M.Tech (Computer Science & Engineering)

MANAS KAINTH

2023

# SELF ATTENTION CONVOLUTIONAL NEURAL NETWORK (SACNN) BASED PREDICTOR FOR REVERSIBLE DATA HIDING IN IMAGES

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

**MASTER OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**


Submitted by:

**MANAS KAINTH**

**2K21/CSE/13**

Under the supervision of

**Dr. RAJEEV KUMAR**

**(ASSISTANT PROFESSOR)**




**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042


JUNE, 2023

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Bawana Road , Delhi -110042

# CANDIDATE'S DECLARATION

I, Manas Kainth, Roll No. 2K21/CSE/13 of M.TECH Computer Science and Engineering, hereby declare that the Project titled **"SELF ATTENTION CONVOLUTIONAL NEURAL NETWORK (SACNN) BASED PREDICTOR FOR REVERSIBLE DATA HIDING IN IMAGES"** is being submitted by me to Delhi Technological University, Delhi, in partial fulfilment of the requirements for the degree of Masters in Technology in Computer Science and Engineering is a legitimate record of my own work and is not copied from any source. The work contained in this report has not been submitted at any other University/Institute for the award of any degree.

Place: Delhi **Manas Kainth**

Date: May, 30, 2023 **2K21/CSE/13**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Bawana Road , Delhi -110042

# CERTIFICATE

I, hereby certify that the project work entitled **"SELF ATTENTION CONVOLUTIONAL NEURAL NETWORK (SACNN) BASED PREDICTOR FOR REVERSIBLE DATA HIDING IN IMAGES",** submitted by Manas Kainth, Roll no 2K21/CSE/13, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirements for award of the degree of M.Tech in Computer Science and Engineering is a genuine record of the work done by him under my supervision. To my best knowledge this work has not been submitted in part or full for award for any other degree in this university or elsewhere.

Place:

Date: May, 30, 2023

**Dr. Rajeev Kumar**

**Assistant Professor**

# ABSTRACT

The traditional image predictors used in reversible data hiding (RDH) schemes are limited by their inability to capture context from a larger set of image pixels. This work, proposes a new predictor for reversible data hiding, that uses the Self-Attention Convolutional Neural Network (SACNN) for improvement in prediction process. With the help of an image division scheme, our method divides a grayscale image into two separate sets. The first set serves as input to the SACNN predictor, which predicts the second set, which will be utilized for data embedding. Our network efficiently captures both local and global dependencies through the use of self-attention mechanism in combination with convolutional layers, allowing for accurate pixel prediction and improving overall prediction accuracy. The predictor is trained on over 1000 images randomly taken from the ImageNet dataset. The experimentation and analysis show that the predictor is able to generate a sharper prediction error histogram and can be utilized for achieving a better embedding performance by using it with a suitable scheme in future.

# ACKNOWLEDGEMENT

I would like to express my gratitude and thanks to my advisor **Dr. Rajeev Kumar**, Assistant Professor, Department of Computer Science and Engineering, Delhi Technological University, Delhi for providing me with invaluable guidance and support throughout this work. His insights and advice have been invaluable to me in completing this work.

Manas Kainth

2K21/CSE/13

# TABLE OF CONTENT

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **RDH** | Reversible Data Hiding |
| **HS** | Histogram Shifting |
| **DE** | Difference Expansion |
| **PEE** | Prediction Error Expansion |
| **PVO** | Pixel Value Ordering |
| **CNN** | Convolutional Neural Network |
| **MEDP** | Median Edge Direction Predictor |
| **GAP** | Gradient Adjusted Predictor |
| **IP** | Interpolation Predictor |
| **MP** | Multiple Predictor |
| **CNNP** | Convolutional Neural Network Predictor |
| **SACNN** | Self-Attention Convolutional Neural Network |
| **PDE** | Partial Difference Expansion |
| **RP** | Rhombus Prediction |

# CHAPTER 1

# INTRODUCTION

## 1.1 Overview

Ensuring solid data security has become very important in the modern digital era. It is now crucial to protect data against unauthorised access, disclosure, and manipulation due to the exponential increase in data generation and exchange. Data security is essential for safeguarding sensitive information, upholding privacy, building trust, and reducing cyber risks. An essential part of data security is data hiding, which involves hiding sensitive information from prying eyes by blending it seamlessly with other data. Traditional data hiding methods suffer from a significant limitation – the irreversible nature of the embedding process. Once data is hidden using conventional techniques, it becomes challenging or impossible to extract the cover data without impacting image quality. To address this, reversible data hiding (RDH) based techniques have been used. RDH techniques play a vital role in information hiding, enabling the embedding of data within digital signals such as images, audio, and videos. The distinctive characteristic of RDH is that the hidden data can be recovered without information being lost, allowing for perfect reconstruction of the original data. Image authentication, medical imaging, and military imaging are just a few of the fields where this technique has found use [1].

Researchers in the field of reversible data hiding have primarily concentrated on two approaches. The first method uses methods like Difference Expansion [2], Histogram Shifting [3][4][5][6], and Prediction Error Expansion [7][8][9] to embed data into images with the least amount of distortion possible. While the second approach gives emphasis to develop predictors with high prediction accuracy. This allows it to accommodate secret data without introducing noticeable changes to the cover media by leveraging redundancy within prediction errors.

Prediction error expansion (PEE) is a popular method employed in the field of reversible data hiding, allowing for the confidential embedding of supplementary information within the errors origination during prediction process of a predictor. The

underlying principle of PEE involves modifying the prediction errors, which represent the discrepancies between the predicted values given by the predictor and the original values of the cover image. By expanding the range of these prediction errors, the embedding capacity is increased while minimising any perceptible impact on the perceptual quality of the host image. The PEE method focuses on increasing prediction accuracy by using predictors that anticipate the current pixel using nearby pixels as a guide. Traditional predictors like the "Difference Predictor (DP) [2], Median Edge Predictor (MEDP) [10], Gradient Adjusted Predictor (GAP) [10], Rhombus Predictor (RP) [8], Partial Difference Expanding Predictor (PDEP) [11], and multiple predictors [12] have been proposed." [14]. These predictors, however, have a significant flaw in that they only take into account a small context when predicting the current pixel. A larger reference should be taken by covering more surrounding pixels in order to increase prediction performance.

The use of convolutional neural networks (CNNs) to increase prediction accuracy has been explored in recent developments in reversible data hiding. Luo et al. [19] presented a CNN-based method for stereo images by leveraging the correlations between the left view and the right view. Recently, CNN-based predictors have been developed, such as Hu et al.'s CNNP [15], which makes use of the large receptive fields [15] and global optimization capabilities of CNNs. To further improve prediction accuracy, Yang et al. [16] suggested a new CNN-based predictor.

CNNs have a local receptive field that is constrained by their kernel size, which makes it difficult for them to capture information beyond their locality. As a result, they have trouble capturing long-range dependencies. Adding more layers or expanding the kernel size are two techniques for expanding a CNN's receptive field, but they frequently increase training time and computational requirements without significantly improving performance. To address these limitations, self-attention based networks have emerged as a powerful approach for capturing long-range dependencies compared to CNNs. Wang et al. [17] proposed a non-local neural network for video classification that uses non-local blocks to capture dependencies between pixels in a longer range. In order to perform image translation tasks, Zhang et al. [18] developed the self-attention generative adversarial network (SAGAN). Convolutional neural networks with self-attention were used by Yan et al. [20] to improve MRI image reconstruction.

In this study, we propose a novel predictor for RDH. The proposed predictor uses an image division scheme and splits a grayscale (512x512) image into two non overlapping sets, with the first set as the input to predict the second set. By introducing self attention mechanisms and CNN layers, the proposed predictor overcomes the limitations of the traditional predictors. The image division scheme enables effective utilisation of the surrounding pixels, leading to accurate predictions and facilitating the embedding of data through an RDH scheme such as PEE. Through experimentation, we observe the effectiveness of the proposed predictor, producing a sharper prediction-error histogram and in comparison to other predictors using the same methods.

The main contribution of this study are as follows :

- Utilising Convolutional Neural Networks (CNNs) and other deep learning techniques for reversible data hiding to show their application to make use of them in this domain.
- A novel architecture has been developed that incorporates self-attention and CNN architectures to enhance the prediction accuracy for error expansion techniques.
- Preliminary experiments have shown promising results in terms of improved prediction accuracy with higher peak points, which is an important metric in prediction based RDH schemes, thus opening more gateways to apply deep learning based approaches to RDH.

This document's remaining sections are organised as follows: The related works in the area of reversible data hiding are covered in Chapter 2. Chapter 3 gives the basic overview of the concepts used. The proposed work is comprehensively explained in Chapter 4. Chapter 5 examines the results of experimentation and compares the proposed predictor with existing work. Finally, chapter 6 concludes the work, summarising the findings and outlining potential avenues for future research.

## 1.3 Problem Statement

Prediction error expansion (PEE) is a technique widely used in the field of reversible data hiding (RDH) to embed information within the prediction errors generated by predictors. Predictors, in RDH, are algorithms designed to estimate the values of pixels

or samples in a given signal, such as images or audio. In PEE, the prediction errors obtained from the predictors are modified by adding or subtracting a small value to carry the hidden data. This modification is carefully performed to ensure that the visual or perceptual quality of the host or carrier remains minimally affected. The concealed data can later be recovered by reverse computations. PEE relies heavily upon the choice and efficiency of predictors used within it.

The precision with which predictions are made directly impacts how much data can be embedded reliably and safely. It goes without saying that proficient predictors facilitate effective data hiding methods. Although there exist multiple methods that aim to enhance prediction accuracy in RDH processes, most tend to focus on predicting from a limited reference, consisting only of a few adjacent pixels. We can make even more accurate predictions by including more pixels for reference. Along with improving prediction accuracy, it's equally important that we select appropriate embedding strategies as well. Selecting an appropriate embedding strategy ensures more data to be embedded while distortion is minimised. Continued research in both prediction techniques and embedding strategies are necessary to further the advancement of the RDH community. To summarise, PEE is a popular technique used in RDH and advancements in predicting techniques that can take more reference pixels into context can lead to improvement in prediction accuracy which, incorporated with a suitable embedding scheme can achieve an optimal tradeoff between embedding capacity and perceptual quality.

On the basis of the above statements, the following research questions are identified.

1. What are the recent algorithms utilised in the prediction process in reversible data hiding ?
2. How can prediction algorithms be improved to take into account a greater number of surrounding pixels, resulting in more accurate predictions and better overall performance in data hiding techniques?
3. What embedding approaches may be used to make most of the predictors and improve the performance of reversible data hiding systems?
4. What evaluation metrics should be employed to assess the performance of predictors ?

## 1.4 Objectives

Based on the statements discussed above the study aims to investigate the recent advancements in reversible data hiding (RDH) techniques, with a particular focus on the prediction process for PEE.

1. The first research question aims to identify and analyse the state-of-the-art algorithms used in the prediction phase of RDH, understanding their strengths, limitations, and performance.
2. The second research question aims to develop novel prediction algorithms that leverage a larger neighbourhood of pixels for accurate prediction
3. Next, the study aims to focus on investigating embedding techniques that use predictors efficiently to maximise embedding capacity while maintaining data dependability and minimising perceptual impact.
4. Finally, the study aims to establish suitable evaluation metrics for assessing the performance of algorithms in the specific domain of PEE . These metrics may include prediction accuracy for predictors and distortion rate on embedding. Mean Squared can be employed to train the deep learning networks to check their performance.

