

An Efficient Video Steganography for Pixel Location Optimization Using Fr-WEWO Algorithm based Deep CNN Model

Shamal Salunkhe*

Ramrao Adik Institute of Technology, Department of Instrumentation Engineering, Navi Mumbai, 400706, India

Email: shamal.salunkhe@rait.ac.in

ORCID iD: <https://orcid.org/0000-0001-5442-0682>

*Corresponding Author

Surendra Bhosale

Veeramata Jijabai Technological Institute, Department of Electrical Engineering, Mumbai, 400031, India

Email: sjbhosale@ee.vjti.ac.in

ORCID iD: <https://orcid.org/0000-0002-5928-0639>

Shubham V. Narkhede

M. S. Data science, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07307, USA

Email: snarkhe1@stevens.edu

ORCID ID: <https://orcid.org/0009-0006-9288-8394>

Received: 29 September, 2022; Revised: 02 November, 2022; Accepted: 18 December, 2022; Published: 08 June, 2023

Abstract: Video steganography is used to conserve the confidential information in various security applications. To give advance protection to the secrete message, pixels locations are optimized using nature inspired algorithm. The input video is separated into a sequence of still image frames then key frames are extracted. The proposed Required Pixel Density (RPD) value calculation and feature extraction are carried out on the extracted frames to perform the frame classification. The frame classification is done using proposed Fractional Water-Earth Worm optimization algorithm based Deep Convolutional Neural Network (FrWEWO-Deep CNN) in order to classify the frames as high, low and medium quality. Thus pixel location prediction is carried out using trained Deep CNN then secret image is hiding within high quality frame with Wavelet Transform (WT) and Inverse WT (IWT). Peak Signal to Noise Ratio (PSNR) and Correlation Coefficient (CC) are performance evaluation parameters. For efficient video steganography better imperceptibility and robustness are required. Imperceptibility is a scale of PSNR value showing similarity between original and stego video frames. The robustness of video steganography is measured by CC between embedded and extracted secret images. The proposed algorithm gives enhanced performance is compared with previous state of art such as WEWO-Deep RNN. The PSNR value is progressed from 41.8492 to 46.5728 dB and CC value improved from 0.9660 to 0.9847.

Index Terms: Deep Convolutional Neural Network, Fractional Calculus, Imperceptibility, Optimization, Wavelet Transform.

1. Introduction

In the recent era, multimedia technology and the internet uses are expanding exponentially. Secure data sharing is basic requirement of many fields such as medical, military or intelligence communication to prevent unauthorised access. Digital videos are the potential tool for information hiding. The video has huge volume which is effectively used to hide secret information. Video Steganography is the process of concealing secret information in the cover image in order to communicate the secret information to receiver. The main aim of this process is to hide the secret information using embedding technique with better imperceptibility and robustness. After the embedding process, the cover image is compared with the embedded image. Both images should be indistinguishable when compared based on statistical features. In the spatial domain video steganography, secret image is hiding directly in cover image. The video frames are transformed into a transform domain, is called as temporal domain video steganography. A stego video is created using frames in a specific sequence along with embedded secret data known as video steganography in raw domain. When compression is performed during creating stego video is known as compressed domain video steganography [1].

Conventional steganography is well thought-out of image structure and distortion function.

Least Significant Bit (LSB), Most Significant Bit (MSB), combination of LSB and MSB substitution, histogram manipulation, Region of Interest (ROI), mapping rule and matrix encoding, Pixel Value Differencing (PVD), are spatial domain techniques [2]. These approaches are not imperceptible and robust. In transform domain video steganography, every video frame is converted into frequency domain by Discrete Cosine Transform (DCT), Discrete Wavelet Transforms (DWT), or other transforms. These approaches have increased embedding capacity, but the clarity of video gets reduced after embedding. In the compressed domain video steganography, video is compressed after embedding so it is resistant to noise [3]. Distortion functions based steganography techniques namely (HUGO), Wavelet Obtained Weights (WOW) and Spatial-UNiversal WAVElet Relative Distortion (S-UNIWARD) are able to embed the secret message in an arbitrary domain with minimum payload [4].

To embed the maximum secret data bits within selected pixels of a key image, bits per pixel (BPP) is proposed by Shumeet Baluja [5]. A bit substitution improves the hiding capacity of the key image without reducing the quality of the embedded image and minimum loss [6]. These concepts motivated us to optimization of pixel location for embedding. Nature inspired algorithms such as Fuzzy logic, Particle Swarm Optimization (PSO), Genetic algorithms (GA), and Neural Networks (NN), are popular approaches used to improve the imperceptibility and payload capacity [7]. Low PSNR value, slow speed and lack of robustness are few of the limitations of conventional steganography techniques could be eliminated using nature inspired algorithms.

The Video steganography with deep learning is more popular in research community, which offers more redundancy and helpful for hiding secret message on video. Presently, the 2D-convolutions with temporal residual modeling were employed to conceal the secret information [8]. Researchers utilized two CNNs for concealing the reference frame within cover frame. VStegNET [9] is one of the efficient models proposed for video steganography. With advancement improvement in deep learning research, the traditional steganography approach provide the steganography as an unsupervised learning task and train deep neural networks that have shown to outperform conventional steganography algorithms to attain a capacity as high as 8 bpp. In [10], the steganography process was carried out using Deep CNN for hiding the secret information with cover information. An auto encoder based deep learning scheme is employed for steganography and the deep learning techniques are trained with the weighted sum of reconstruction loss among the secret as well as revealed secret image and the cover as well as container image. The deep learning technique, HiDDeN method was devised, which was employed for both steganography and watermarking scheme [11].

The major contribution of this research is to design and develop the FrWEWO-Deep CNN for video steganography. The proposed FrWEWO algorithm is designed by incorporating Fractional Calculus (FC) with Water wave optimization (WWO) and Earth worm optimization (EWO). The series of steps followed by the proposed model are key frame extraction, frame classification, pixel prediction, embedding and extraction. At first, the input video is forwarded to the key frame extraction phase in order to convert the video into series of frames, and then the frame classification is done based on the proposed RPD values and the extracted features. Then, the pixel location prediction is done using Deep CNN, where the weight of Deep CNN is trained using proposed FrWEWO algorithm. Once the pixel prediction is done, then the secret image embedding and extraction is carried out using wavelet transform technique in order to retrieve the secret image.

The major research objectives of this paper is given by,

- **Proposed FrWEWO-Deep CNN for pixel prediction in video steganography:** In this research, the proposed FrWEWO algorithm is designed to predict the pixel location for video steganography. It is the optimization algorithm used to train Deep CNN. Proposed RPD values are considered for pixel classification along with feature extraction. Optimization in pixel location selection improved the PSNR and CC value.

This paper is organized as below. Section 2 describes the motivation and literature survey of video steganography based on various techniques, section 3 elaborates the proposed FrWEWO algorithm, the results and discussion are presented in section 4 and section 5 provides the conclusion and future work of this research.

2. Literature Review

Video steganography is the process of concealing secret information with the cover video. The conventional video steganography techniques have several limitations due to the lack the pre-processing stages. The traditional steganography techniques cause numerous constraints in terms of embedding capacity, robustness and imperceptibility against several attacks. These observed problems in the existing video steganography are stimulated the development of proposed FrWEWO, to improve PSNR and CC.

