain watermarking methods embed the watermark in the coefficients of transformed image. DCT and DWT are the most widely used transforms. Implementing watermarking technique in frequency domain has few advantages: higher capacity, better robustness against certain attacks, very near to human visual system (HVS), and better frequency localization of coefficients of the cover image where the watermark is to be embedded.

A three level decomposition of cover image was carried out using discrete wavelet transform (DWT) with Haar wavelet. DWT divided the image in to set of fixed frequency bands namely vertical, horizontal and diagonal frequency bands. Then, block based DCT was applied on the DWT coefficients. DCT reduced the image to have less compression so that the image got less compression effect. The DWT and DCT transformations were applied for both cover image and watermark. Then, the watermark coefficients were inserted into the original image. The new values were then applied to PNN for training. The trained PNN was then used to recover the watermark during watermark extraction process.

The Probabilistic Neural Network is a supervised learning network. It implements the Bayes-approach. PNN uses the radial function for the estimation of the sample likelihood values. It is a feed forward neural network and does not have the feedback connection. Therefore, training PNN was very fast and easy to learn compared to neural networks having feedback, such as BPNN.

The extraction process was an inverse process of that of embedding procedure. The trained PNN was used in the extraction process. This neural network remembered the relation between coefficients of watermarked image and corresponding pixel in the watermark image. A three-level decomposition was performed on the watermarked image using DWT. These coefficients were then divided into small blocks of size 4x4 pixels and applied DCT on each block. Then, they extracted the content of these coefficients and used them as input to the trained PNN. Inverse transforms were performed to obtain the watermark data.

The performance metrics used were:

1. Mean Square Error (MSE): It is the average of square of the difference between original image and watermarked image.
2. Peak Signal to Noise Ratio (PSNR): PSNR is calculated as the between original and watermarked image. Higher value of PSNR indicated that watermarked image is closely similar to original image.
3. Normalized Correlation Coefficient (NCC): This metric is used to analyze compatibility of original image and recovered image. The value ranged from 0 to 1.

The experiment was carried out for different attacks like JPEG compression, Gaussian noise, Salt and pepper noise and image rotation. [5] used a method of watermarking that produced better results and was robust to indicated attacks. Also, the computational cost was better because Haar transform and PNN was used.

A bit more insight into robustness required in digital image watermarking along with the use of a popular attack, is mentioned in "A robust watermarking scheme to JPEG compression for embedding a color watermark into digital images," by D. Muñoz-Ramirez, V. Ponomaryov, R. Reyes-Reyes, V. Kyrychenko, O. Pechenin and A. Totsky, at the *2018 IEEE 9th International*

Conference on Dependable Systems, Services and Technologies (DESSERT), Kiev, 2018, pp. 619-624. [6]
Technology advances and easy access to multimedia tools with digital content have increased the number of issues in copyright procedures. Digital watermarks are a set of techniques that are used to protect the copyright of digital content. In this paper, a robust watermarking framework to embed a color watermark was presented. To make a color image as watermark, a method based on Discrete Cosine Transform (DCT) and Quantization Index Modulation (QIM) had been designed. Also, the color watermark was encoded in such a way that the data to represent the colors were reduced. Additionally, the coded watermark was embedded into the mid-frequency coefficients of DCT to ensure the robustness and imperceptibility of the watermark. The efficiency of the proposed scheme against the most common attacks such as JPEG compression, impulsive and Gaussian noises, scaling, etc., had been tested in terms of Peak Signal-to-Noise Ratio (PSNR), Similarity Structural Index Measure (SSIM) and Normalized Correlation Coefficient (NCC) demonstrated good performance.