The above objectives provide a framework for conducting study in the field of RDH domain, with a focus on exploring recent prediction algorithms, improving prediction accuracy, and defining appropriate evaluation metrics for assessing performance.

# CHAPTER 2

# LITERATURE REVIEW

This chapter examines the related works, beginning with predictors used in the prediction process followed by the embedding techniques used in reversible data hiding.

## 2.1 Image Predictors

Prediction methodologies, and prediction-error expansion methods have all advanced in the realm of reversible data hiding. Researchers have investigated techniques such as content-adaptive predictors, pixel-value-ordering prediction, and pairwise PEE to achieve high embedding capacity with minimal distortion. A predictor is used by prediction error expansion-based techniques to forecast pixel values. In this section, the various predictors for predicting pixels are discussed.

Difference prediction used in [2] analyses pixel values by using the difference between pair of pixels, such as the horizontal difference ($d_h$) computed as the absolute difference between the current pixel and its left neighbour. Expansion values are estimated based on these differences, determining the amount of adjustment needed to embed additional data. The difference value ($d_h$) is then classified into two categories for data embedding i.e Expandable Differences, Changeable differences. This method is based on reversible integer transform where for :

$$(m,n), m, y \epsilon P, 0 \leq m,n \leq 255, \qquad l = [(m+n)/2], h = m - n \tag{1}$$

can be recovered from their inverse transform by

$$m = l + [(h+1)/2], n = l - [h/2] \tag{2}$$

where m,n are pixel values and l = average mean of m,n and h = difference b/w m,y.

Another popular technique for estimating original pixel values and computing prediction errors is the median edge detector (MED). To predict the values of the target pixels, the median edge direction predictor (MEDP) [8] makes use of a built-in

edge detection mechanism involving three nearby pixels. It is a crucial part of the LOCO-I algorithm, which is renowned for its efficient compression and low computational complexity.

$$P = \begin{cases} \min(n, w) & \text{if } nw < \max(n, w) \\ \max(n, w) & \text{if } nw > \max(n, w) \\ n + w - nw & \text{otherwise} \end{cases} \tag{3}$$

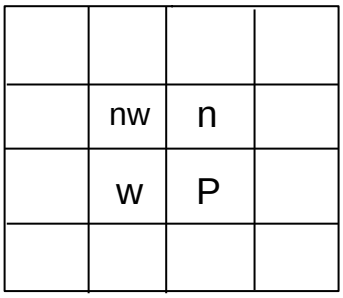



**Fig. 2.1 Reference Pixels for prediction in MED**

According to , the MED predictor selects a median value from the nearby pixels ($w, n + w = nw$). According to the template's first two rows, an edge is considered to exist when the value of $nw$ is either minimum or maximum in comparison to the other neighbouring pixels of $x$. The prediction for the currently shown pixel is either the $n$ value for vertical edges or the $w$ value for horizontal edges. The MED predictor creates the predicted value for the current pixel $x$ by taking into account the context of these three nearby pixels. This scheme is displayed in Fig. 2.1

A rhombus pattern prediction scheme was put forth in [8].The Rhombus predictor in predicts the pixel value at position $(i, j)$by considering four reference pixels: the left neighbour ($P_{i,j-1}$), the right neighbour ($P_{i,j+1}$ ), the top neighbour ($P_{i-1,j}$), and the bottom neighbour ($P_{i+1,j}$).

The predicted pixel value ($P'_{i,j}$) is obtained as follows:

$$P'_{i.j} = \frac{P_{i,j-1} + P_{i,j+1} + P_{i-1,j} + P_{i+1,j}}{4} \tag{4}$$

The prediction error ($E_{i,j}$) is calculated as the difference between the actual pixel value ($P_{i,j}$)and the predicted pixel value:

$$E_{i,j} = P_{i,j} - P'_{i,j} \quad (5)$$

During the data embedding process, the expanded prediction error ($E'_{i,j}$) is computed by adding the embedding value (embedding_value) to the prediction error:

$$E'_{i,j} = E_{i,j} + embedding\_value \quad (6)$$

Ou et al. [11] predict the target pixel using a partial differential equation (PDE)-based predictor. The method makes use of PDE to prioritise pixels with higher correlation over those with lower correlation. The four nearest pixels ($X_{i-1,j}, X_{i,j-1}, X_{i+1,j}, X_{i,j+1}$), are used to predict the centre pixel $X_i$. Using the gradients between each pixel and the target pixel, the context of the four pixels is calculated as:

$$P_{i,j}^{k+1} = P_{i,j}^{k} + \lambda \sum_D \quad (C_D \Theta_D P_{i,j}^{k}) \quad (7)$$

where D represents the gradient for top, left ,right and both pixels , k is the iteration number,(i,j)gives the pixel coordinates. gives the gradient and the CD represents the weight of the gradient . The value of P is updated till convergence.

The gradient adjusted predictor (GAP) was used by D. Coltec [10] to predict pixels and mploys the gradient adjusted predictor (GAP) for pixel prediction. In this predictor, the difference of the gradient between the target pixel and its neighbours are used to predict the pixel value. This method takes reference of seven neighbouring pixels of the target pixel. It assumes that neighbouring pixels have similar values and that the intensity or colour changes gradually across the image. By considering the differences in horizontal and vertical gradients, the GAP predictor predicts the value of a target pixel. It employs conditions based on the gradient differences to assign the interpolated pixel value, adjusting it further based on additional conditions. The GAP predictor aims to provide a reasonable estimate for missing or corrupted pixels based on the available neighbouring information. The GAP algorithm works in the way given in Fig 2.2 to predict the value $\hat{P}_{i,j}$ for target pixel $P_{i,j}$

$dh = |P_{i-1,j} - P_{i-2,j}| + |P_{i,j-1} - P_{i-1,j-1}| + |P_{i,j-1} - P_{i+1,j-1}|$
$dv = |P_{i-1,j} - P_{i-1,j-1}| + |P_{i,j-1} - P_{i,j-2}| + |P_{i+1,j-1} - P_{i+1,j-2}|$
**IF** $(dv - dh > 80)$, $P_{i,j} = P_{i-1,j}$
**ELSE IF** $(dv - dh < -80)$, $P_{i,j} = P_{i,j-1}$
**ELSE**
$P'_{i,j} = \frac{P_{i-1,j}+P_{i,j-1}}{2} + \frac{P_{i+1,j-1}+P_{i-1,j-1}}{4}$
**IF** $(dv - dh > 32)$, $P'_{i,j} = \frac{P'_{i,j}+P_{i-1,j}}{2}$
**ELSE IF** $(dv - dh > 8)$, $P'_{i,j} = \frac{3P'_{i,j}+P_{i-1,j}}{4}$
**ELSE IF** $(dv - dh < -32)$, $P'_{i,j} = \frac{P'_{i,j}+P_{i,j-1}}{2}$
**ELSE IF** $(dv - dh < -8)$, $P'_{i,j} = \frac{3P'_{i,j}+P_{i,j-1}}{4}$
**END IF**

**Fig. 2.2 GAP algorithm for predicting pixels**

| | | | |
|---|---|---|---|
| | $P_{i-2,j}$ | $P_{i-2,j+1}$ | $P_{i-2.j+2}$ |
| $P_{i-1,j-1}$ | $P_{i-1,j}$ | $P_{i,-1j+1}$ | |
| $P_{i,j-1}$ | $P_{i,j}$ | | |

**Fig. 2.3 Reference Pixels for GAP**

Interpolation Predictor (IP), developed by Luo et al. in [12], embeds the bits using the interpolation error, which is the difference between the interpolated value and the corresponding pixel value, either by expanding it additively or leaving it unaltered. This method uses the interpolation error for data embedding rather than the inter-pixel differences or prediction errors found in the majority of difference expansion (DE) approaches. The difference is also expanded by addition rather than bit shifting. This method has the advantages of minimal distortion from conservative expansion (each pixel is only changed by one), the lack of a location map to identify expanded interpolation errors, and the higher expandability of interpolation errors compared to inter-pixel differences or prediction errors. This method consequently guarantees great image quality while keeping a sizable embedding capacity.. To more accurately predict the pixel values, Jafar et al. in [13] use the method of applying multiple predictors, such as edge-based predictors, linear predictors, etc.

Convolutional neural networks are used by Hu et al. in CNNP [15] to compute the predicted values. To predict the pixel values corresponding to the target pixel, their method makes use of CNN's multi-receptive field and global optimisation property. They also use an image coding scheme to divide images into two sets for prediction, as well as convolution-based feature extraction and image prediction blocks. To train the network they used over 1000 random images from ImageNet and followed the histogram shifting and error expansion mechanism for embedding data. Fig. 2.4 shows the architecture of their predictor containing CNN layers included in feature extractions and image prediction modules. In New CNNP [17], the authors use a deep CNN model to predict pixel values, which is combined with a new image coding scheme to improve prediction accuracy. They use batch normalisation layers to improve predictor performance and have also illustrated an improved new image division scheme. Their feature extraction block contains 3 CNN blocks connected in parallel to gather more context and the image prediction module contains 14 CNN blocks connected in a series. They also include an image reconstruction block to get the generated image. This is described in figure 2.5.