The literature review is based on video steganography techniques along with their upgrading and drawbacks are discussed in this section. The Deep CNN model is proposed with temporal residual modeling for performing the video steganography [12]. The proposed model comprised of five computational processes, such as residual frame labeling, secret image hiding, secret image retrieval, frame classification and the reconstruction of residual frame. Although, this method is effectively impressive, the computational complexity is high. The hybrid adaptive video steganography approach, which is the combination of the embedding probability of cover location as well as the distortion cost function [13]. This method increased the video bit rate after embedding the information. The Multiple Object Tracking (MOT) algorithm and Error Correcting Codes (ECC) for performing video steganography is proposed [14]. The

Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) are employed to decompose the cover image. Here, the secret data was concealed on the coefficients of DCT and DWT region. The proposed method required more time for processing and PSNR could be improved. The Botnet communication approach for video steganography in n Social Network Service (SNS) messenger [15]. The video steganography is carried out using Divide-Embed Combine Method. For embedding, the input and secret video was converted into multiple frames. The final embedding was attained by combining cover frames and stego frames in order to prepare the stego video. The payload allocation mechanism for video steganography based on distortion analysis [16]. The motion vector modification distortion was employed to perform the embedding. The processing time is high. The DWT-based object tracking mechanism for video steganography. 2D-DWT scheme is employed to embed the secret information on video for object tracking. Although, this method had high imperceptibility and robustness against various attacks, the processing time is high [17].

The transform block decision approach for performing video steganography. This method was improved the visual quality of video. Moreover, the embedding mapping rule was utilized to embed the message bits on pixels. In addition, the proposed method attained the large embedding capacity and good visual quality. The proposed method is vulnerable to various kinds of attacks [18]. The content adaptive data hiding algorithm have been widely utilized because of its better security [19]. In the content adaptive data hiding scheme, the particular region is adapted based on the cost function of content-based distortion. For reducing the distortion in embedding process, the optimized coding techniques have been selected to embed the secret information in particular region. The large amount of information is to embed in the complex texture region and no information is embedding in the simple texture area. The complex as well as simple texture area is analyzed using steganalysis. However, the content adaptive embedding scheme has certain limitations as it does not satisfy the Kirchhoff's Principle [20].

During the transmission of secure data, the attackers try to read the information. If the attacker knows the embedding rule, then it is a severe threat to the steganographic algorithm. Another issue of steganographic algorithm is, it mainly depends on the distortion function, and the distortion function is calculated based on the statistical model of cover. However, the statistical modeling of cover is very complex and not possible to execute for some cases. The major aim of steganography process is to conceal the secret message on multimedia cover, and then the steganalysis process determines whether the information is concealed on the multimedia cover. The deep learning techniques provide a better result in video steganography.

Challenges

- The deep CNN [12] was proposed for performing video steganography. However, the proposed method did not provide the optimum result. The challenge lies on including the refined deep learning model for avoiding the failure cases.
- In order to improve the performance of deep learning scheme, the content adaptation scheme was devised in [13]. The performance of embedding was poor in content adaptation scheme. Moreover, the computational complexity of this method was poor.
- For enhancing the computational complexity, Multiple Object Tracking (MOT) algorithm and Error Correcting Codes (ECC) scheme was devised. The challenges of proposed model in [14] were limited to the poor visual quality, security and efficiency. The curvelet transform provides a better solution for this issue.
- In order to improve the efficiency and visual quality of embedded image, Botnet Communication Model was devised in [15]. This method was failed to process with complex environments. However, the proposed method did not provide the better result with various mediums, such as audio, documents and so on.
- In [17], DWT-based object tracking scheme was employed to perform video steganography. This method as failed to process with various level of DWT decomposition. Hence, this method could replace by applying various level of DWT decomposition using various wavelet filters.

3. Methodology of the Proposed Video Steganography

This section describes methodology of the proposed FrWEWO-Deep CNN technique for video steganography using wavelet transform based embedding. The processing steps are key frame extraction, RPD computation, feature extraction, frame classification, pixel prediction, embedding and extraction. Initially, the input video is partitioned into multiple frames based on key frame extraction method using Structural Similarity Index (SSIM). From the extracted frames, the Required Pixel Density (RPD) values and the relevant features are extracted for further effective processing. The RPD values and extracted features are given to the input of proposed FrWEWO-Deep CNN technique, which classifies the frame as either low or high or medium based on the embedding strength with respect to RPD value and extracted features. Then, the pixel prediction is carried out using proposed FrWEWO-Deep CNN algorithm for embedding the secret image. At the end, the hidden image is extracted from the frame using DWT. A block diagram of the proposed FrWEWO-Deep CNN technique for video steganography is shown (refer Fig.1).

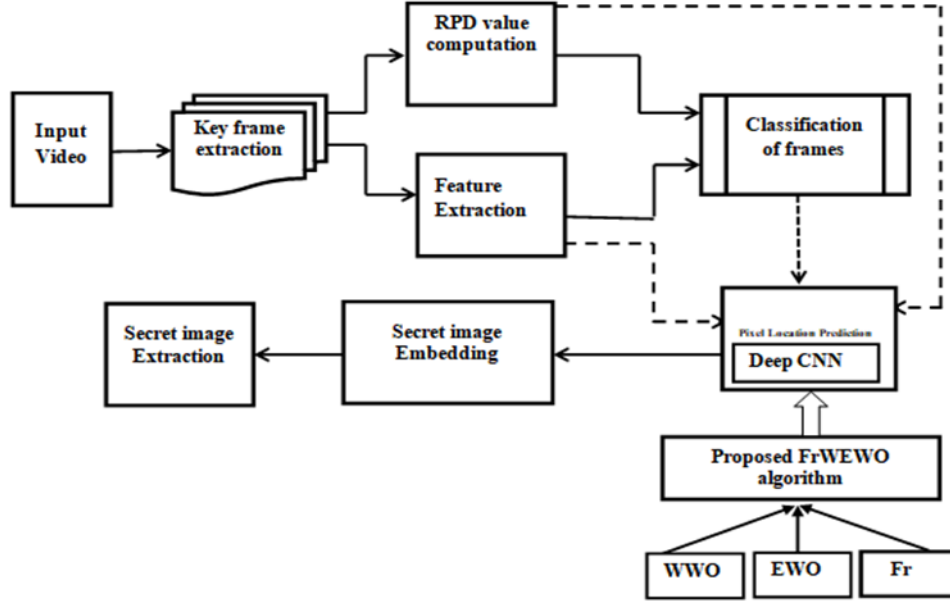


Fig. 1. Block diagram of proposed FrWEWO-based Deep CNN for video steganography.

3.1 Input of video steganography

Proposed FrWEWO-Deep CNN model is proposed to optimize performance of Deep learning model. Input of the steganography is an input video which is selected from dataset. Research contribution aimed to perform efficient embedding. Following discussion is based on video steganography process.

3.1.1. Input Data

Let us assume the database Q contains s number of videos, which is mathematically expressed as,

$$Q = \{P_1, P_2, \dots, P_d, \dots, P_s\} \quad (1)$$

where, s is the total count of videos and P_d is the d^{th} video from dataset. In this research, the d^{th} video is considered as input for key frame extraction phase.

3.1.2. Key frames extraction from video

The extraction of key frame is an essential step to split the input video into multiple frames. Generally, the video refers to the sequence of multiple frames with respect to specific instant of time. The pre-processing using key frame extraction is an important step since it improves the effectiveness of embedding and extraction process. The number of frames produced by the key frame extraction method is based on the time duration of videos since the frame extraction is based on the frame per second. The extracted frames are represented with following expression,

$$Z_x = n \times fps \quad (2)$$

where, Z_x indicates the total count of frames, fps specifies the frames per second and n indicates the duration of video. Here, the key frames are extracted based on DTT with SSIM, which are described as below.