A different approach to using domains is achieved by using contours and statistical distributions, in a robust manner, is elaborated in "A Robust Image Watermarking Scheme Using Local Statistical Distribution in the Contourlet Domain," by H. Sadreazami and M. Amini, at the *IEEE Transactions on Circuits and Systems II: Express Briefs*. [7]
Data security is a main concern in everyday data transmissions in the internet. A possible solution to guarantee a secure and legitimate transaction is via hiding a piece of tractable information into the multimedia signal, i.e., watermarking. [7] proposed a new multiplicative image watermarking scheme in the contour-let domain by taking into account the local statistical properties and inter-scale dependencies of the contour-let coefficients of images. Although the contour-let coefficients were non-Gaussian within a sub-band, their local distribution fitted the Gaussian distribution very well. In addition, it was known that there existed cross-scale dependencies among these coefficients. Hence they decided to use of bivariate Gaussian distribution to model the distribution of the contour-let coefficients. An optimum blind watermark decoder was designed in the contour-let domain using the maximum likelihood

method. By means of carrying out a number of experiments, the performance of the proposed decoder was investigated with regard to the bit error rate and compared to other decoders. It was shown that their decoder built upon the bivariate Gaussian model was superior to other decoders in terms of rate of error. It was also shown that their decoder provides higher robustness in comparison to other decoders in presence of attacks such as filtering, compression, cropping, scaling and noise.

Finally, a project based on robust watermarking in regard to Discrete Wavelet Transform domain alone, was projected in "A Robust Watermarking Scheme Over Quadrant Medical Image in Discrete Wavelet Transform Domain," by O. Göker, N. Nazli, M. M. Erol, R. Choupani and E. Dogdu, at the *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, Thessaloniki, 2018, pp. 277-282. [8]
The diffusion of digital content is very fast in today's technology. The velocity of gathering data might cause unlawful distribution of content. The major problem in content authorization is the robustness of methodology. Since frangible methodologies result unauthorized content access quicker, more robust solutions are essential for copyright protection. Watermarking technology is considered as a robust solution for copyright protection and authentication. In watermarking, quality of the image is a challenge. Applying a watermark on a medical image might cause corruption in original image, which leads to misleading content. [8] used Quad-tree algorithm which found a region of non-interest to apply watermark on medical image in Discrete Wavelet Domain to provide authentication of the content without altering region of interest. Furthermore, the visual quality of watermark implemented medical images and sample values were also discussed with the experimental results as well.

CHAPTER 3
REQUIREMENT SPECIFICATION

The proposed work requires the following system requirements:

- System with 4GB RAM minimum

These are the required installations:

- Python installation with 2.6 version minimum
- Keras
- OpenCV
- Numpy
- PIL
- Pandas
- Matplotlib
- Sklearn
- random, os, requests, beautifulSoup, Scipy

CHAPTER 4
ARCHITECTURE

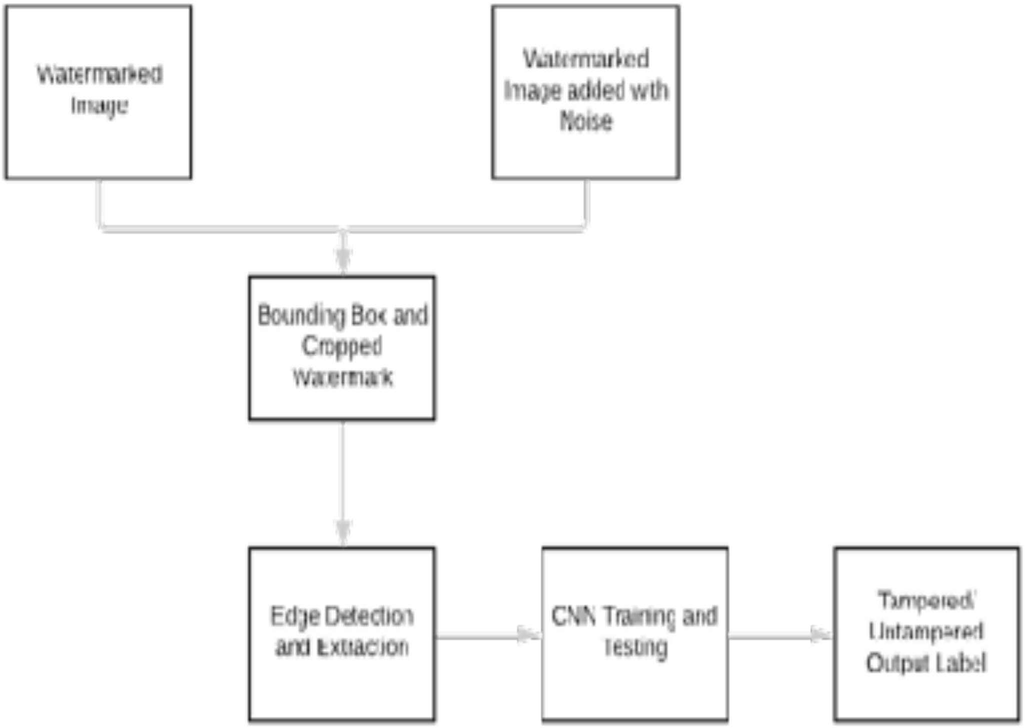
Watermarked Image
Watermarked Image added with Noise
Bounding Box and Cropped Watermark
Edge Detection and Extraction
CNN Training and Testing
Tampered/ Untampered Output Label