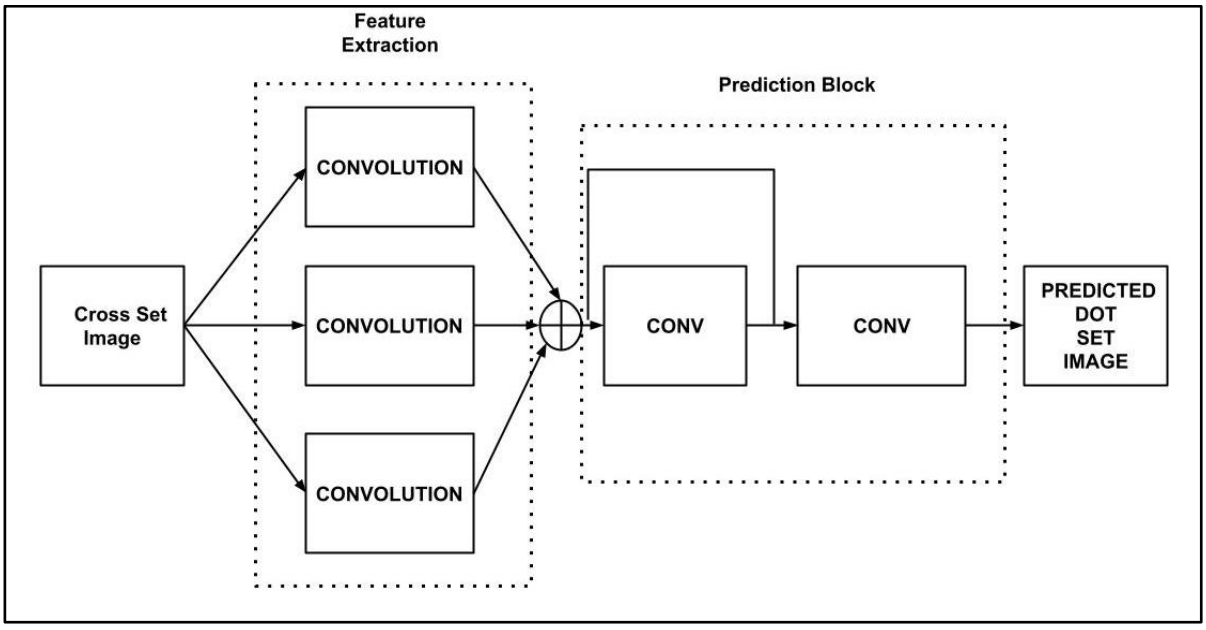



**Fig. 2.4 CNNP**

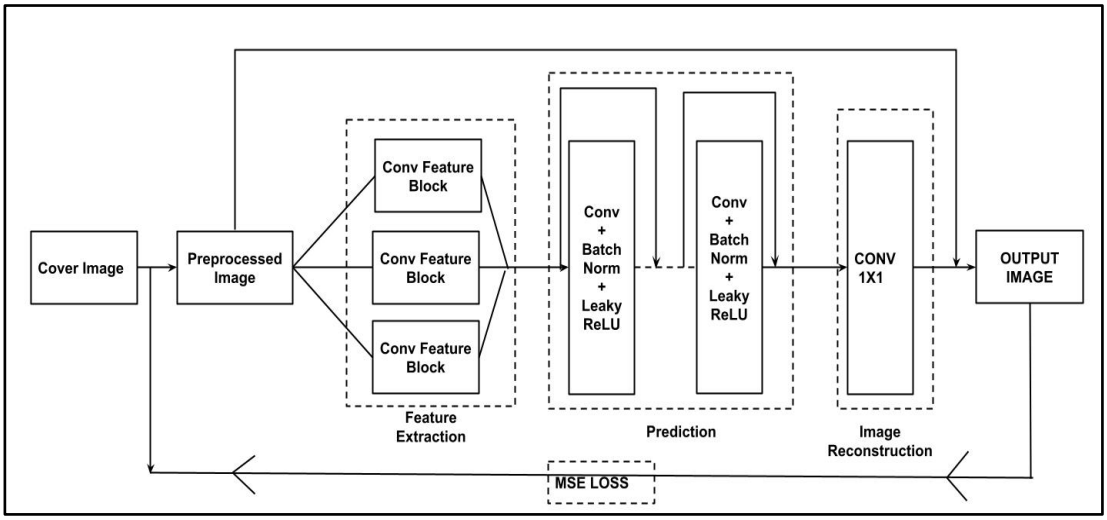



**Fig 2.5 New CNNP**

A summary of the predictors is given in table 1.

**Table 1. A review of various image predictors used for prediction in RDH**

| Predictor | Methodology | Advantages | Disadvantages |
|---|---|---|---|
| **Difference Predictor** | Uses the difference of neighbouring pixels for difference expansion. | High embedding capacity, minimal distortion. | Limited accuracy, sensitive to noise. |
| **Rhombus Predictor (RP)** | Utilises four neighbouring pixels to predict the centre pixel value. | Improved prediction accuracy, efficient embedding. | Limited context information may not handle complex patterns well. |
| **Median Edge Direction Predictor (MEDP)** | Employs three neighbouring pixels for edge detection-based prediction. | Effective edge detection, suitable for image regions with clear edges. | Limited adaptability to non-edge regions, lower accuracy for smooth regions. |

| | | | |
|---|---|---|---|
| **Partial Differential Equation Predictor (PDE)** | Uses a partial differential equation-based approach, emphasising pixel correlation. | Considers pixel correlations, more accurate in correlated regions. | Computationally intensive, may introduce artefacts in low-correlation regions. |
| **Gradient Adjusted Predictor (GAP)** | Utilises gradient-adjusted pixels for prediction, considering pixel gradients. | Improved handling of gradient variations, better adaptation to image content. | Complex implementation, higher computational requirements. |
| **Interpolation Predictor (IP)** | Predicts pixel values using an interpolation error difference scheme. | Accurate prediction in interpolated regions, handles smooth transitions well. | Limited performance in non-interpolated regions, susceptible to interpolation errors. |
| **Convolutional Neural Network Predictor (CNNP)** | Employs a deep CNN model for pixel prediction using multi-receptive fields and global optimization | High accuracy, effective utilisation of global image features. | Computationally intensive, requires extensive training data. |
| **New CNN Based Predictor** | Utilises a deep CNN model with an improved image division scheme and batch normalisation layer | Improved prediction accuracy, better performance with batch normalisation. | Higher computational requirements, additional complexity in image division scheme. |

## 2.2 Data Embedding Schemes

Reversible data hiding (RDH) depends critically on the choice of an embedding scheme, which affects elements like embedding capacity, perceptual quality, robustness, compatibility, and security. The chosen scheme decides how much information may be concealed while minimising perceived distortion. It must be resistant to assaults and appropriate for the cover signal kind. The incorporation of security elements and steganalysis considerations is also possible. Therefore, to attain the ideal balance in RDH, a careful selection process is required. The three main techniques for reversible data hiding are difference expansion, histogram shifting, and prediction error expansion. The most recent research on these techniques is examined in this section.

Reversible data embedding through difference expansion, introduced by Tian [2], ensures minimal distortion and precise recovery of the original image. However, this method is only applicable to grayscale images. One of the most widely used techniques in RDH is histogram shifting. Ni et al. [23] present a“histogram modification-based lossless embedding and high storage capacity reversible data hiding method.”However, during the data hiding procedure, distortion could happen. Pan et al. [24] propose a method for reversibly hiding data that combines multilayer embedding with local histogram shifting. This approach guarantees a continuous histogram and better security. The embedded cover image's histogram is made nearly identical to the original cover image's histogram using localization techniques to hide it from attackers. Wang et al.'s [25] proposed rate and distortion optimisation approach employs several histogram shifts and adaptively calculates the peak and zero bins while improving distortion. It requires a lot of computing power and time to assess this procedure. Kim et al. [28] advise skewed histogram shifting to lessen distortion and enhance the utilisation of skewed histograms in RDH

Predicting a given image and extending the prediction errors are two major methods used for prediction-error expansion. Using the “prediction-error expansion and histogram shifting,” Thodi et al. [7] present a novel reversible watermarking method. It provides a lossless authentication solution for digital photographs, allowing the original content to be recovered following watermark extraction.

"A lossless watermarking algorithm" for images is put forth by Sachnev et al. [8] that, in the majority of cases, does not call for a position map. The method employs prediction mistakes to merge input into an image. The method allows for more data to be included in the image while minimising distortion by prioritising prediction mistakes based on their local variance magnitude and, on occasion, adopting a smaller location map. The reversibility of this approach is ensured by the use of a double-layered embedding mechanism. Initially, the cover image is divided into two sets in this mechanism: 'cross' and 'dot'. Following the separation procedure, both cross and dot sets are inserted within the half portion of the secret message. Additionally, a sorting mechanism is incorporated to boost performance.

Li et al. [11] present "a pixel-value-ordering (PVO), a novel prediction strategy, and the well-known prediction-error expansion (PEE)"technique-based high-fidelity reversible data hiding system for digital images. Ou et al.'s [9] proposal of a prediction-error expansion based RDH uses a novel predictor based on partial differential equations. PEE uses data embedding to individually adjust prediction-errors. Instead of considering each prediction-error individually, Ou et al[11] propose that every two adjacent prediction-errors be considered jointly to generate a sequence of prediction-error pairs. This will enable for more effective use of these correlations. Zheng et al. [26] developed a local predictor that uses the content-adaptive block size for each to-be-predicted pixel to obtain a more accurate prediction by accounting for the presence of edges and textures in natural images. To provide as much information as possible, Caciula et al. [27] considered the distinct prediction error expansion (PEE) of each pixel up to the overflow/underflow limit. He et al. [30] also propose a flexible spatial location-based PEE predictor that makes the best use of spatial correlation. Although some of these prediction-based techniques are computationally expensive and only work with digital images, they allow for high fidelity steganography and improved prediction accuracy. He et al. [32] studied pixel value ordering prediction-based prediction-error expansion.. By enhancing PVO prediction in the fields of spatial correlation and correlated pixel pair, a novel prediction method is suggested. RDH was examined by He et al. [33] using dual pairwise prediction-error expansion. In order to fully utilise pairwise PEE, a dual pairwise PEE technique is created. Hu et al. [34] uses prediction error ordering (PEO) based adaptive embedding technique for RDH. The above studies are summarised in Table 2.

**Table 2. Summary of reversible data hiding techniques**

| Author | Methodology | Advantages |
|---|---|---|
| Tian [2] | difference expansion | Low distortion, exact recovery of original image |
| Thodi and rodriguez [7] | prediction-error expansion | Lossless authentication, recovery of original content |
| Ni et al. [23] | histogram modification | Lossless embedding, high data capacity |
| Sachnev et al. [8] | sorting and prediction using two stage embedding | High embedding capacity, less distortion |
| Ou et al. [21] | Pairwise PEE for efficient reversible data hiding | Better utilisation of prediction error correlations |
| Pan et al. [24] | local histogram shifting with multilayer embedding | Continuous histogram, enhanced security |
| Wang et al. [25] | Rate and distortion optimization using multiple histogram shifting | Adaptive determination of peak and zero bins, distortion optimization |
| Zheng et al. [26] | Content-adaptive local predictor using variable block size | Accurate prediction |
| Caciula et al. [27] | Distinct PEE for each pixel up to overflow/underflow limit for maximum data embedding | High embedding capacity |
| Kim et al. [28] | Skewed histogram shifting for using a pair of extreme predictions | Reduced distortion, improved utilisation of skewed histograms |

| | | |
|---|---|---|
| He et al. [29] | PVO prediction considering spatial correlation and correlated pixel pair | Improved prediction accuracy, high fidelity steganography |
| He et al. [30] | Flexible spatial location based PVO predictor | Optimises utilisation of spatial correlation, high fidelity steganography |
| Shaik et al. [31] | wavelet transform and PEE | High embedding capacity, improved image quality |
| He et al [32] | Dual pairwise PEE strategy | Fully exploits potential of pairwise PEE, high embedding capacity |
| He et al. [33] | PEE and adaptive block size enable reversible data hiding | High embedding capacity, improved image quality |
| Hu et al. [34] | PEO based adaptive embedding | Intelligent predictor combined with adaptive two stage embedding gives high embedding capacity |

## 2.3 Discussion on literature

The predictors and the prediction process play a crucial role in prediction error expansion (PEE) within reversible data hiding. Predictors must accurately estimate pixel values in order to give reliable prediction errors, which serve as the foundation for embedding hidden data. The embedding strategy equally plays an important part in RDH for better utilisation of the predictors. The following aspects highlight the importance of predictors and embedding strategy:

1. **Prediction Accuracy**: The accuracy of predictors directly impacts the quality of the prediction error expansion technique. The underlying dependencies and patterns in the cover image can be effectively captured by highly accurate predictors, leading to predictions that are more accurate. Predictors need to incorporate the mechanisms that enable the utilisation of both local and global references. Traditional predictors frequently use a small set of nearby pixels in

their calculations. Convolution and self-attention, on the other hand, enable the prediction process to cover a larger pixel neighbourhood. This broader perspective enables a more thorough analysis of image characteristics and improves prediction precision, leading to an improvement in PEE embedding performance.

2. **Reversibility and Extraction of Hidden Data:**. The usage of the embedding scheme is also important to facilitate better embedding capacity. The maximisation of embedding capacity and reduction of visual impact in reversible data hiding heavily relies on the embedding scheme chosen for prediction error expansion (PEE). The prediction process in PEE facilitates the reversibility of data hiding. The original cover image can be completely recovered without any loss or distortion. This manipulation enables additional data to be hidden while maintaining the visual integrity of the cover image. Visual quality and embedding capacity are intertwined and choosing the right embedding scheme is crucial. An efficient embedding scheme must strike a balance between different factors such as capacity, visual quality, robustness etc. The two stage embedding scheme introduced in [8] is a go to scheme for many approaches.