The DTT [21] is an orthogonal transformation technique, which is based on the concept of the concept of Chebyshev polynomials. The DTT of an input video P is indicates as $H(f)$, and it is expressed as,

$$H(f) = \sum_{m=0}^{M-1} t(I, m)P \quad (3)$$

where, $t(I, m)$ indicates the orthogonal basis of discrete Chebyshev polynomial. Likewise, the inverse of DTT is expressed as,

$$P = \sum_{m=0}^{M-1} t(I, m)H(F) \quad (4)$$

After the completion of DTT processing, the similarity between two consecutive frames is computed. Here, the function of computing SSIM is computes the quality of image by identifying the similarity between uncompressed image and reference image. The SSIM makes the value '1' for similar frames and '0' for dissimilar frames. Let us assume the two frames p and q with similar size $J \times J$. SSIM is expressed as,

$$SSIM = \frac{(2\alpha_p\alpha_q + \gamma_1)(2\beta_{p,q} + \gamma_2)}{(\alpha_p^2 + \alpha_q^2 + \gamma_1)(\beta_p^2 + \beta_q^2 + \gamma_2)} \quad (5)$$

where, α_p indicates the average value of p , α_q represents the average value of q , β_p^2 indicates the variance of p , β_q^2 signifies the variance of q , $\beta_{p,q}$ indicates the covariance of p and q , γ_1 and γ_2 are constants. Hence, the extracted key frames from the input video are specified as,

$$P = \{I_1, I_2, \dots, I_d, \dots, I_s\} \quad (6)$$

where, s be the total count of frames and I_d be the d^{th} frame from video. Once the key frames are extracted, then both the RPD computation and feature extraction are performed for each key frame.

1) RPD computation

The RPD computation is proposed to measure the density of pixels exists in the frame. Every key frame is partitioned into a count of blocks with size $s \times a$, and it is mathematically expressed as,

$$RPD = \sum_{z=1}^a \frac{d_z}{y} \quad (7)$$

where, d_z indicates the number of similar pixels in a block, y represents the total count of pixels in a block and a depicts the total count of blocks in a frame. The outcome of RPD computation is illustrated as R .

2) Feature extraction

The advantage of feature extraction method is to improve the effectiveness of embedding and extraction. This process extracts the relevant and effective features, such as wavelet energy, edge feature and LDP feature from key frame. Features are explained as below,

i) Wavelet energy: The wavelet energy [22] is computed by taking the average of entropies of required region. It is expressed as,

$$B = \frac{1}{l} \sum_{f=1}^l \chi(D_f) \quad (8)$$

where, D_f indicates f^{th} region of an image. The wavelet energy feature is indicated as F_1 .

ii) Edge feature: The edge feature [22] is calculated from the edges of blocks. Here, the edge feature is expressed as,

$$G = \frac{1}{l \times C} \sum_{f=1}^l \sum_{g=1}^C G_{f,g} \quad (9)$$

where, $G_{f,g}$ represents the edges of an pixel, C indicates the total count of pixels in f^{th} region, G depicts the edges of block. Moreover, the edge feature is indicated as F_2 .

iii) LDP feature: The feature Local Directional Pattern (LDP) [23] is an enhanced local pattern descriptor. It is incorporating the directional components using Kirsch compass kernels. The advantage of adapting LDP feature is that it poses less susceptible to noise. The LDP feature with pixel (r_a, w_a) is expressed as,

$$LDP(r_a, w_a) = \sum_{j=0}^{e-1} h(i_j - i_a) 2^j \quad (10)$$

where, $LDP(r_a, w_a)$ depicts the LDP feature at mid pixel position. If $r \geq 0$, then the value of h is 1, otherwise the value of h becomes 0.

$$h(r) = \begin{cases} 1 & \text{if } r \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where, e depicts the overall responses of Kirsch masks and i_a indicates the a^{th} maximum Kirsch activation. The LDP feature is indicated as F_3 .

Thus, the outcome of feature extraction process is a final feature vector represented as F_t , which is formed by concatenating the various extracted features, such as wavelet energy, edge feature and LDP feature. It is mathematically expressed as,

$$F_t = \{F_1, F_2, F_3\} \quad (12)$$

where, F_t stands for total feature, F_1 represents the wavelet energy feature, F_2 indicates the edge feature and F_3 depicts the LDP feature. The final feature vector F_t and the outcome of RPD computation R is given to the input of Deep CNN for frame classification. The frame classification for embedding based on RPD value cannot provide the accurate classification outcome. Hence, the RPD value along with effectively extracted features provides an accurate classification outcome using Deep CNN.

3.2. Frame classification using proposed FrWEWO-Deep CNN

This section explains the frame classification using Deep CNN. The weights inputs of Deep CNN are trained using proposed FrWEWO algorithm. The advantages of proposed method are high convergence rate and optimization. The structure of DCNN classifier [24, 25] is signified (refer Fig. 2).

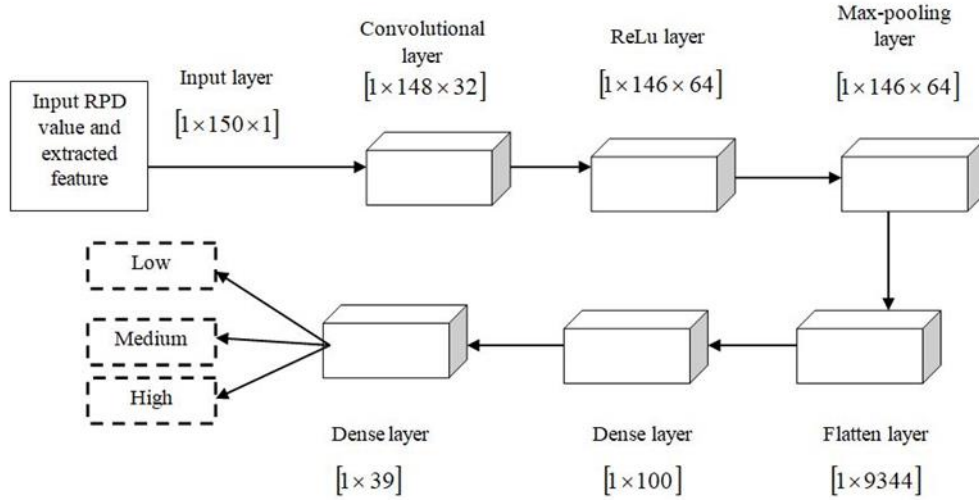


Fig. 2. Design of Deep CNN for video steganography

Moreover, this proposed method attained the better classification results. The Deep Classifier classifies the extracted frame as either low, high or medium quality for embedding. Thus, the extracted feature F_t and RPD value R is given to the input of Deep CNN for further classification. The structural design of Deep CNN classifier is given below.