Fig 3.1 CNN Architecture

The proposed solution follows this working architecture:

- Initially the original watermarked image is taken and the same image added with various attacks.

- On these images, an automated algorithm is run where a bounding box is created over the watermarks despite whether the image is attacked or not. The watermarks are then cropped out and placed in respective datasets.
- We then perform edge detection on the images in these datasets by using the Canny Edge Detector, and these edges are used as input features into our Convoluted Neural Network.
- After the network is trained, using the keras library among many others mentioned in the Python library requirements, we provide some test images for the network to decipher. The final output will be a binary output label of 0 or 1 where 0 refers to Untampered and 1 refers to Tampered.
- Based on the results, the accuracy between the two different types of images can be compared and analyzed.

CHAPTER 5
RESULTS AND DISCUSSIONS

The results that have been achieved are given in chronological order:

1. Initially the proper pre-processing of the watermarked images was done. The images were segregated into datasets and attacked with noises like Salt and Pepper, Speckle, Blur and Gaussian and other forms of attack like image rotation by different degree measurements and image compression or complete removal of the watermark. These are some attacks that have been performed on these images, to use to feed the neural network in the training and testing phase.

2. Next, an implementation which will place a bounding box over the watermarks was written, after which the watermarks were automatically cropped and added to their respective new datasets.

3. Consequently, a CNN was created using the Keras library which took the images from the training dataset and performed Canny Edge detection on them. These edges were used as an input feature into the CNN. Based on the dimension of the edges, the watermarks were then normalised and used for training the network. After this, test images are given, and the output is predicted. The output labels are 0 or 1, representing ‘Tampered’ or ‘Untampered’ respectively.