Overall, because they have a significant impact on both the quality of predictions made and their accuracy, predictors and the prediction process are essential elements of PEE. The use of local and global references, the incorporation of the self-attention mechanism and convolution operations, and the alleviation of limitations in classical predictors allow PEE to produce more reliable and accurate data embedding. Predictor accuracy is crucial to the prediction error expansion technique because it affects both the accuracy of hidden data extraction and the process's ability to be undone. Secondly an appropriate RDH scheme must be used to make better use of the predictors. The choice of RDH scheme will have a significant impact in the data hiding process and the right scheme can allow for greater embedding capacity with less distortion which can make full use of an improved predictor.

# CHAPTER 3

# PRELIMINARIES

This chapter discusses the basic theoretical concepts involved in this study, to enable the reader to grasp the underlying principles and ideas included.

## 3.1 Data Security

Data protection from unauthorised access, manipulation, and theft has grown to be a serious challenge as a result of the growing reliance on digital platforms for communication, storage, and transactions. Confidentiality, Availability and Integrity are an important issue in data security and must be ensured to meet the security standards. Together these three are referred to as the CIA triad and provide a comprehensive framework for data security requirements. Confidentiality ensures the data to remain secret and can only be accessed by authorised parties. The main aim of confidentiality is to restrict unauthorised access and leakage of the confidential data. Different confidentiality measures like encryption , access control mechanisms, and secured networks are applied to maintain confidentiality. Integrity protects the data from unintended modification and relates to preserving data till its existence from data corruption and non authorised modification. To ensure integrity is preserved, various mechanisms can be employed such as access control, checksums, digital signatures etc. The third principle 'Availability' refers to the availability of information without failure to those who need it and are authorised to see it. The major requirements to ensure availability requires fault tolerance, data protection and recovery, preventing outage or denial of service.

## 3.2. Data Hiding

An essential component of data security, data hiding refers to methods for securing the transmission, storing, and protection of sensitive data by blending data into other data. Numerous data-hiding techniques have developed over time, including steganography, watermarking, and cryptography, each with their own advantages and goals.

Cryptography, protects confidentiality and integrity, preventing unauthorised access, and maintaining data integrity during transmission or storage by encrypting the original data with a cryptographic key.Application areas for cryptography include secure communication, data protection, and access control, offering a strong defence against data breaches. Contrarily, the main functions of watermarking are content protection and copyright enforcement. Watermarking prevents unauthorised use, keeps track of copyright violations, and provides proof of ownership. Another data hiding technique is steganography, which conceals information within a carrier medium, making the presence of hidden data undetectable. Unlike cryptography, which makes data unreadable, steganography keeps data undetectable. Steganography makes sure that hidden data evades visual or statistical analysis by taking advantage of human perception limitations and the statistical characteristics of the cover medium. Steganography can be less reliable than watermarking or cryptography, despite still being an effective method of data hiding.Hybrid approaches combine several data encryption methods, minimising the drawbacks of each while maximising the benefits. [38]

Sensitive data is protected from breach by security, which offers protection from unauthorised access, when evaluating data storage techniques. Security protects sensitive data from breach by providing protection from unauthorised access. The ability of a method to withstand any modification or attack without compromising hidden data is referred to as robustness. This promise that the concealed information t won't be visible to viewers who are humans. The capacity of a carrier medium refers to the volume of data that can be carried in it. For the embedding and extraction procedures, computational efficiency necessitates computational resources. Researchers and practitioners are assisting in the development of safe and effective ways to protect sensitive information in an increasingly connected world by continuously improving data security techniques.

## 3.3 Steganography

The art of steganography involves concealing private data in seemingly harmless objects like photos, audio files, films, or text. Steganography has a long history, dating back to the time when messages were hidden in various ways. These techniques included hidden compartments, tattooed hair, microdots, and invisible ink.

Steganography in the digital age has been developed in order to exploit the resources and limitations of digital media. Steganography is used in numerous fields and has numerous applications. It is widely used in situations when secrecy is essential, such as secret operations, military-intelligence communications, and information hiding in digital forensics. Steganography is also useful in digital watermarking, which embeds undetectable information into media files to claim ownership or confirm authenticity. Steganographic techniques must constantly advance in order to avoid detection and fend off attacks from enemies and steganalysis.

Steganography can be used to embed confidential information in the manner described below:

1. Select a cover medium, like an image.
2. Select data to embed, called secret message
3. A function *f(i)* for embedding secret message into the cover medium

Fig 3.1 shows this process, where $C_m$ is the cover media, $S_m$ is the secret message to be embedded, *F(x)* is the function to embed called the steganographic function and $S_o$ is the output stego data.

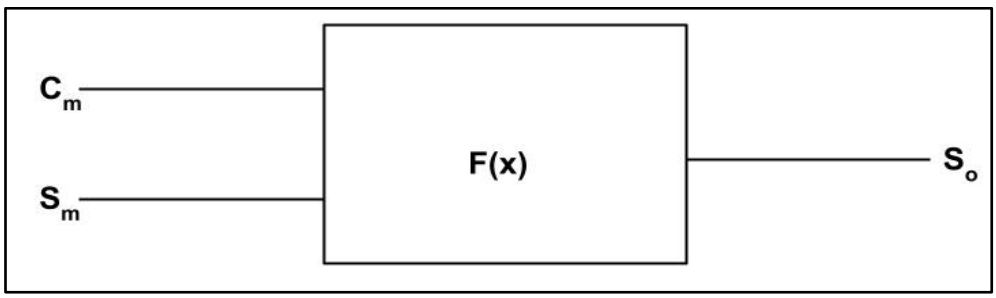



**Fig. 3.1 Steganography Process**

## 3.4 Data Hiding Techniques

In the realm of information hiding, two main categories can be identified: reversible data hiding and irreversible data hiding. Reversible data hiding methods allow the full recovery of the original data, whereas irreversible data hiding techniques involve embedding data in a way that prevents its full restoration.

1. Irreversible data hiding
2. Reversible data hiding

The classification of data hiding schemes is shown in Fig. 3.2, which also highlights the most popular data hiding techniques.

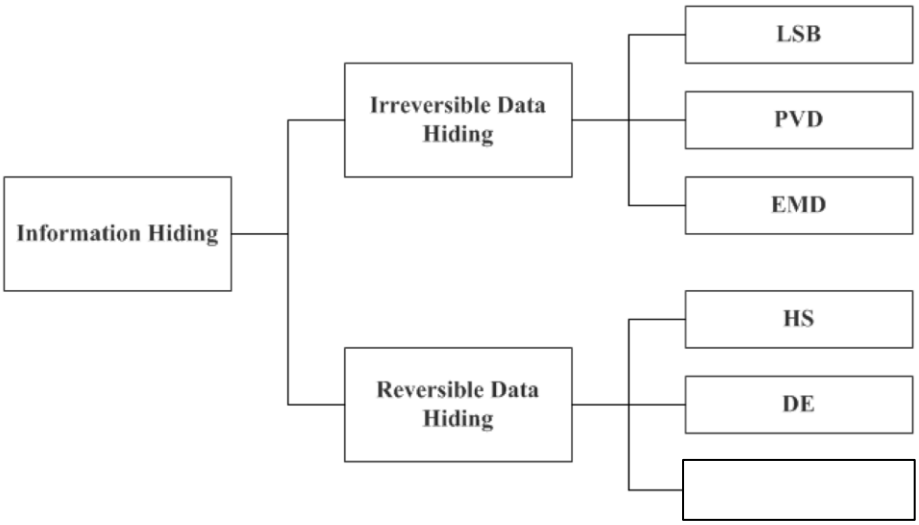



**Fig 3.2. Data Hiding Techniques**

### 3.4.1 Irreversible Data hiding

Irreversible data hiding techniques involve embedding extra information into a host signal or data stream in such a way that it cannot be completely restored to its original form. These methods are frequently employed when precise data recovery is not necessary but when greater embedding capacity and undetectability of the hidden data are desired. The irreversible data hiding process is shown in Fig. 3.3. In irreversible data hiding, there are several techniques such as LSB substitution, PVD, and EMD. Each method has its own characteristics and applications, providing different schemes to hide information with different requirements for capacity and quality. The least-redundant method of bit substitution (LSB) is a well-known method of irreversible data hiding. This approach provides a large payload capacity and minimal distortion. This is achieved by replacing the cover pixels' least significant bits with secret bits.. Wu and Tsai devised the pixel-value differencing (PVD)[39] approach by calculating the difference between two pixels in the cover image. Using a predefined range table, the number of bits to be embedded in these pixels is then computed. The PVD approach achieves a high degree of imperceptibility while embedding a considerable amount of secret bits into the cover image.

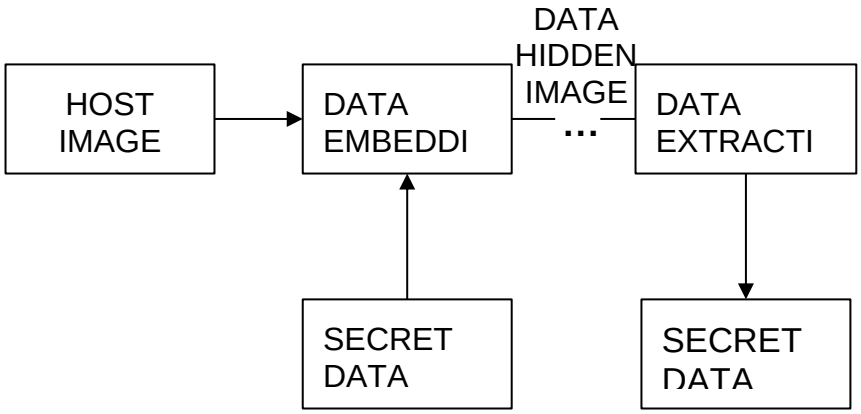



**Fig. 3.3 Irreversible Data Hiding Process**

The irreversible data hiding techniques can be applied for use depending on the type of cover signal, required capacity and distortion acceptable. Although irreversible data hiding techniques have advantages in terms of embedding capacity, they are not appropriate for applications that require full recovery of original data which takes us into the domain of reversible data hiding.

### 3.4.2 Reversible Data Hiding

Reversible data hiding schemes are used to hide secret information into a cover data and ensures that the original cover data can be recovered after the extraction of the hidden data. These techniques have a special advantage over irreversible data hiding methods because they allow for original data recovery while preventing major distortions. Fig. 3.4 depicts the reversible data hiding process. The reversible data hiding mechanism involves embedding data in such a manner that the embedded information does not cause any noticeable changes and loss of visual quality of images.

The main strategies used in reversible data hiding are :

1. Difference Expansion
2. Histogram Shifting
3. Prediction Error Expansion

A quick overview of the above techniques is provided below.