3.2.1 Structural Design of Deep CNN

The DCNN involves of three layers, like convolution (conv) layer, pooling (pool) layer and Fully Connected (FC) layer in which every layer executes the various operations. The advantage of Conv layer is to generate the feature map depending on the gathered input vector. Once the feature map is formed, then the feature map is sampled, and then forwarded to the pooling layer. Finally, the frame classification is performed in FC layer. The attained accuracy is depending on the Conv layers count in DCNN.

i) *Conv layers:*

The function of conv layer is to gather the suitable pattern of input information based on the conv filters, and connects the gathered information between the preceding neuron layers with the trainable neuron weights. Let the input of DCNN is portrayed as V , and the output acquired from Conv layer is specified as,

$$(M_e^i)_{j,o} = (y_e^i)_{j,o} + \sum_{G=1}^{I_1^{e-1}} \sum_{E=-I_1^e}^{I_1^E} \sum_{N=-I_2^e}^{I_2^e} (f_{m,G}^e)_{E,N} * (M_e^{i-1})_{j+E,O+N} \quad (13)$$

where, $*$ represents the convolutional operator, $(M_e^i)_{j,o}$ states the fixed feature map or the predicted outcome of e^{th} conv layers, which is centered as (j,o) and $(M_e^{i-1})_{j+E,O+N}$, and y_e^i portrays the kernel function in e^{th} conv layers.

ii) *Pooling layers:*

Pooling layer refers to the non-parametric layer without the inclusion of any weights or bias, henceforth the operations are performed effectively.

iii) *FC layers:*

The data acquired from pooling layer is exposed to input of FC layer for establishing the classification of gathered data. Finally, the output acquired through FC layer is expressed as,

$$W_e^i = \mu(F_e^i) \text{ with } F_e^i = \sum_{G=1}^{I_1^{e-1}} \sum_{E=-I_1^e}^{I_1^E} \sum_{N=-I_2^e}^{I_2^e} (f_{m,G}^e)_{E,N} * (M_e^{i-1})_{j+E,O+N} \quad (14)$$

The output acquired through DCNN is represented as R_L^* . Here, the weights of DCNN are optimally tuned with proposed FrWEWO algorithm. The Deep CNN classifier provides the output as low, medium or high quality frames for embedding the secret message.

3.3 The proposed FrWEWO algorithm

The training process of Deep CNN is carried out using proposed optimization algorithm, namely FrWEWO. Here, the proposed FrWEWO algorithm is designed by incorporating WWO [26] and EWO [27] with the FC [28] in order to train the weights and bias of Deep CNN. The EWO algorithm is designed based on the reproduction process of earthworms. Two types of earthworm reproduction are crossover-1 and crossover-2. The crossover -1 creates only one offspring, whereas crossover -2 creates greater than one offspring per iteration. The advantage of EWO is that the exploration and exploitation rate is high. Moreover, this method resolves the discrete constrained optimization issues. The WWO algorithm is designed based on the characteristics of shallow water wave theory. The advantage of WWO algorithm is simple and easy to execute. The proposed algorithm is a modified WEWO-DRNN where fractional calculus is applied on final crossover and mutation equation [29] to get best solution. The combination of crossover and mutation operator is proposed in WEWO algorithm. In addition, the Fractional Calculus (FC) is an analytical model, which helps to resolve the complex computational issues. Hence, the proposed FrWEWO algorithm is designed to take the advantage of FC, EWO and WWO algorithm. The proposed FrWEWO algorithm is given below.

Step 1) Initialization:

The initialization of solution is expressed as,

$$H = \{H_1, H_2, \dots, H_x, \dots, H_n\}; 1 \leq x \leq n \quad (15)$$

where, n depicts the total count of solution and s_x depicts the x^{th} solution

Step2) Evaluation of fitness

The optimal solution is based on the fitness function. Mean Square Error (MSE) a criterion is used to select fitness of each pixel. The fitness with the minimal value of MSE is considered as an optimal solution. The MSE is calculated as,

$$MSE_{rr} = \frac{1}{m,n,l} \sum_{c=1}^{m,n,l} [r_{(m,n,l)} - s_{(m,n,l)}]^2 \quad (16)$$

where, r is the original frame and S is the embedded image, c specifies number of video frames, where $1 < c \leq m, n, l$.

Step3) Weight determination

In this step, the optimal solution is determined using the update equation of proposed FrWEWO algorithm, which is effectively designed by the incorporation of FC concept with the final update function of WEWO algorithm. The WEWO algorithm is constructed by combining EWO and WWO algorithm [29]. The final updated equation of WEWO algorithm is derived as,

$$H_{uv}^{b+1} = \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} \quad (17)$$

The application of FC resolves the complex computational issues of optimization problems. In order to apply FC, let subtract H_{uv}^b on both sides of equation (17), then the expression becomes,

$$H_{uv}^{b+1} - H_{uv}^b = \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} - H_{uv}^b \quad (18)$$

Apply FC in equation (18), then the expression becomes,

$$V^k(H_{uv}^{b+1}) = \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} - H_{uv}^b \quad (19)$$

$$H_{uv}^{b+1} - kH_{uv}^b - \frac{1}{2}k.H_{uv}^{b-1} - \frac{1}{6}(1-k)H_{uv}^{b-2} - \frac{1}{24}k(1-k)(2-k)H_{uv}^{b-3} = \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} - H_{uv}^b \quad (20)$$

$$H_{uv}^{b+1} = kH_{uv}^b - H_{uv}^b + \frac{1}{2}k.H_{uv}^{b-1} + \frac{1}{6}(1-k)H_{uv}^{b-2} + \frac{1}{24}k(1-k)(2-k)H_{uv}^{b-3} + \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} \quad (21)$$

$$H_{uv}^{b+1} = H_{uv}^b(k-1) + \frac{1}{2}k.H_{uv}^{b-1} + \frac{1}{6}(1-k)H_{uv}^{b-2} + \frac{1}{24}k(1-k)(2-k)H_{uv}^{b-3} + \frac{H_{\max,v} + H_{\min,v}}{1 + H_{uv}^b} \quad (22)$$

where, $H_{\max,v}$ and $H_{\min,v}$ are the upper as well as lower bound location of earth worm, H_{uv}^b be the present location of water wave at time b and H_{uv}^{b+1} be the present location of water wave at time $b+1$. Equation (22) represents the final updated equation of proposed FrWEWO algorithm, which optimally selects the exact pixel solution such that the performance of system is enhanced. The output acquired from the proposed FrWEWO algorithm is used to train the weights of Deep CNN for classifying the frames in order to detect the best embedding pixel location from the frame.

Step 4) Re-evaluation of fitness:

The optimal solution is re-evaluated based on the fitness function given in equation (16), in which the minimal value of MSE is considered as an optimal solution.

Step 5) Terminate:

The optimum weights are evaluated in a repetitive manner till the highest iteration W_{mzx} is reached. Thus optimization of pixel location is performed using proposed algorithm.

The pseudo code of proposed Fr-WEWO algorithm is given bellow,

The optimal region for embedding is selected using proposed FrWEWO algorithm, which is designed by the integration of EWO, WWO and FC algorithm. The effectiveness of secret image embedding is increased using proposed algorithm.