4. A 85% accuracy was achieved as an end result, given the different types of attacks the network was trained with.

Fig 5.1 Watermarked Image

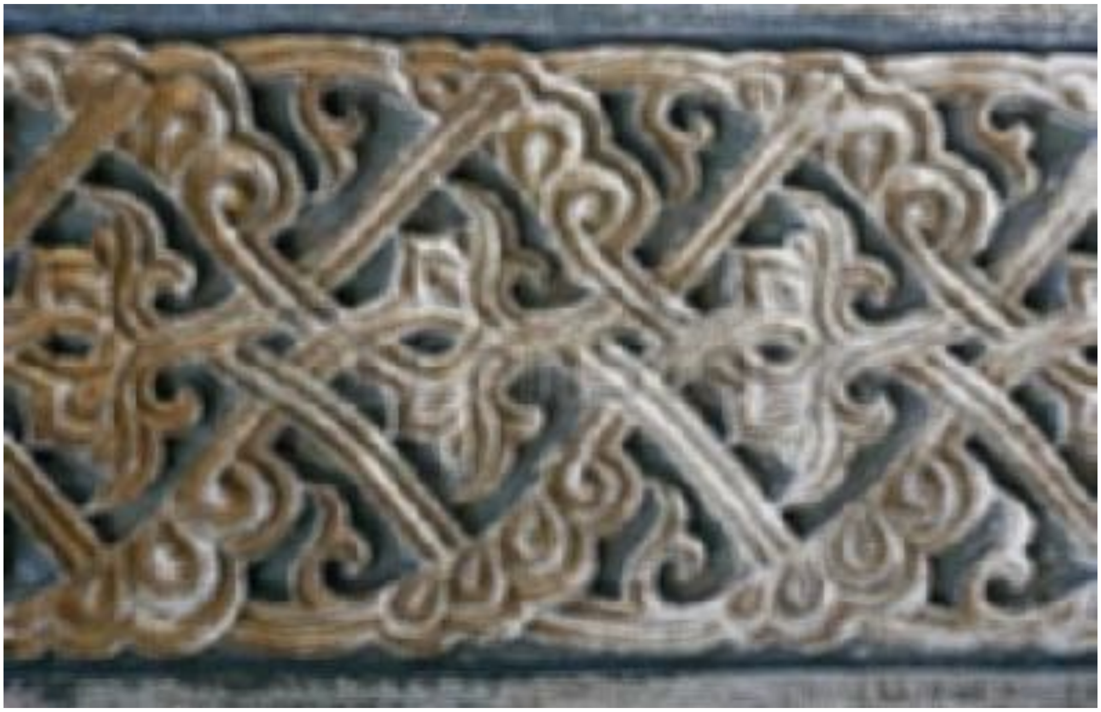