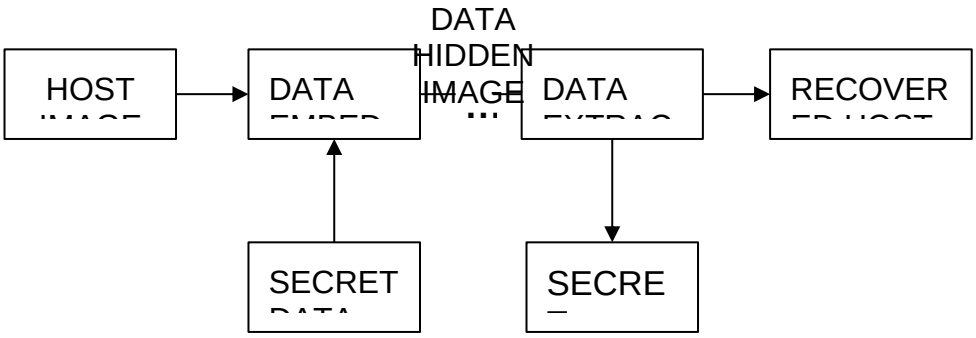



**Fig. 3.4 Reversible Data Hiding**

### 1. Difference Expansion

Difference expansion is a popular method used in RDH. This method utilises the difference between the neighbouring pixels to hide the secret data. The method works by finding the difference between pairs of adjacent pixels in the cover image and then this difference is modified to hide the secret data. During the extraction process, the modified difference is extracted and the original cover data can be restored. The process of difference expansion has several benefits for reversible data hiding. First off, it offers a high embedding capacity because the modification of differences enables the embedding of a sizable amount of data. Second, by making the changes at the pixel level and minimising visual distortions, the imperceptibility of the embedded data is maintained. Once the hidden data is extracted, the original cover image can be completely recovered without any loss or distortion.

### 2. Histogram Shifting

A common method for reversible data hiding that allows for the embedding of extra information into a cover image while preserving its visual quality is histogram shifting. This technique works by reversibly altering the cover image's histogram to make room for secret data to be hidden without introducing obvious distortions. The frequency distribution of pixel values is represented by the histogram of an image. In a grayscale image, the histogram plots the number of pixels with each possible intensity value. In order to accommodate the hidden data, histogram shifting involves moving the positions of pixel values within the histogram. Several benefits of reversible data hiding are provided by histogram shifting. First off, it offers a high embedding capacity because the modification of histogram bins enables the embedding of a sizable

amount of data. Second, rather than directly changing the pixel values, the modifications take place at the level of the pixel value, maintaining the imperceptibility of the embedded data. By doing this, the cover image's visual distortions are reduced. Histogram shifting is a versatile technique that can be used on both colour and grayscale images. However, shifting the histogram comes with some difficulties. Finding a balance between the stego image's visual quality and embedding capability can be difficult.

## 3. Prediction Error Expansion

Prediction error expansion (PEE) is a prominent technique in the field of reversible data hiding, allowing for the confidential embedding of supplementary information within the prediction errors of a predictor. The underlying principle of PEE involves modifying the prediction errors, which represent the discrepancies between the predicted values generated by the predictor and the original pixel values of the cover image. By expanding the range of these prediction errors, the embedding capacity is increased while minimising any perceptible impact on the visual quality of the cover image.

The primary benefit of PEE is that, after the embedded data is extracted, the original cover image can be completely restored without loss or distortion. Due to this characteristic, PEE has become a popular method. The prediction of pixel errors is a crucial step in many data hiding techniques, including PEE. To reduce the discrepancies between the predicted values and the actual pixel values, the accurate and advanced predictors need to be used. With the advancements in machine learning algorithms, prediction models based on artificial neural networks or other learning techniques have been developed. These models can make more precise predictions, especially in complex and non-linear image regions, because they learn the patterns and dependencies in the image data. The prediction errors, which serve as the foundation for embedding hidden data, are directly impacted by the predictor’s accuracy. Precise prediction accommodates the embedded information while reducing the visual impact on the cover image. Moreover the embedding strategies can also optimise the use of prediction for efficient embedding.

### 3.4.3 RDH embedding Scheme

Although many embedding schemes have been proposed in the literature, we give an overview of a popular two stage embedding scheme in. The two stage embedding technique based on [8], and proposed in[14] can be used for data embedding because of its effectiveness in reversibly hiding the data. The hidden data is embedded in two steps where a part of the hidden message is embedded in the first stage and the second part embedded in the next stage. After encoding the given data, the reverse process is used for data extraction..

### 1 Encoder

Two subsets of the original image are employed in the first stage, together with a trained predictor. The predictor uses the "Cross" set image ($I_c$) as input to produce the anticipated "Dot" set image ($I'_d$). The outcome is the data hidden "Dot" set image ($I_{DW}$), which is produced after a piece of the information ($W_1$) is reversibly embedded into both the initial "Dot" set image ($I_d$) and the predicted "Dot" set image ($I'_d$ ). The network is once more used in the second step to create the predicted "Cross" set image ($I'_c$) from the data-hidden "Dot" set image ($I_{DW}$) and hidden with information ($W_2$), resulting in the creation of the "Cross" set picture with hidden data ($I_{CW}$). The entire data hidden image ($I_W$) is created by merging the data hidden "Cross" set image ($I_{CW}$) and the data hidden "Dot" set image ($I_{DW}$). The structure of the encoder is given in Fig. 3.5

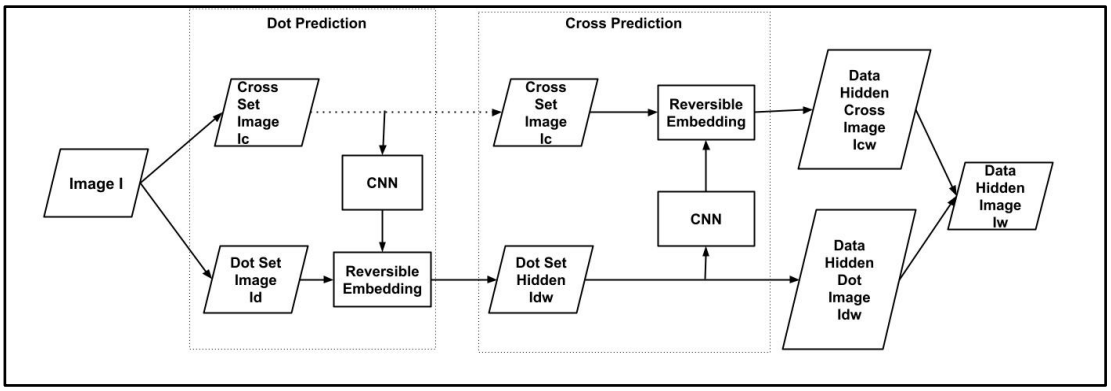



**Fig 3.5. Encoder**

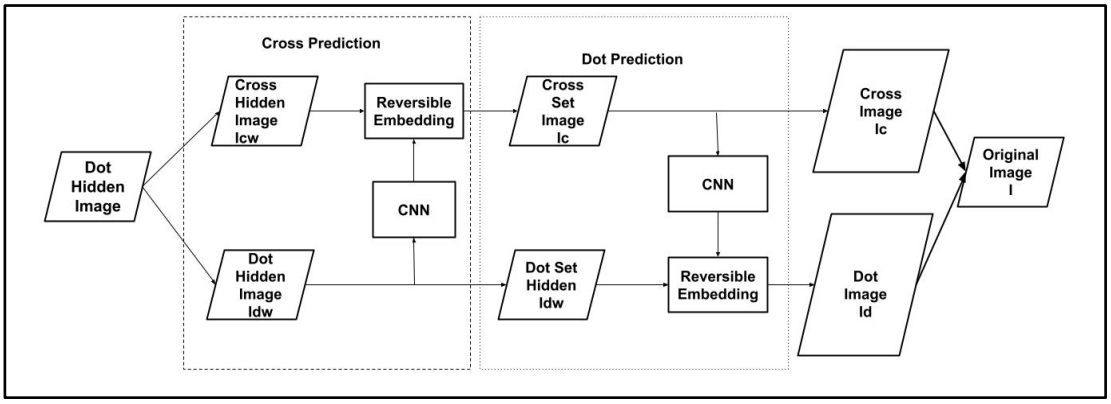



**Fig. 3.6 Decoder**

## 2 Decoder

The decoding scheme works in reverse to the encoding scheme . Firstly, the data hidden image ($I_w$) is divided into two subset images, $I_{CW}$and $I_{DW}$, following the partition pattern described later. Then, the data hidden "Dot" set image ($I_{DW}$) is processed through the proposed predictor, generating the predicted "Cross" set image ($I'_C$). By incorporating $I_{CW}$, the information ($W_2$) is extracted and used to recover the original "Cross" set image ($I_C$). Subsequently, the recovered "Cross" set image ($I_C$) is fed into the predictor, resulting in the predicted "Dot" set image ($I'_D$), which aids in recovering the original "Dot" set image ($I_D$) and extracting the information ($W_1$). Finally, the recovered images, $I_D$ and $I_C$, are spatially combined to reconstruct the original image ($I$), while the information bits ($W_1$ and $W_2$) are combined to retrieve the hidden information (W). Fig 3.6. Depicts the decoding process. Fig 3.7 and 3.8 give a brief description of the encoding and decoding process.

| Encoding Process |
|---|
| 1: Divide the original image into subsets: IC and ID. |
| 2: Generate ˜ID using the CNNP with IC as input. |
| 3: Embed information W1 into ID and ˜ID, resulting in IDW. |
| 4: Generate ˜IC using the CNNP with IDW as input. |
| 5: Embed information W2 into IC and ˜IC, resulting in ICW. |
| 6: Combine ICW and IDW to obtain the data hidden image IW. |

**Fig 3.7 Encoding Process**

| Decoding Process |
| --- |
| 1: Divide the data hidden image IW into subsets: ICW and IDW. |
| 2: Generate ˜IC using the CNNP with IDW as input. |
| 3: Extract information W2 and recover IC using ICW and ˜IC. |
| 4: Generate ˜ID using the CNNP with IC as input. |
| 5: Extract information W1 and recover ID using IDW and ˜ID. |
| 6: Combine IC and ID to reconstruct the original image I. |

**Fig. 3.8 Decoding Process**

## 3.5 Deep Learning Networks

Whether it's making Netflix recommendations or forecasting the weather, machine learning techniques are being used in almost every industry. Machine learning algorithms are a group of algorithms that learn from data and carry out a variety of tasks using statistical techniques. The machine learning algorithms can be further classified into supervised and unsupervised algorithms along with reinforcement learning. These methods, especially the deep neural networks can be applied to the RDH because of their ability to learn complex patterns and optimisation characteristics.

### 3.5.1 Convolution Neural Networks

Convolutional neural networks (CNN) are a subset of deep neural networks, which are essentially supervised learning algorithms. The capabilities of neural networks, which were originally inspired by the biological neuron, have greatly increased. Deep neural networks are now being used for a variety of tasks when there are sufficient computational resources available.

A CNN is a combination of many sequential layers that transforms images by applying activations to learn information. In a CNN, a layer is essentially made up of neurons in the width, height, and depth dimensions.

Three different kinds of layers make up a CNN architecture:

1. Convolutional Layers
2. Pooling Layers
3. Fully Connected Layers

The primary component of CNN, convolutional layers, perform convolutional operations on the input data. Two sets of data are put through a convolution operation, which enables their merging. In order to perform the convolution operation, a filter/kernel component is used. A kernel resembles a matrix that is placed over an input image to draw out its features. The kernel strides over the input image until it covers the entire image. This stride is a hyper parameter that can be changed while training. Convolution layer output is next activated using a function like the Rectified Linear Unit (ReLU).