The Proposed FrWEWO algorithm
Input: Population H , iteration counter w , RPD, Final Feature Vector F_i
Output: H_{uv}^{b+1} optimal solution
Initiate
Initialization of solution $H = \{H_1, H_2, \dots, H_x, \dots, H_n\}; 1 \leq x \leq n$
Calculate the error using equation (16)
Arrange every earthworm as per the fitness value, $H = \{H_{Best}, \dots, H_{worse}\}; 1 \leq x \leq n$
Check Termination Criteria
While H_{uv}^{b+1} not found or
($w < w_{max}$) do
Randomly choose an earthworm for crossover
Compute H_{uv}^{b+1} using equation (22)
Calculate the fitness value of offspring H_{uv}^{b+1} using equation (16)
If $H_{uv}^{b+1} > H_{Best}$ Then replace with H_{Best}
Else if $H_{uv}^{b+1} > H_{worse}$
Then replace with H_{worse}
Else discard
Evaluate population based on newly updated position using (16)
$w = w + 1$
Return optimal solution H_{uv}^{b+1}
End

3.4. Pixel prediction using Deep CNN

The pixel prediction is utilized to select the optimal pixel for obscuring the secret image on the input frame. Here, the pixel prediction is carried out using Deep CNN for embedding the secret image on the frames acquired from video. For predicting the optimal pixel, the suitable regions are selected based on the estimated RPD values, extracted features F_i and the frame classification outcome L_e , H_e and M_e . The advantage of Deep CNN is that it provides the better classified outcome. The Deep CNN classifier effectively predicts the optimal pixel for embedding based on the RPD value, extracted feature and the frame classification outcome. The output of pixel prediction process is represented as H_x in which the selected pixels are utilized for secret image embedding. Moreover, the structural design of Deep CNN is described in previous section C.

3.5. Wavelet transform-based secret image embedding

The video steganography is the process of concealing a secret image on the selected region. After predicting the appropriate pixel, the embedding process is carried out by inserting a secret image on the predicted region of pixel using wavelet transform. The wavelet transform is utilized for embedding as well as extraction of secret image on the frame. High processing speed is main advantage of wavelet transform.

WT: The wavelet coefficient is determined to embed and extract the image. While executing the embedding process, the image is subdivided into eight sub-bands with two levels of decompositions. In the first level of decomposition, the frame is partitioned into four sub-bands, like LL , HL , HH and LH . Here, LL signifies the coefficients at coarse level, HL , HH and LH designates the finest scale wavelet coefficient. Here, the LL sub band is selected to perform second level of decomposition for embedding. The partitioned sub-bands are expressed as,

$$\{LL^*, LH^*, HL^*, HH^*\} = X(E_a) \quad (23)$$

where, E_a specifies the key frame, X represents the wavelet transform, and $\{LL, LH, HL, HH\}$ denotes the wavelet coefficients. The second level decomposition is carried out on all the bands of image E_a , thereby the 16 sub-bands are attained and is expressed as,

$$\{LL_1^*, LH_1^*, HL_1^*, HH_1^*\} = X(LL^*) \quad (24)$$

$$\{LL_2^*, LH_2^*, HL_2^*, HH_2^*\} = X(LH^*) \quad (25)$$

$$\{LL_3^*, LH_3^*, HL_3^*, HH_3^*\} = X(HL^*) \quad (26)$$

$$\{LL_4^*, LH_4^*, HL_4^*, HH_4^*\} = X(HH^*) \quad (27)$$

Thus, LL^* and HH^* coefficients are selected to perform the embedding and extraction process. Consequently, the embedding is carried out on the in selected region such that the embedding function is articulated as,

$$E_f = K_f + S * A_f \quad (28)$$

where, E_f represents the embedded image, K_f indicates the key frames, S denotes the embedding strength and A_f demonstrates the secret image. Here, the embedding strength S is determined based on the condition of required pixel density. The various conditions of pixel density frames (PDF) are given below.

- **Condition 1:** If the PDF is high, then the embedding strength of particular frame is also high ($S = 0.5$). Hence, more embedding changes can be done in the particular frame.
- **Condition 2:** If the PDF is medium, then the embedding strength of particular frame is also medium ($S = 0.25$). Therefore, the medium level embedding changes can be done in the particular frame.
- **Condition 3:** If the PDF is low, then the corresponding frame has low embedding strength ($S = 0.1$). Consequently, the low level embedding changes can be done in the respective frame.

IWT: After the embedding, the IWT is applied to the embedded image in order to extract the hidden data. Thus, the embedded image is indicated as, $E_f = \{LL^{**}_1, LH^{**}_1, HL^{**}_1, HH^{**}_1\}$. Moreover, the IWT is stated as,

$$IWT(E_f) = \{LL^{**}, LH^{**}, HL^{**}, HH^{**}\} \quad (29)$$

where, IWT designates the inverse wavelet transform. Accordingly, second level of decomposition is stated as,

$$M^* = IWT(E_f) \quad (30)$$

where, M^* designates the embedded image.

3.6. Extraction of secret message

Once the embedding process is done, then the embedded image is sent to the destination side over the network. In order to extract the hidden image from embedded image, the receiver requires the key frames for extraction. Hence, the WT is employed to retrieve wavelet coefficients, thereby the receiver extracts the secret image at the receive side. Moreover, the initial level decomposition is given by,

$$\{LL^{**}, LH^{**}, HL^{**}, HH^{**}\} = X(M^*) \quad (31)$$

where, M^* represents the embedded image. Moreover, the second level of decomposition is carried out on each sub-bands, which is expressed as,

$$\{LL_{a1}^*, LH_{a1}^*, HL_{a1}^*, HH_{a1}^*\} = X(LL_a^*) \quad (32)$$

$$\{LL_{a2}^*, LH_{a2}^*, HL_{a2}^*, HH_{a2}^*\} = X(LH_a^*) \quad (33)$$

$$\{LL_{a3}^*, LH_{a3}^*, HL_{a3}^*, HH_{a3}^*\} = X(HL_a^*) \quad (34)$$

$$\{LL_{a4}^*, LH_{a4}^*, HL_{a4}^*, HH_{a4}^*\} = X(HH_a^*) \quad (35)$$

At last, the embedded image comprised of wavelet a sub band, which is stated as $F^* = \{LL_{a1}^*, HH_{a1}^*, \dots, LL_{a4}^*, HH_{a4}^*\}$ and it is used to mine the secret image. Henceforth, the extraction process is expressed as,

$$A = K_f^* - K_f \quad (36)$$

where, A signifies the retrieved message. After the application of WT,

$$K = O^* - O \quad (37)$$

where, O^* specifies band hidden by secret image, O designates the pixel position, and the secret message is recovered as,

$$N = \begin{cases} 1; & \text{If } K > \delta \text{ \& } O > \delta \\ 0; & \text{Else} \end{cases} \quad (38)$$

Here, the pixel position O is anticipated based on Deep CNN, which is used to estimate the position of image pixels for embedding the secret message

4. Simulation and Results

This section describes the results and discussion of proposed FrWEWO-Deep CNN for video steganography based on evaluation metrics, such as Correlation Coefficient and PSNR. The assessment is done with the number of frames with and without the salt and pepper noise. The experimental setup, dataset description, performance metrics, comparative techniques, comparative analysis, and discussion on result are presented in this section.

4.1. Experimental setup

The experimentation of proposed Fr-WEWO-Deep CNN method is done with Pycharm tool with Windows 10 OS, 4GB RAM and Intel i3 core processor.

4.2. Dataset description

The dataset employed for the proposed FrWEWO-Deep CNN method for video steganography is CAVIAR database [30]. The video clips were filmed for the CAVIAR database with a wide angle camera lens in the entrance lobby of the INRIA Labs at Grenoble, France. The resolution is half-resolution PAL standard (384 x 288 pixels, 25 frames per second) and compressed using MPEG2. The file sizes are mostly between 6 and 12 MB, a few up to 21 MB. CAVIAR video database is a set of small video clips recorded under various acts. In this work two video clips from the CAVIAR database are used for the comparative analysis.

4.3. Performance metrics

The effectiveness of proposed FrWEWO-DeepCNN using correlation coefficient and Peak Signal to noise ratio (PSNR) are described. The analysis is performed without noise and then salt-pepper noise is added for computing the performance of the proposed algorithm.