Fig 5.2 Blur Noise Attack

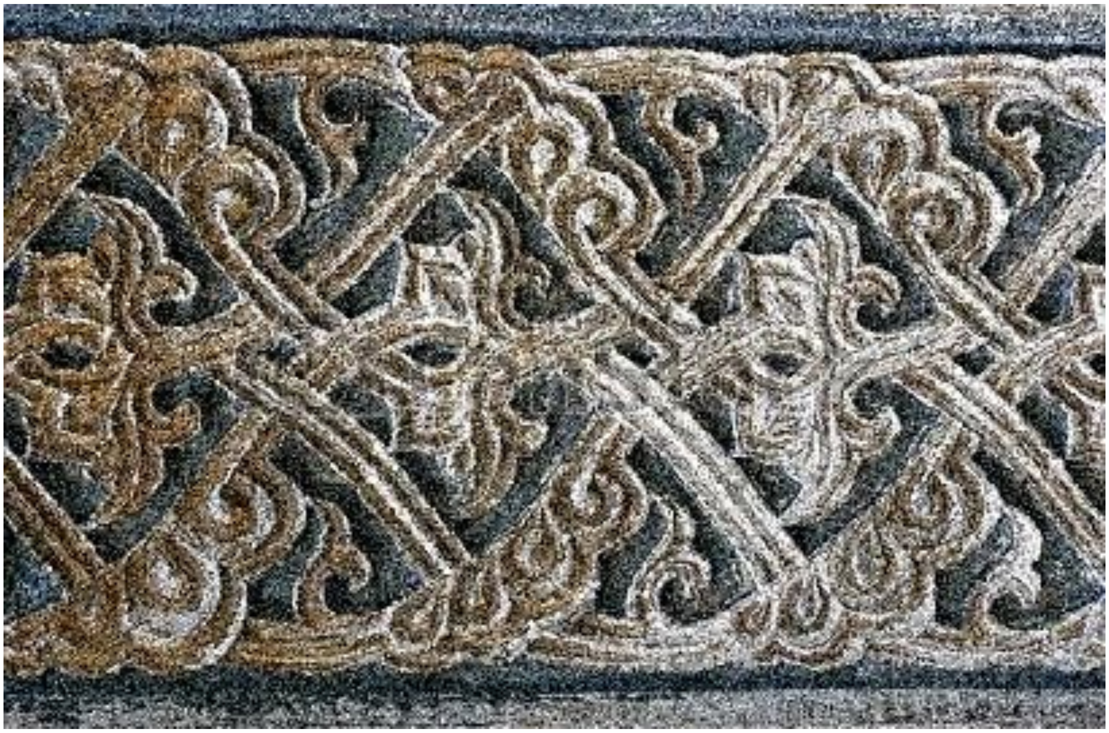

Fig 5.3Gaussian Noise Attack

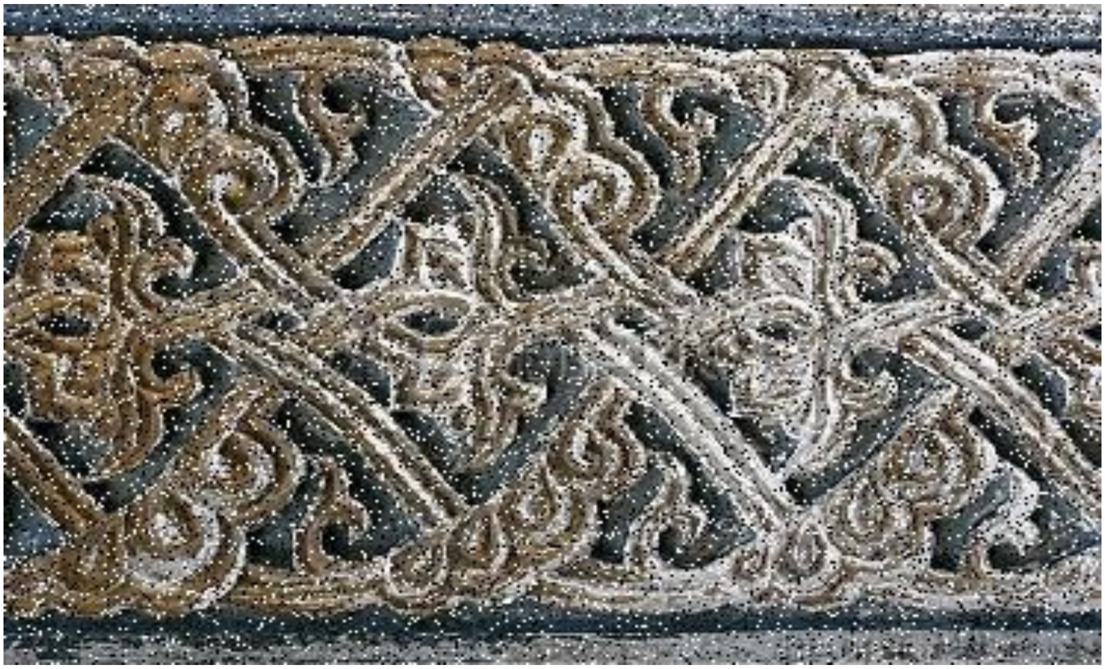

Fig 5.4Salt and Pepper Noise Attack

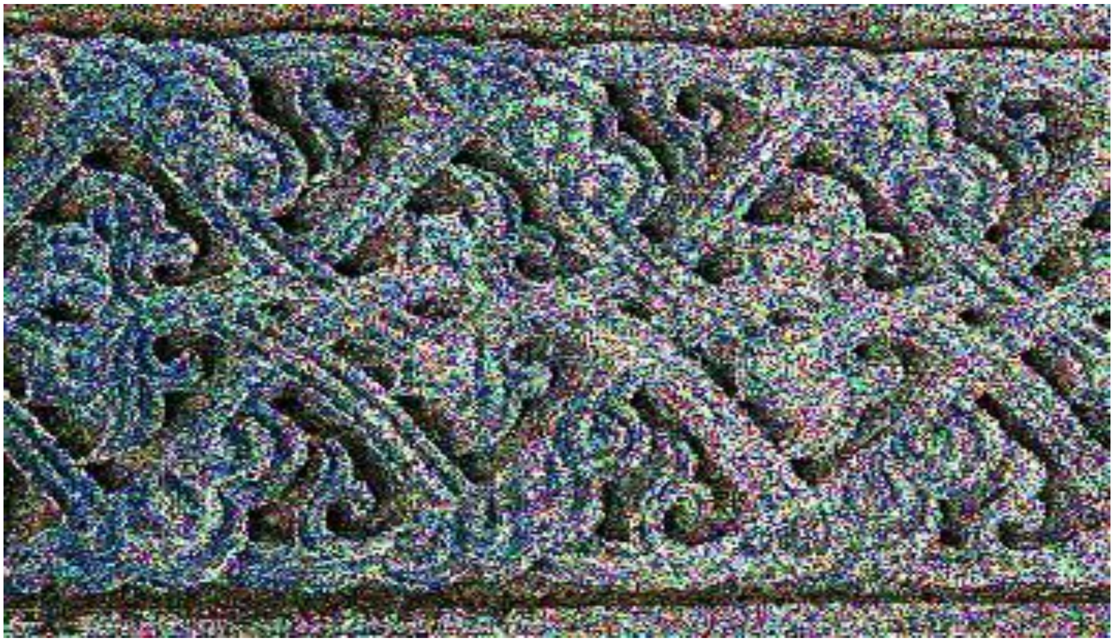