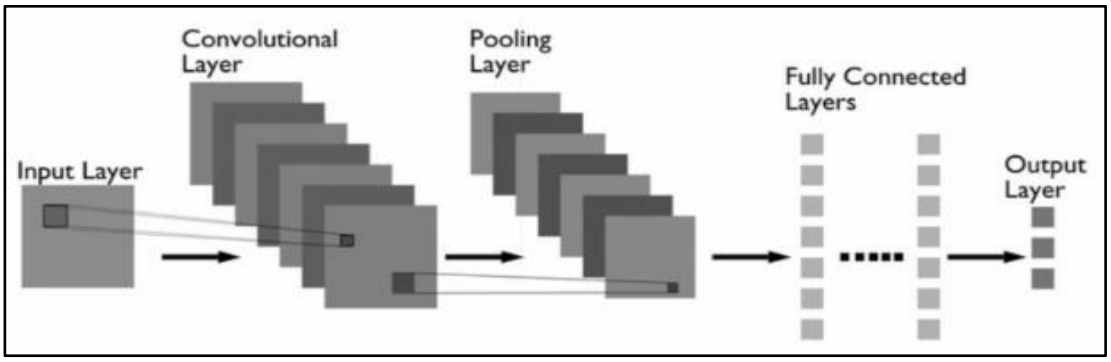



**Fig. 3.9 Convolutional Neural Network Architecture**

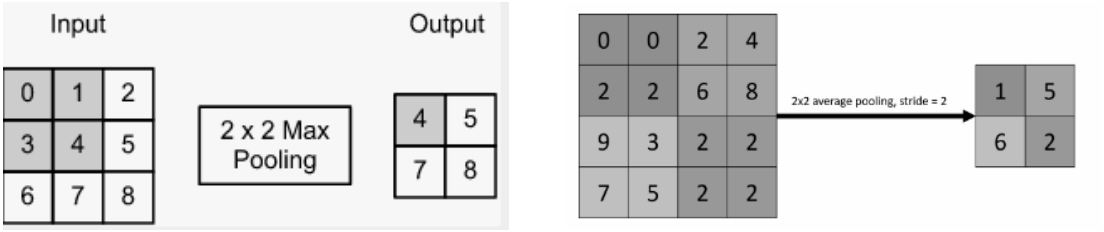



**Fig. 3.10. Pooling Operations (Max Pool and Average Pool)**

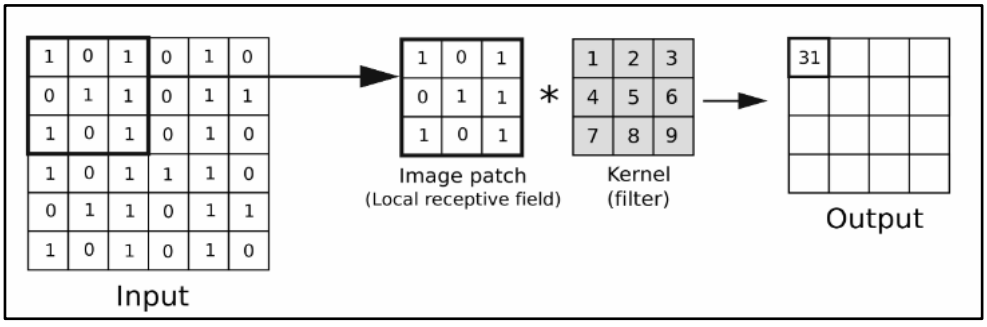



**Fig. 3.11 Convolution Operation**

The pooling layers in a CNN are used to reduce parameters in the network and perform dimension reduction. The stacking up of layers and interconnections between them can build up huge numbers of parameters. The pooling operations are performed to reduce these parameters. Maximum pooling and average pooling are the two most popular pooling methods that are used today. In maximum pooling, the maximum value from the region covered by the kernel is given as the output. In contrasts to maximum pooling, the average pooling gives the average of all parts enclosed by the kernel. Fully Connected Layers comes at the later part in the CNN. Theselayers deal with the flattened input. The neurons in these layers are fully connected with neurons in the previous layer and are mainly used for classification purposes.

**Model Parameters**

a. **Filter Dimensions** - $F\ X\ F$ sized filters with $C$ channels produces $F \times F \times C$ kernel that is applied on $I \times I \times C$ input and results in $O \times O \times 1$ size.

b. **Stride -** It denotes the number of pixels by which a kernel slides over the image after each convolution operation.

c. **Padding -** Padding applies P zeros to the boundary of the input image to balance the convolution operations.The feature map is given by

$$O = \frac{(I - F + 2P)}{S} + 1 \qquad (1)$$

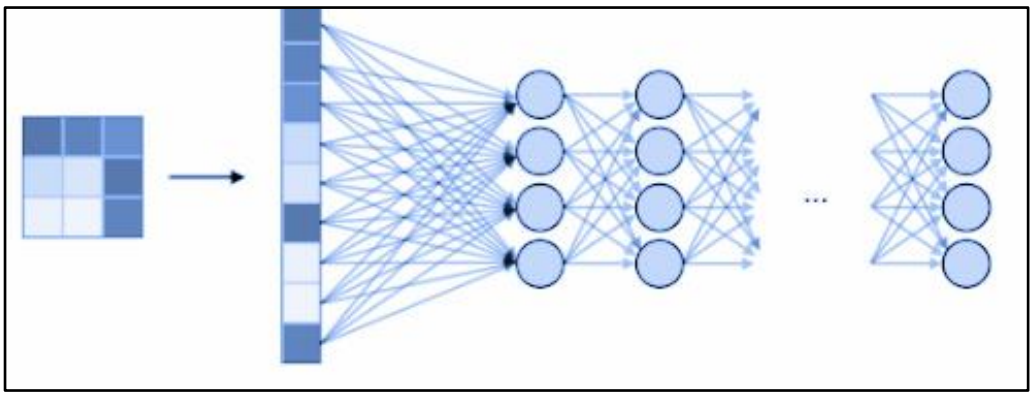


**Fig 3. 12. Fully Connected Layers**

| **Table 3. Complexity of a CNN network** | | | |
|---|---|---|---|
| | CONV | POOL | FC |
| Input Parameter | $I \times I \times C$ | $I \times I \times C$ | $N_{in}$ |
| Output Parameters | $O \times O \times C$ | $O \times O \times C$ | $N_{out}$ |
| Trainable Parameters | $(F \times F \times C + 1).K$ | 0 | $(\square_{\square\square} + 1) \times N_{out}$ |

The model complexity can be evaluated by determining the number of parameters that the model can have. This is shown in table 1.

### 3.5.2 Self-Attention

The mechanism of self-attention has drawn a lot of interest in the field of computer vision, particularly in image processing tasks. Self-attention, which was first used in the context of natural language processing, has shown to be very good at identifying long-range dependencies and modelling connections between various sequence elements. Self-attention can be used to analyse both global and local relationships in an image, which improves performance on a variety of tasks like object detection, image classification, and image generation.

Local filters are used to extract features from image regions in traditional convolutional neural networks (CNNs). Although CNNs have been incredibly successful at analysing images, they have limitations when it comes to capturing long-range dependencies across the image. With self-attention, spatial relationships between image regions can be gathered by modelling relationships between the pair of positions in the image[39], even though these regions may be distanced from each other

The calculation of attention weights, which determine the significance of different positions within the image, is the key task in self-attention. These attention weights are produced by calculating the similarity between various positions. . Query,

Key, and Value are the three inputs of self-attention. These are created from each pixel in the image and capture different aspects of image features. The query Q is the parameter representing the information for which attention is calculated. The key K is the parameter against which the attention is calculated. The value V represents the value of attention of key K for the query Q. The comparison between query and key is calculated using a metric which can be the dot product, cosine similarity etc. To obtain the attention weights, the resulting similarity metrics are processed and then passed through a softmax function. The attention weights in the end contain the modelling of the query with respect to each position of the image. Fig 3.13. displays the self-attention mechanism. The integration of self-attention into computer vision tasks have led to improvement in performance of the computer vision systems. In image recognition tasks self-attention has worked in improving the performance[40]. It has also been used in image generation tasks such as image super resolution[41]. Standalone self-attention has also been employed to replace convolution blocks in the state of the art CNNs for image related tasks.

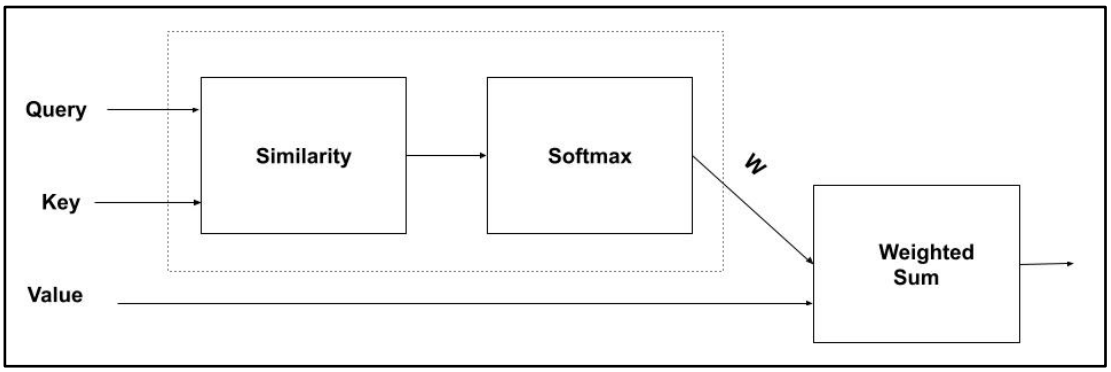



**Fig. 3. 13 Self-Attention Mechanism**

# CHAPTER 4

# PROPOSED WORK

This chapter gives a detailed overview of the proposed approach beginning with the description of the image division technique, followed by proposed architecture, the embedding technique and their implementation.

## 4.1 Image Division

Image division is performed on an image $I$ according to the chessboard pattern. Two sets of images namely, the cross set image and dot set image are created by dividing the pixels of a single image. In the cross set image $I_c$, the pixels corresponding to the dot positions are marked as 0. Similarly in the dot set image $I_d$, the pixels representing the cross positions are marked as 0. Fig. 4.1 illustrates this division procedure using a $8 \times 8$ sized block.

This division of an image into a cross and dot set makes the images in two sets independent, while still maintaining a relationship between the adjacent pixels as the value of the target pixels can be calculated by surrounding pixels. The images from the cross set are used to predict the images from the dot set till convergence.

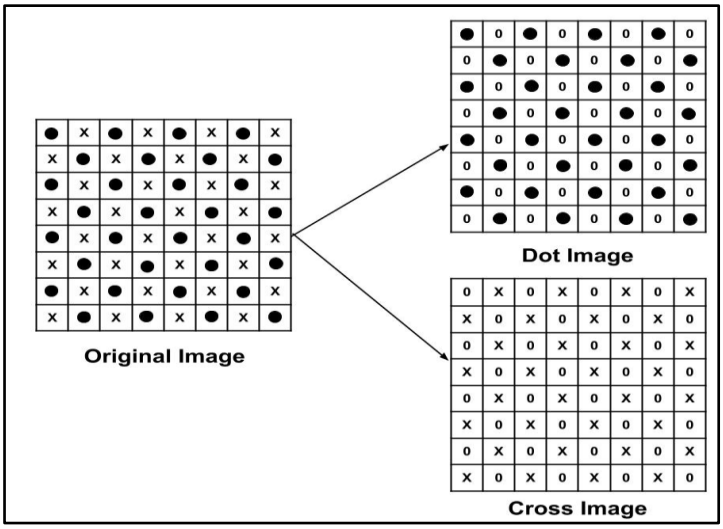



**Fig. 4.1 Image Division Scheme**

## 4.2 Proposed Architecture

Our proposed model is divided into two modules (i.e feature extraction module, image prediction module). These are shown in Fig. 4.2. The cross image $I_C$ and dot image $I_d$ are initially obtained by dividing the pixels of a grey scale cover image $I_{cov}$ as described in the above section . After feeding the cross image to the suggested predictor, the output predicted image $I_{dp}$ is created. To optimise the model's parameters, the mean squared error (MSE) between the dot image $I_d$ and the anticipated image $I_{dp}$ is used. Fig. 4.2 shows the framework in detail. Later data embedding, prediction-error expansion technique with double encoding scheme proposed in [8] has been used to evaluate embedding performance.