4.3.1. Imperceptibility Analysis:

Image noise is an arbitrary variation in the brightness or colour of the images. The impulse noise is one of the external noise-induced deterioration of the image. Noise corrupts images by replacing some of the original image's pixels with new pixels having different brightness values. Impulse noise causes pixel deterioration at random locations. Impulse noise-corrupted pixels have intensity levels that are either very high or very low. They typically contrast sharply with the smooth, pristine, unblemished surroundings. A noisy image to be modelled as follows: $g(x, y) = f(x, y) + \eta(x, y)$, where $f(x, y)$ is the original image pixel, $\eta(x, y)$ is the noise term and $g(x, y)$ is the resulting noisy pixel. The fixed-valued impulse noise is called as salt-and pepper noise. Random bright pixels and random dark pixels are added to an image to add salt and pepper noise. Data drop noise is another name for it since it statistically reduces the original data values. For grey scale image, the gray scale values of noisy pixels distorted by random-valued impulse noise are equally distributed in the range [0,255]. If the amount of bits for transmission is 8, pixels are replaced with corrupted pixels with maximum or minimum values, i.e., 255 or 0, respectively.

The common assessment used to measure the imperceptibility of video steganography are Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM)

4.3.2. PSNR:

The proportion of maximum signal power and the noise which affect the fidelity is computed using PSNR and is given by equation 39,

$$PSNR = \frac{1}{2} [PSNR[A(g, h), T(g, h)] + PSNR[B(g, h), T(g, h)]] \quad (39)$$

where, $A(g,h)$ and $B(g,h)$ indicate two input images and $T(g,h)$ represent the embedded image.

4.3.3. Correlation coefficient:

Correlation coefficients are a widely-used as a performance evaluation measure in image processing. The correlation between two images is measured using correlation coefficients. Although there are other kinds of correlation coefficients, Pearson's correlation coefficient is the most common. The correlation coefficient known as Pearson's is frequently applied in linear regression. With a correlation value of 1, there is a fixed proportional increase in one variable for every positive increase in the other. A correlation coefficient of -1 denotes a fixed proportional negative decline in the other variable for every positive gain in the first. If the value is zero, the two photos are simply unrelated. Equation 40 provides the correlation coefficient as shown below,

$$\gamma = \left(\frac{\sum (xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum (xi - \bar{x})^2 \sum (yi - \bar{y})^2}} \right) \tag{40}$$

The statistical relation between two variables is given as correlation coefficient and it ranges between -1 to +1 where ± 1 represent strongest possible agreement and 0 indicate disagreement between two images.







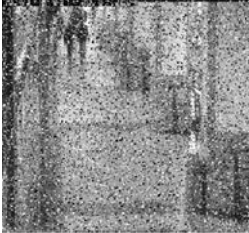

4.4. Comparative techniques

The performance of proposed FrWEWO-Deep CNN technique is analyzed using various existing techniques, such as DWT+DCT [6], DCNN+DWT [31], DCNN+LSB [12], proposed WEWO-DeepRNN [29] and 3D-CNN [9].

4.5. Comparative assessment

The comparative assessment of various techniques is done based on evaluation metrics, Correlation coefficient and PSNR. The assessment of proposed technique is based on frames with salt and pepper noise and frames without noise. The videos are mentioned as, video case-1 and video case-2, named view of The mall shop areas and Walk by shop 1 corridor respectively. The Lena (512 x 512) and Peppers (256 x 256) are secret images used for embedding. The simulation results of key frames are shown in Table 1.

Table 1. Video case-1 and video case-2 key frames, stego frames with secret images.

Video	Without Noise Key frame	Secret image	With noise stego frame	Extracted image
Video Case-1				
Video Case-2				

In Table 1, one key frame from every video case is selected to visualize the embedding and extraction outcomes. The key frame no. 7 from video case 1 and key frame no. 10 from video case 2 are used. The key frames without noise and with salt and pepper noise stego frames are presented. Original images and extracted images are also presented. The frame wise PSNR and CC variations are discussed in next part of the paper.

4.5.1. Comparative assessment based on frames with salt and pepper noise using video case-1

The comparative assessment of proposed FrWEWO-Deep CNN technique based on frame with salt and pepper noise using video case-1 is shown (refer Fig. 3). Frame 10, 1, 13, 7, and 19 are high, medium and low quality frames. The comparative assessment of proposed technique with respect to correlation coefficient is indicated (refer Fig. 3a). When salt-pepper noise is 0.05, the corresponding CC values evaluated by DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 0.5598, 0.7439, 0.5738, 0.7819, 0.8694 and 0.8762. When amount of salt-pepper noise is 0.250, the CC values obtained for DWT+DCT, DCNN + DWT,

DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 0.5075, 0.69361, 0.5286, 0.7209, 0.8453 and 0.8455. The comparative assessment of proposed technique in terms of PSNR for video 1 are shown (refer Fig.3b). When salt-pepper noise is 0.05, the evaluated PSNR values by existing DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 31.48 dB, 38.22 dB, 31.86 dB, 39.83 dB, 41.58 dB and 43.87 dB. When salt-pepper noise is 0.250, the PSNR values for DWT+DCT, DCNN+DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 27.07dB, 32.58dB, 26.42dB, 34.13dB, 39.67 dB and 41.54 dB.

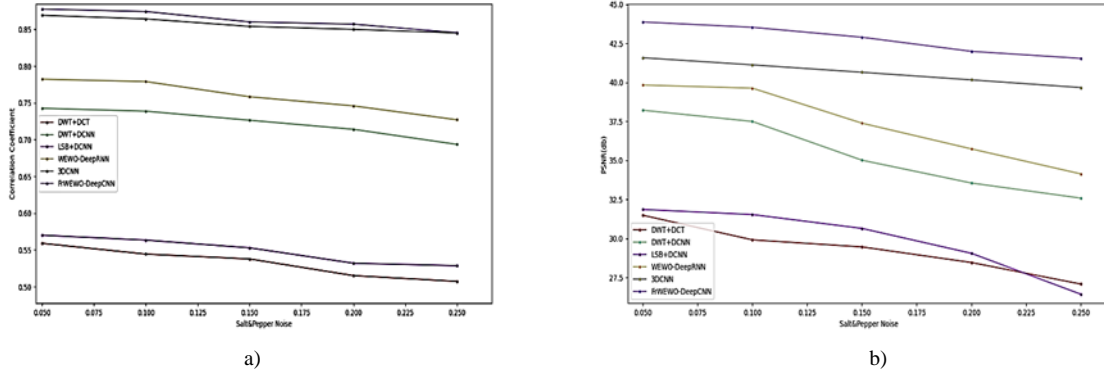


Fig. 3. Comparative assessment by adjusting frames with noise using a) Correlation coefficient, b) PSNR

4.5.2. Comparative assessment based on frames without noise using video-1

The comparative assessment of proposed FrWEWO-Deep CNN technique by adjusting the frame without noise using video-1 is demonstrated (refer Fig. 4). The assessment of proposed FrWEWO-Deep CNN technique in terms of correlation coefficient is given (refer Fig. 4a). The correlation coefficient of proposed model is 0.9975, whereas the existing techniques, such as DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 0.7172, 0.9674, 0.9348, 0.9741 and 0.9822, correspondingly for frame 10. The correlation coefficient of proposed model is 0.9713, whereas the existing techniques, such as DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 0.7014, 0.9526, 0.9159, 0.9602 and 0.9624, correspondingly for frame 19.