Fig 5.5Speckle Noise Attack

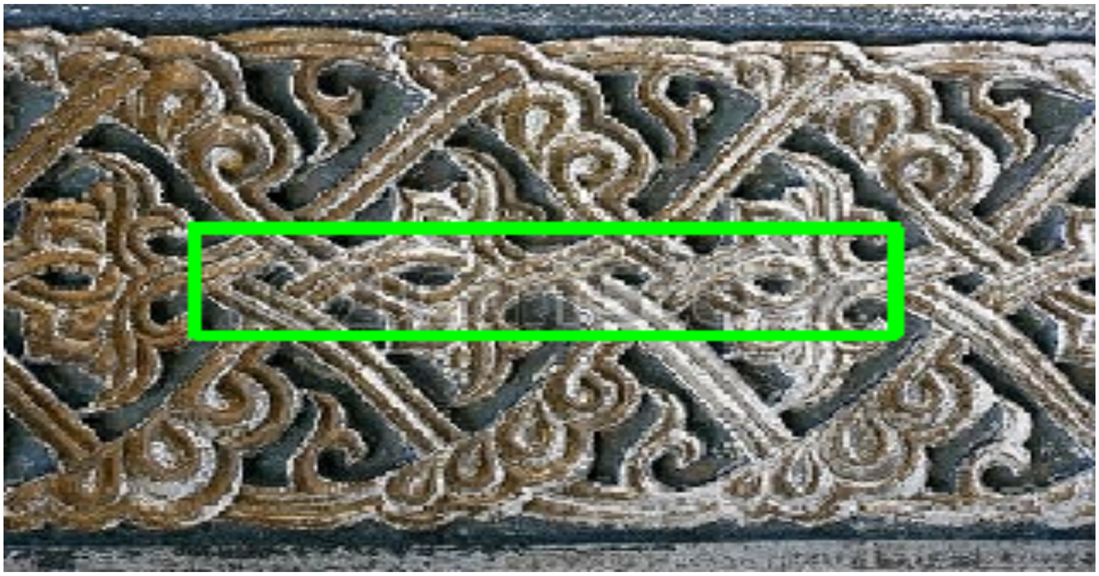

Fig 5.6 Original Watermark with bounding box

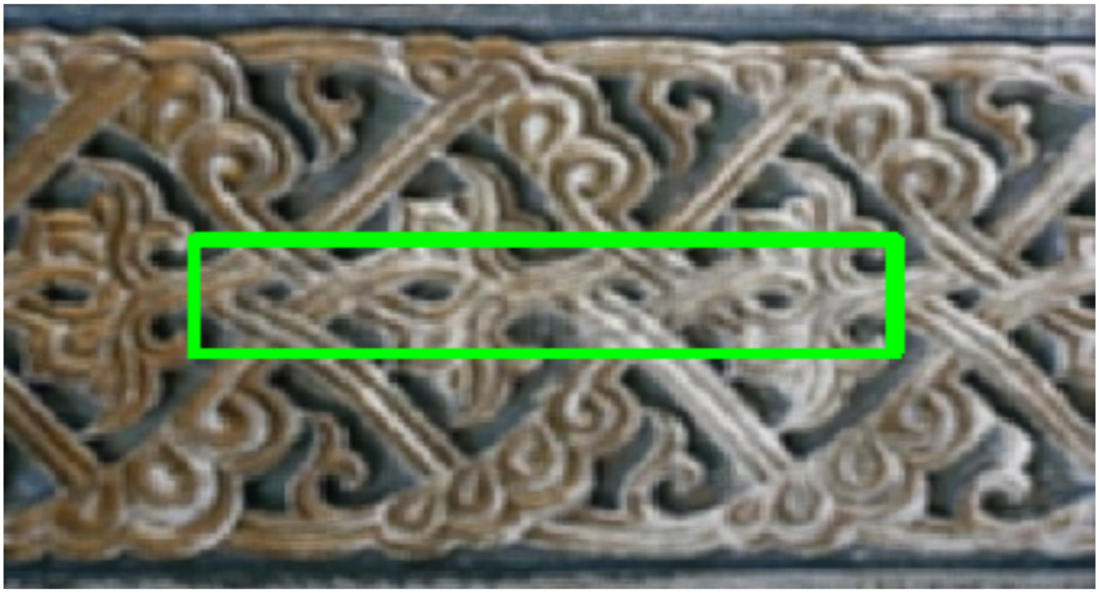