### 4.3.1 Feature Extraction Module

The feature extraction module uses a combination of convolution blocks and self-attention blocks. A convolution block contains *2* convolution operations and a leaky rectified linear unit ($LeakReLU$) operation. The first convolution operation has a $K \times K(K = 5,7)$ kernel size and the second convolution operation uses a kernel size of 3 respectively. The LeakyReLU layer is applied in between the two convolution operations as shown in fig. 4.3(a). The value of output channels are set to 32. Two convolution blocks ($K = 5,7$) are applied in parallel to gather more context from the images.

The feature maps generated from the convolution blocks are fed respectively to two self-attention blocks to calculate the self-attention feature maps. In a self-attention block, image feature maps $(x)$ from the previous layers are fed into three *1 x 1* convolution operations to generate the feature maps $f(x), g(x), h(x)$. Then, matrix multiplication operation is performed between the transpose of $f(x)$ and $g(x)$. The result is then fed to a SoftMax operation to generate the attention map. Then another matrix multiplication operation is performed between the transpose of the attention map and $h(x)$. At last, element wise sum of feature maps $(x)$ and transpose of attention map is computed to obtain the self-attention feature maps. The structure of the self attention block is shown in fig. 4.3(b). The feature maps obtained from convolution and self-attention blocks are then combined and fed to the image prediction module for pixel prediction.

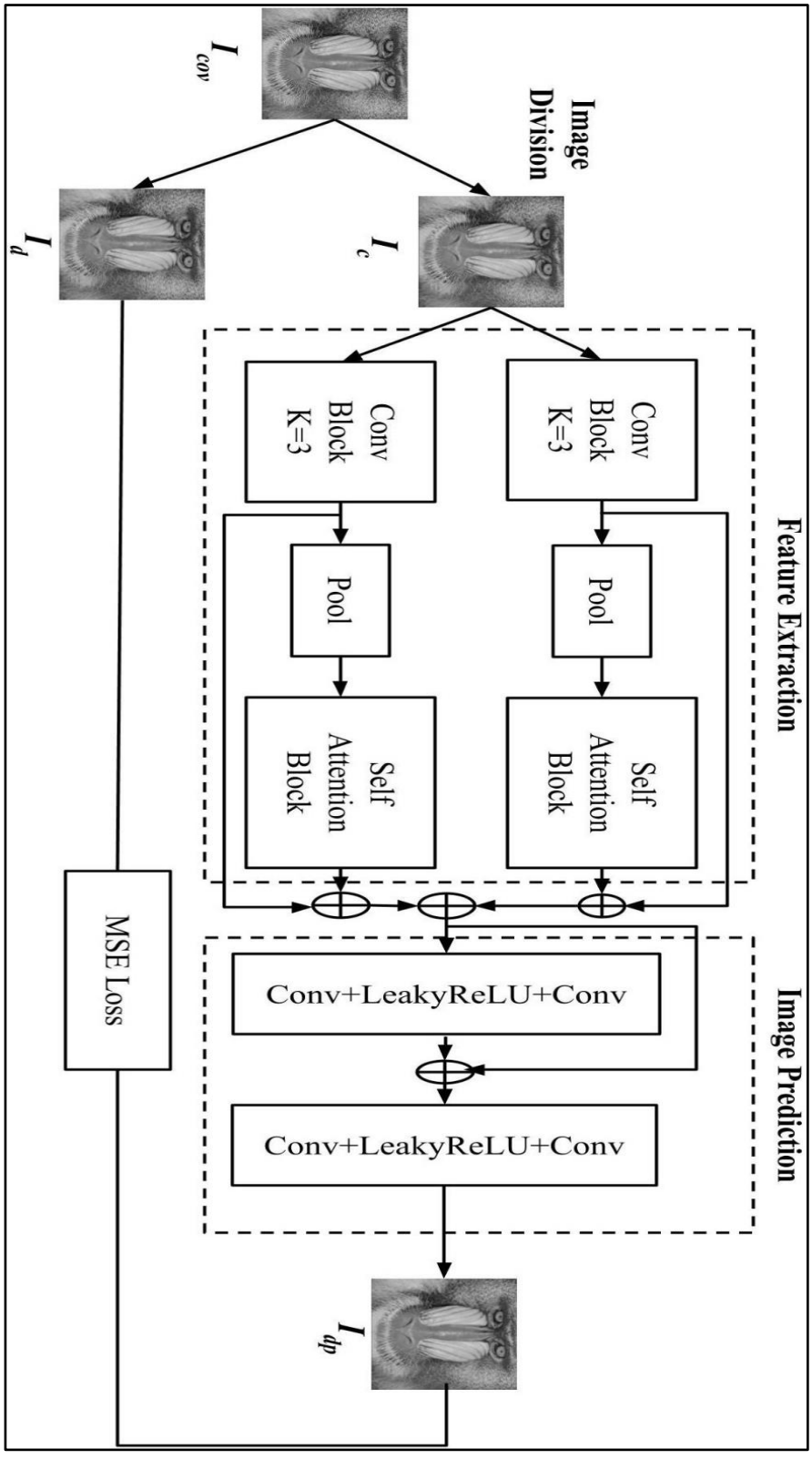
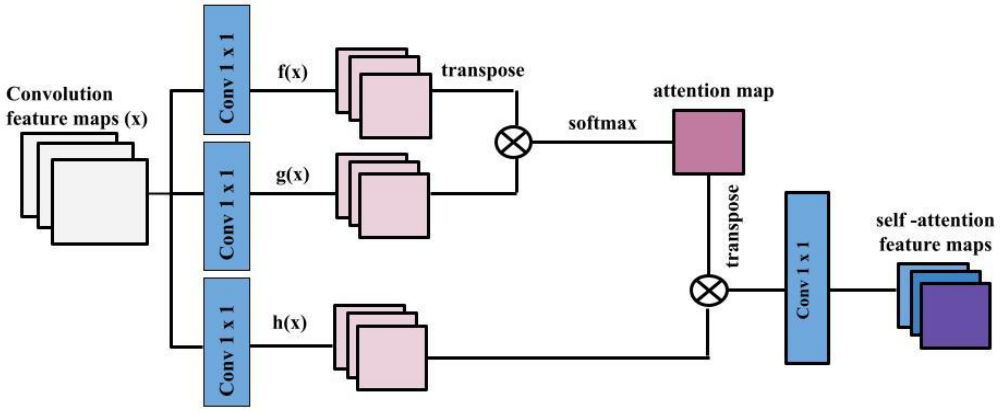



**Fig. 4.2. Proposed Architecture**


(a) Convolution Block




(b) Self-Attention Block

**Fig 4.3 Feature Extraction Block**

### 4.3.2 Image Prediction Module

The feature maps obtained from the feature extraction step are fed to two convolution blocks arranged in a sequential manner. The first block consists of a 3 x 3 convolution operation, followed by a LeakyReLU operation, followed by another 3 x 3 sized convolution operation. There are a total of 32 output channels. The output from the first block is combined with the input and fed to the second convolution block to generate the predicted image. The second block follows a similar structure to the previous block. To produce the final anticipated image, the last convolution operation sets the number of output channels to 1.

### 4.3.3 Training

The prediction model proposed is trained on over 1000 images selected randomly from the ImageNet [22] dataset. The images are first resized to *512 x 512* size and subsequently converted to grayscale. The cross set image $I_C$ is used to generate the predicted dot set image $I_{dp}$. The mean squared error (M.S.E) between the target image $I_d$ and predicted dot image $I_{dp}$is used as the loss function:

$$loss = \frac{1}{N}\sum_{i=1}^{N}(I_d - I_{dp})^2 + \lambda \parallel \omega^2 \parallel \tag{4}$$

where N refers to the count of training samples, $\lambda$ represents the weight decay applied to avoid overfitting, $\omega$ and represents all of the network's weights. In order to optimise the loss function for training with a learning rate of $10^{-3}$ and a batch size of 4, the ADAM optimizer [34] is employed. $\lambda$ is configured to have a value of $10^{-3}$. The suggested model undergoes 90 rounds of training on Kaggle's NVIDIA TESLA P100 GPU with 16 GB of memory.

## 4.4 Dataset Description

With more than 14 million labelled images spread across roughly 21,000 categories, ImageNet[23] is a highly influential dataset in computer vision with images ranging from 200 x 200 to over 1000 x 1000 pixels in resolution. Its hierarchical structure makes it possible to categorise and comprehend visual concepts at a finer level. The use of ImageNet as a benchmark for assessing computer vision algorithms enables researchers to follow developments and assess model performance.

Due to the wide usage of ImageNet in computer vision tasks, random images from the ImageNet dataset are used to train the proposed CNN based predictor by converting ImageNet images to cross and dot set images. Furthermore, to train the proposed model, a problem specific data loader is created by making image pairs (X,Y) such that X belongs to the cross set images and Y belongs to the dot set images. This data loader can be utilised for training networks for cross-dot division schemes.

# CHAPTER 5

# RESULTS

This chapter discusses and analyses the performance of the proposed SACNN based predictor on 7 standard test images frequently used in benchmarking RDH methods. To evaluate the performance of the proposed predictor we have tested it on several test images used for benchmarking purposes. These images are single channel grayscale images with pixel intensity lying in the range [0,255]. Each of these images are 512 x 512 sized images. The test images (baboon, lena , boat, peppers, barbara, ship, airplane) are displayed in fig. 5.1

**Fig. 5.1 Standard Test Images**

## 5.1. Prediction Accuracy

We have calculated the prediction error between the cover picture and corresponding predicted image for each test image. Zero prediction error denotes the ability to anticipate the current pixel with accuracy, and the amount of zero prediction errors is closely correlated with data hiding capacity, which is typically used as a benchmark to

assess the effectiveness of RDH approaches. The sharper prediction error distribution leads to less distortion in the embedded image.

In fig. 5.1, we have shown the prediction error histogram for baboon image. The value of prediction errors is represented in the horizontal axis and the vertical axis represents the occurrence of predicted errors. The Fig. 5.2 displays the prediction error histogram of the baboon image.