The assessment based on PSNR value is illustrated for video 1 (refer Fig. 4b). When DWT+DCT, DCNN + DWT, DCNN+LSB and proposed WEWO-DeepRNN, 3D-CNN are achieved the PSNR values of 38.89 dB, 43.92 dB, 41.24 dB, 43.95 dB and 45.00 dB, whereas the proposed technique achieved the PSNR values of 47.89 dB for frame 10. The PSNR of proposed model is 45.29 dB, whereas the existing techniques, such as DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 36.57 dB, 39.12 dB, 37.21 dB, 40.64 dB, correspondingly for frame 19.

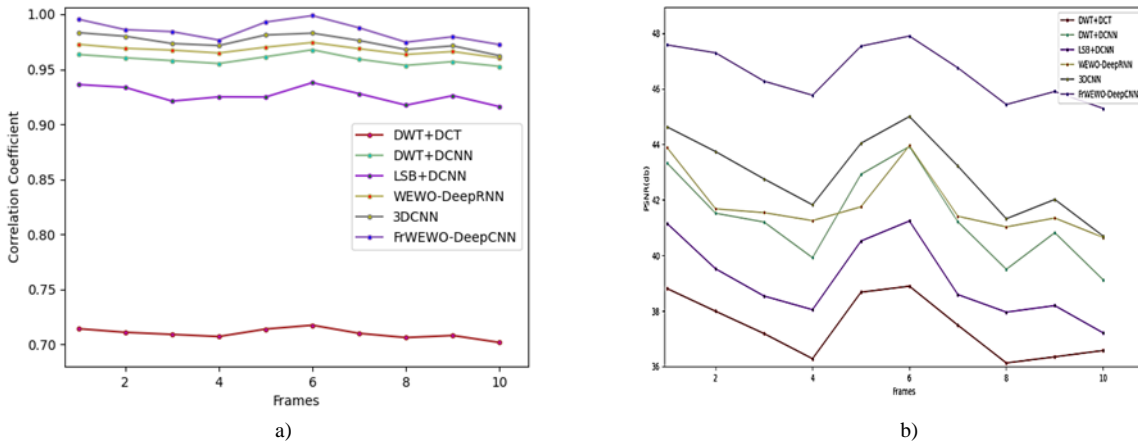


Fig. 4. Comparative assessment by adjusting frames without noise using a) Correlation coefficient, b) PSNR

4.5.3. Comparative assessment based on frames with salt and pepper noise using video-2

The comparative assessment of proposed FrWEWO-Deep CNN method based on frame with salt and pepper noise using video-2 is shown (refer Fig. 5). The comparative assessment of proposed technique with respect to correlation

coefficient is illustrated (refer Fig. 5a). When salt-pepper noise is 0.05, the corresponding CC values evaluated by DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 0.5318, 0.7613, 0.5830, 0.7839, 0.8560 and 0.8697. When amount of salt-pepper noise is 0.250, the CC values obtained for DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 0.5059, 0.7356, 0.5436, 0.7422, 0.8351 and 0.8487.

The comparative assessment of proposed technique in terms of PSNR for video 2 is shown (refer Fig. 5b). When salt-pepper noise is 0.05, the evaluated PSNR values by existing DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 31.51 dB, 38.74 dB, 32.41 dB, 39.12 dB, 41.37 dB and 43.08 dB. When salt-pepper noise is 0.250, the PSNR values for DWT+DCT, DCNN+DWT, DCNN+LSB, proposed WEWO-DeepRNN, 3D-CNN and proposed FrWEWO-Deep CNN are 26.34 dB, 35.02 dB, 27.56 dB, 35.57 dB, 39.43 dB and 40.88 dB.

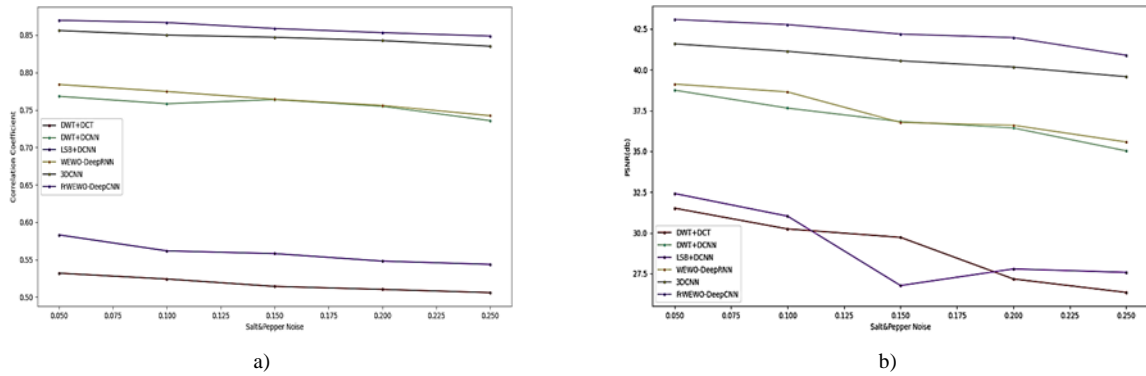


Fig. 5. Comparative assessment by adjusting frames with noise using a) Correlation coefficient, b) PSNR

4.5.4. Comparative assessment based on frames without noise using video-2

The comparative assessment of proposed FrWEWO-Deep CNN technique based on frame without noise using video-2 is illustrated (refer Fig. 6). The assessment of proposed FrWEWO-Deep CNN technique in terms of correlation coefficient is shown (refer Fig. 6a). The correlation coefficient of proposed model is 0.9871, whereas the existing techniques, such as DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 0.7115, 0.9717, 0.9339, 0.9741 and 0.9714, correspondingly for frame 3. The correlation coefficient of proposed model is 0.97, whereas the existing techniques, such as DWT+DCT, DCNN + DWT, DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 0.7040, 0.9513, 0.9246, 0.9608 and 0.961, correspondingly for frame 10.

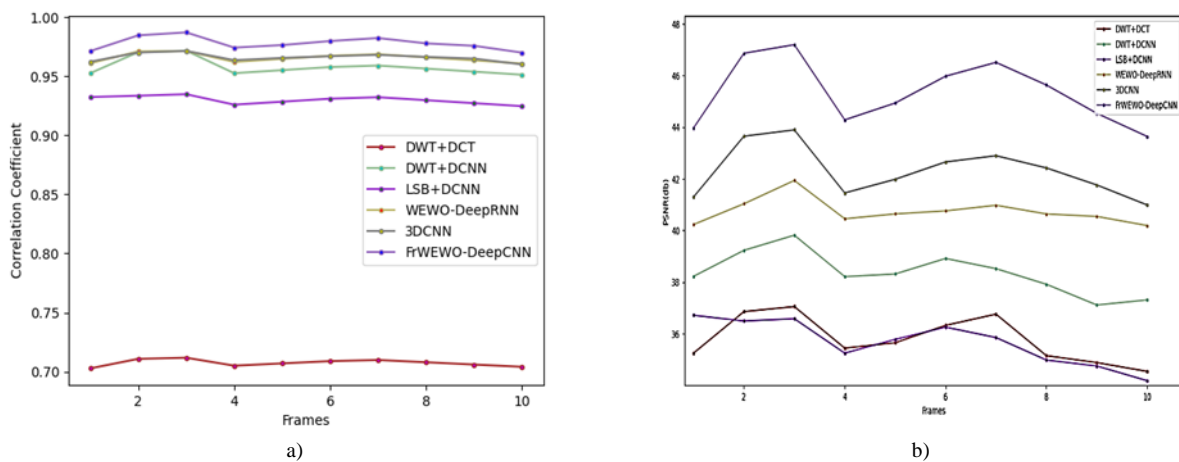


Fig. 6. Comparative assessment by adjusting frames without noise using a) Correlation coefficient, b) PSNR

The assessment based on PSNR value for video 2 is given (refer Fig. 6b). When DWT+DCT, DCNN + DWT, DCNN+LSB and proposed WEWO-DeepRNN, 3DCNN are achieved the PSNR values of 37.05 dB, 39.82 dB, 36.58 dB, 41.94 dB and 43.89 dB, whereas the proposed technique achieved the PSNR values of 47.19 dB for frame 3. The PSNR of proposed model is 43.64 dB, whereas the existing techniques, such as DWT+DCT, DCNN + DWT,

DCNN+LSB, proposed WEWO-DeepRNN and 3D-CNN are attained the correlation coefficient of 34.52 dB, 37.31 dB, 40.19 dB, 40.98 dB, correspondingly for frame 10.