Fig 5.7 Attacked Watermark(Blur) with bounding box

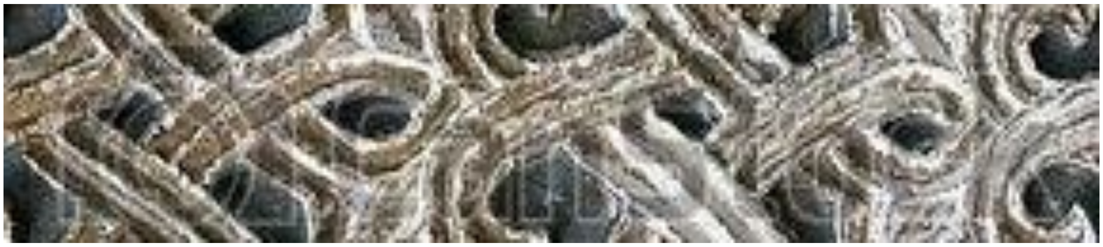

Fig 5.8 Extracted Original Watermark

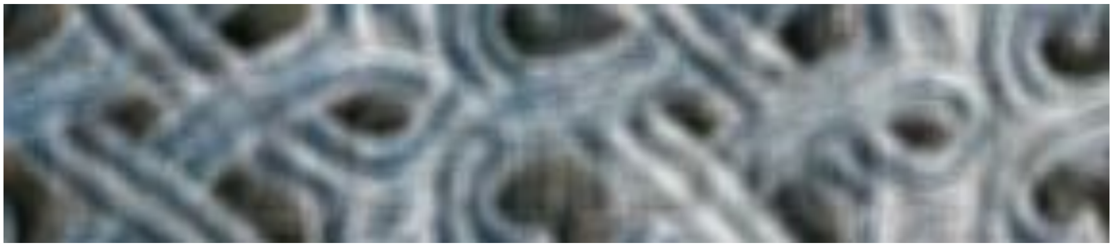

Fig 5.9 Extracted Attacked Watermark(Blur)