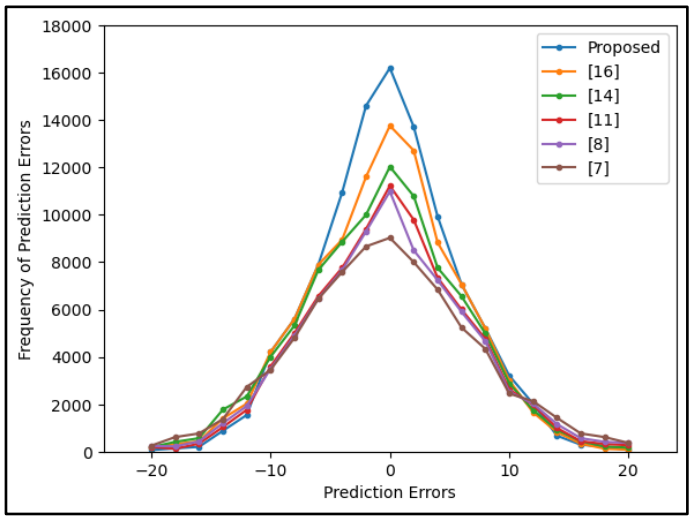



**Fig 5.2 Prediction Error Histogram for Baboon**

**Table 4. Comparative Analysis of proposed Predictor with other predictors**

| Image | [7 ] | [8 ] | [11 ] | [14 ] | [16 ] | Proposed |
|---|---|---|---|---|---|---|
| **Airplane** | 44894 | 51108 | 56740 | 57968 | 62261 | 65000 |
| **Baboon** | 9029 | 10988 | 11228 | 12032 | 13765 | 16204 |
| **Lena** | 29067 | 34886 | 35694 | 36203 | 39375 | 56497 |
| **Lake** | 15202 | 19657 | 20472 | 20409 | 25934 | 37380 |
| **Boat** | 16991 | 20712 | 21047 | 22111 | 25078 | 34673 |
| **Barbara** | 22200 | 27536 | 28530 | 31896 | 43135 | 39754 |
| **Peppers** | 16567 | 21844 | 22064 | 21959 | 28038 | 35337 |

Table 6 compares the proposed model for zero prediction errors on various test images against other predictors for comparison. Each row represents a different image, and the columns show the zero prediction errors for the images. As we can see from table 6, the proposed predictor has better zero prediction accuracy than the other predictors and thus produces a sharper error histogram as compared to the existing predictors.

Overall on comparing the given data, we can see that the proposed model achieves a higher peak point in making zero errors predictions which can be used to achieve better PSNR values using a suitable embedding scheme. Although, the predictor shows promising results, further experimentation needs to be done by combining the predictor with an embedding scheme to verify its performance.

# CHAPTER 6

# CONCLUSION AND FUTURE SCOPE

In this study, we proposed a brand new predictor for reversible data hiding. It has the ability to capture the long range of dependencies and provide precise predictions for each and every pixel. Our suggested predictor uses the self-attention mechanism and convolution operations for predicting . Both local as well as global references are taken into account by fusing convolutional operations and self-attention, which enhances its overall prediction capability. We have seen and proven through experimentation how the predictor is able to predict accurately and generate a sharper prediction error histogram. Overall, our results make it obvious that our approach is effective and has a potential to advance the field of reversible data hiding.

Even though the self-attention based predictor has shown promising improvements in image prediction accuracy, there are a number of potential areas which can be further taken for research and development for this predictor. Future research could be focused on optimising the architecture of the self-attention-based predictor. In order to improve the model's capacity size for capturing long-range dependencies for various variations of self-attention mechanisms, such as multi-headed attention or self attention with proper positioning encoding. Self-attention mechanisms can be incorporated into recurrent neural networks or transformer-based architectures for improving performance Further, this research can be focused on creating effective inference and training methods for the self-attention-based predictor. It can include techniques like approximate self-attention, sparse attention, or hierarchical attention, aiming to achieve the goal where computational complexity is low while preserving performance.A variety of embedding techniques such as adaptive embeddings, can be applied for embedding and contribute to the reversible data hiding community and also demonstrate application of machine learning techniques for the same.

# BIBLIOGRAPHY


[1] Y. -Q. Shi, X. Li, X. Zhang, H. -T. Wu and B. Ma, "Reversible data hiding: Advances in the past two decades," in IEEE Access, vol. 4, pp. 3210-3237, 2016.

[2] Jun Tian, "Reversible data embedding using a difference expansion," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 8, pp. 890-896, Aug. 2003.

[3] Zhicheng Ni, Yun-Qing Shi, N. Ansari and Wei Su, "Reversible data hiding," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354-362, March 2006

[4] J. Wang, J. Ni, X. Zhang and Y. -Q. Shi, "Rate and Distortion Optimization for Reversible Data Hiding Using Multiple Histogram Shifting," in *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 315-326, Feb. 2017

[5] S. Kim, X. Qu, V. Sachnev and H. J. Kim, "Skewed Histogram Shifting for Reversible Data Hiding Using a Pair of Extreme Predictions," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 11, pp. 3236-3246, Nov. 2019

[6] Zhibin Pan, Sen Hu, Xiaoxiao Ma, Lingfei Wang,
Reversible data hiding based on local histogram shifting with multilayer embedding,
Journal of Visual Communication and Image Representation,Volume 31,2015,

[7] D. M. Thodi and J. J. Rodriguez, "Expansion Embedding Techniques for Reversible Watermarking," in IEEE Transactions on Image Processing, vol. 16, no. 3, pp. 721-730, March 2007.

[8] V. Sachnev, H. J. Kim, J. Nam, S. Suresh and Y. Q. Shi, "Reversible Watermarking Algorithm Using Sorting and Prediction," in IEEE

Transactions on Circuits and Systems for Video Technology, vol. 19, no. 7, pp. 989-999, July 2009.

[9] B. Ou, X. Li, Y. Zhao, R. Ni and Y. -Q. Shi, "Pairwise Prediction-Error Expansion for Efficient Reversible Data Hiding," in *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5010-5021, Dec. 2013

[10] D. Coltuc, "Low distortion transform for reversible watermarking," in IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 412-417, Jan. 2012, doi: 10.1109/TIP.2011.2162424

[11] B. Ou, X. Li, Y. Zhao and R. Ni, "Reversible data hiding based on PDE predictor", *J. Syst. Softw.*, vol. 86, no. 10, pp. 2700-2709, 2013.

[12] L. Luo, Z. Chen, M. Chen, X. Zeng and Z. Xiong, "Reversible Image Watermarking Using Interpolation Technique," in IEEE Transactions on Information Forensics and Security, vol. 5, no. 1, pp. 187-193, March 2010

[13] I. F. Jafar, K. A. Darabkh, R. T. Al-Zubi and R. A. Al Na'mnehR, "Efficient reversible data hiding using multiple predictors", *IEEE Comput. J.*, vol. 59, no. 3, pp. 423-438, Mar. 2016

[14] R. Hu and S. Xiang, "CNN prediction based reversible data hiding", *IEEE Signal Process. Lett.*, vol. 28, pp. 464-468, 2021.

[15] W. Luo, Y. Li, R. Urtasun and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks", *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 4898-4906, 2016

[16] X. Yang and F. Huang, "New CNN-Based Predictor for Reversible Data Hiding," in *IEEE Signal Processing Letters*, vol. 29, pp. 2627-2631, 2022

[17] X. Wang, R. Girshick, A. Gupta and K. He, "Non-local Neural Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7794-7803

[18] Zhang, H. Goodfellow, I. J. Metaxas, D. N. Odena, A. “Self-attention generative adversarial networks”. In: Proceedings of the 36th International

Conference on Machine Learning, 7354–7363, 2019.

[19] T. Luo, G. Jiang, M. Yu, C. Zhong, H. Xu and Z. Pan, "Convolutional neural networks-based stereo image reversible data hiding method", *J. Vis. Commun. Image Representation*, vol. 61, pp. 61-73, 2019.

[20] Yan Wu, Yajun Ma, Jing Liu, Jiang Du, Lei Xing, “Self-attention convolutional neural network for improved MR image reconstruction,” Information Sciences,
Volume 490, 2019.

[21] B. Ou, X. Li, Y. Zhao, R. Ni and Y. -Q. Shi, "Pairwise prediction-error expansion for efficient reversible data hiding", *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5010-5021, Dec. 2013.

[22] J. Deng, W. Dong, R. Socher, L. -J. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 248-255, 2009.

[23] Zhicheng Ni, Yun-Qing Shi, N. Ansari and Wei Su, "Reversible data hiding," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354-362, March 2006

[24] Zhibin Pan, Sen Hu, Xiaoxiao Ma, Lingfei Wang, “Reversible data hiding based on local histogram shifting with multilayer embedding,” Journal of Visual Communication and Image Representation,Volume 31,2015

[25] J. Wang, J. Ni, X. Zhang and Y. -Q. Shi, "Rate and Distortion Optimization for Reversible Data Hiding Using Multiple Histogram Shifting," in *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 315-326, Feb. 2017

[26] Hongchang Zheng, Chuntao Wang, Junxiang Wang, Shijun Xiang, “A new reversible watermarking scheme using the content-adaptive block size for prediction,” Signal Processing,Volume 164, 2019

[27] Ion Caciula, Henri George Coanda, Dinu Coltuc, “Multiple moduli prediction error expansion reversible data hiding,” Signal Processing: Image Communication,Volume 71,2019

[28] S. Kim, X. Qu, V. Sachnev and H. J. Kim, "Skewed Histogram Shifting for Reversible Data Hiding Using a Pair of Extreme Predictions," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 11, pp. 3236-3246, Nov. 2019

[29] W. He and Z. Cai, "An Insight Into Pixel Value Ordering Prediction-Based Prediction-Error Expansion," in *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3859-3871, 2020

[30] He W, Cai Z, Wang Y. "Flexible spatial location-based PVO predictor for high-fidelity reversible data hiding." Information Sciences. 2020 May 1;520:431-44.

[31] G. Mamatha and A. Shaik, "Reversible Data Hiding Based on Integer Wavelet Transforms and Prediction Error Expansion," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 330-335

[32] W. He and Z. Cai, "An Insight Into Pixel Value Ordering Prediction-Based Prediction-Error Expansion," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3859-3871, 2020

[33] W. He and Z. Cai, "Reversible Data Hiding Based on Dual Pairwise Prediction-Error Expansion," in IEEE Transactions on Image Processing, vol. 30, pp. 5045-5055, 2021

[34] R. Hu and S. Xiang, "Reversible Data Hiding By Using CNN Prediction and Adaptive Embedding," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 10196-10208, 1 Dec. 2022.

[35] Kingma, Diederik & Ba, Jimmy. "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations. 2014

[36] D. Ruan, Y. Shi, J. Wen, N. Zheng and M. Zheng, "Spatially-Aware Context Neural Networks," in IEEE Transactions on Image Processing, vol. 30, pp. 6906-6916, 2021

[37] Y. Zhang, J. Jiang, Y. Zha, H. Zhang and S. Zhao, "Research on Embedding

Capacity and Efficiency of Information Hiding Based on Digital Images," International Journal of Intelligence Science, Vol. 3 No. 2, 2013, pp. 77-85

[38] Varghese, F., Sasikala, P. “A Detailed Review Based on Secure Data Transmission Using Cryptography and Steganography.” Wireless Pers Commun 129, 2291–2318 2023

[39] Guo, MH., Xu, TX., Liu, JJ. et al. “Attention mechanisms in computer vision: A survey.” Comp. Visual Media 8, 331–368, 2022

[40] H. Zhao, J. Jia and V. Koltun, "Exploring Self-Attention for Image Recognition," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10073-10082

[41] L. Jiang, M. Zhong and F. Qiu, "Single-Image Super-Resolution based on a Self-Attention Deep Neural Network," 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Chengdu, China, 2020, pp. 387-391

PAPER NAME

**MANAS_KAINTH_THESIS_FINAL.pdf**

WORD COUNT

**11795 Words**

CHARACTER COUNT

**67689 Characters**

PAGE COUNT

**56 Pages**

FILE SIZE

**1.6MB**

SUBMISSION DATE

**May 30, 2023 12:10 PM GMT+5:30**

REPORT DATE

**May 30, 2023 12:11 PM GMT+5:30**

## 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database

## Excluded from Similarity Report

- Submitted Works database
- Bibliographic material
- Cited material
- Small Matches (Less then 10 words)