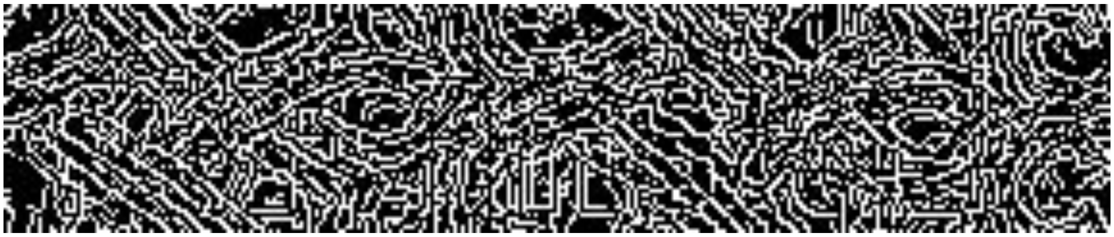

Fig 5.10 Extracted Original Edge Feature for CNN

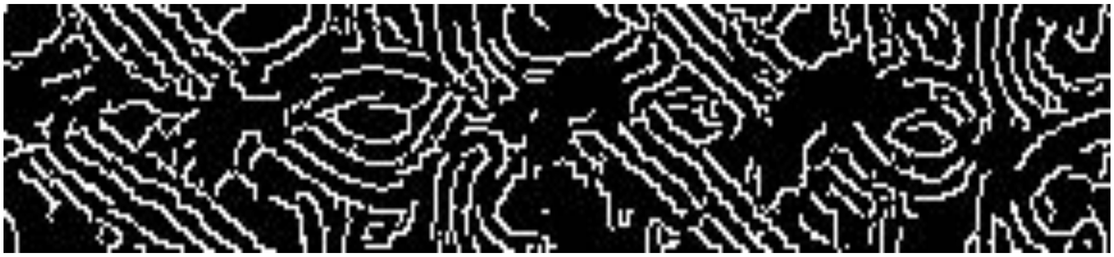

Fig 5.11 Extracted Edge Feature for CNN (Blur)

Fig 5.12 Test Images into CNN

CHAPTER 6
CONCLUSION

The application created in this paper works towards correctly classifying an extracted watermark as attacked or not attacked by making use of a Convoluted Neural Network.

This application can be further developed by training the network with any kind of watermark, (i.e) varying sizes of watermarks or different watermarks all together. Also the application can be developed to work with more difficult attacks as well. The attacks used to train the network in this application, are only a few and much more can be implemented to make the application more scalable.

Implementing an application of this scale, would require a much larger database of images containing different attacks as well. A larger database would in fact help improve the accuracy of the labels and strengthen the network's abilities.

CHAPTER 7
REFERENCES

[1] Mahsa Afsharizadeh, Hossei Ebrahimpor-komleh, "Improving Prediction Based Digital Image Reversible Watermarking By Neural Networks" Second International Congress On Technology, Communication And Knowledge (ICTCK 2015) November, 11-12, 2015

[2] Asmaa Qasim Shareef, Ph.D Roaa Essam Fadel, "An Approach of an Image Watermarking Scheme using Neural Network", University of Baghdad University of Baghdad at International Journal of Computer Applications (0975 – 8887) Volume 92 – No.1, April 2014.

[3] W. Zheng *et al*., "Robust and high capacity watermarking for image based on DWT-SVD and CNN", *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Wuhan, 2018, pp. 1233-1237.

[4] E. Elbasi and V. Kaya, "Robust Medical Image Watermarking Using Frequency Domain and Least Significant Bits Algorithms", *2018 International Conference on Computing Sciences and Engineering (ICCSE)*, Kuwait City, 2018, pp. 1-5.

[5] G. Kulkarni and S. Kuri,“Robust digital image watermarking using DWT, DCT and probabilistic neural network”, *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, Mysuru, 2017, pp. 1-5.

[6] D. Muñoz-Ramirez, V. Ponomaryov, R. Reyes-Reyes, V. Kyrychenko, O. Pechenin and A. Totsky, “A robust watermarking scheme to JPEG compression for embedding a color watermark into digital images”, *2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Kiev, 2018, pp. 619-624.

[7] H. Sadreazami and M. Amini, “A Robust Image Watermarking Scheme Using Local Statistical Distribution in the Contourlet Domain”, *IEEE Transactions on Circuits and Systems II: Express Briefs.*

[8] O. Göker, N. Nazli, M. M. Erol, R. Choupani and E. Dogdu, “A Robust Watermarking Scheme Over Quadrant Medical Image in Discrete Wavelet Transform Domain”, *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, Thessaloniki, 2018, pp. 277-282.

[9] Naskar R., Chakraborty R.S. (2011) Reversible Image Watermarking through Coordinate Logic Operation Based Prediction. In: Jajodia S., Mazumdar C. (eds) Information Systems Security. *ICISS 2011*. Lecture Notes in Computer Science, vol 7093. Springer, Berlin, Heidelberg.