4.6. Comparative discussion

The comparison of proposed model based on evaluation metrics in terms of correlation coefficient and PSNR is disused here. In this paper, the performance of proposed technique is analyzed based on presence the salt and pepper noise and without noise with video- case 1 and video case 2 in terms of evaluation metrics. Average values of correlation coefficient and PSNR are computed from frame 10, 1, 13, 7, and 19 in video case-1. Average values of correlation coefficient and PSNR are computed from frame 3, 10, 7, 15 and 25 in video case-2. Table 2 is showing comparative analysis for video case-1 and video case-2.

Table 2. Comparative analysis for video case-1 and video case-2

Video Input	Variations	Metrics	DWT+ DCT	DWT+ DCNN	LSB+ DCNN	Proposed WEWO-DeepRNN	3DCNN	Proposed FrWEWO-DeepCNN
Video case-1		CC	0.7096	0.9584	0.9264	0.966	0.9748	0.9847
	Without noise	PSNR	37.4399	41.347	39.0976	41.8492	42.9234	46.5728
	Salt and	CC	0.5327	0.7231	0.5494	0.7585	0.8565	0.8629
	Pepper noise	PSNR	29.2785	35.3763	29.9028	37.3491	40.6429	42.7624
Video case-2		CC	0.7073	0.9581	0.9301	0.9656	0.9659	0.9775
	Without noise	PSNR	33.7943	38.3608	35.6876	40.7439	42.3031	45.3576
	Salt and	CC	0.5172	0.7571	0.5588	0.7642	0.8462	0.8594
	Pepper noise	PSNR	29.2785	35.3763	29.9028	37.3491	40.6429	42.7624

For video- case 1, the correlation coefficient and PSNR values attained by the proposed model are 0.9748 and 46.5728 dB, whereas the existing techniques acquired the correlation coefficients 0.7096, 0.9584, 0.9264, 0.9660 and 0.9748, and the PSNR values 37.4399 dB, 41.3470 dB, 39.0976 dB, 41.8492 dB and 42.9234 dB, respectively. When salt and pepper noise is added to the frames, the correlation coefficient and PSNR values attained by the proposed model are 0.8629 and 42.7624 dB, whereas the existing techniques acquired the correlation coefficients 0.5327, 0.7231, 0.5494, 0.7585 and 0.8565, and the PSNR values 29.2785dB, 35.3763dB, 29.9028 dB, 37.3491 dB and 40.6429dB, respectively.

For video- case 2, the correlation coefficient and PSNR values attained by the proposed model are 0.9775 and 45.3576 dB, whereas the existing techniques acquired the correlation coefficients 0.7073, 0.9581, 0.9301, 0.9656 and 0.9659, and the PSNR values 33.7943 dB, 38.3608 dB, 35.6876 dB, 40.7439 dB and 42.3031 dB, respectively. When salt and pepper noise is added to the frames, the correlation coefficient and PSNR values attained by the proposed model are 0.8594 and 42.1783 dB, whereas the existing techniques acquired the correlation coefficients 0.5172, 0.7571, 0.5588 and 0.8462, and the PSNR values 28.9465 dB, 36.9339 dB, 29.1143 dB, 37.3395 dB and 40.6014 dB, respectively.

More key frames for training developed the more reliable DCNN model. In the research frames are classified based on optimum region (pixel location). Better values of PSNR and CC are presenting reliable and accurate results as compared to existing methods of video steganography.

5. Conclusion and Future Work

5.1. Conclusion

The proposed FrWEWO-Deep CNN technique for video steganography is presented in this paper. The pixel locations are predicted to facilitate embedding process of steganography. Frame classification is also carried out using a deep learning classifier DCNN which indicated best key frame for secrete image embedding. The classification of frames is based on RPD values and extracted features. The proposed concept of RPD value is attained to find out block wise similar pixels of every frame. The secret image is embedded and retrieved from the cover video using wavelet transform technique. Embedding at optimum pixel location is improving the efficiency of video steganography. The experimental results demonstrates that the proposed methodology attained the better performance based on CC and PSNR. Here, the proposed FrWEWO-Deep CNN algorithm is showing better imperceptibility. It also resolves the computational and analytical issues, which reduced the error produced thus, the PSNR value improved. The pixel location selection and prediction improved the correlation coefficient value. All the techniques are performing better without noise added in the key frames. But when salt and pepper noise is added in the frame, PSNR value and correlation coefficient got affected.

5.2. Limitations

This paper focused on the optimization using evolutionary algorithm and training of Deep CNN model for video steganography. A still image video steganography technique is analyzed with performance index such as PSNR and CC. The video steganography in the uncompressed domain has a good hiding capacity but not strongly robust against signal processing, noise, and attacks by an intruders. So, we recommend investigating a compressed domain video steganography to enhance the robustness and imperceptibility.

5.3. Future Work

In future, the effectiveness of embedding will be enhanced by applying an effective optimization technique with macro blocks (compressed domain video steganography). Payload capacity and execution time will also consider for performance evaluation. More number of videos will be analyzed for more precise results.

References

- [1] Rachna Patel, Kalpesh Lad and Mukesh Patel, "Study and investigation of video steganography over uncompressed and compressed domain: a comprehensive review", *Multimedia Systems*, 27:985–1024, 2021, pp. 986-1024.
- [2] Wang Xiang-yang, Wang Chun-peng, Yang Hong-ying, Niu Pan-pan, "A robust blind color image watermarking in quaternion Fourier transform domain", *Journal of Systems and Software*, 2013, vol. 86, no. 2, pp. 255-277.
- [3] Rohit S. Malladar and Sanjeev R. Kunte, "Selective Video Encryption Using the Cross Coupling of One-dimensional Logistic Maps", *I. J. Computer Network and Information Security*, 2021, 5, pp. 40-54
- [4] Jan Butora and Jessica Fridrich, "Effect of JPEG Quality on Steganographic Security, IHMMSec-19, 2019, TROYES, France Association for Computing Machinery, ACM ISBN 978-1-4503-6821-6/19/06.
- [5] Shumeet Baluja, "Hiding Images in Plain Sight: Deep Steganography, In *Advances in Neural Information Processing Systems*", 2017, pp. 2069-2079.
- [6] Ramadhan J. Mstafa, Khaled M. Elleithy and Eman Abdelfattah, "Video Steganography Techniques: Taxonomy, Challenges, and Future Directions", In *proceedings of IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, IEEE, 2017, pp. 1-6.
- [7] Yunxia Liu, Shuyang Liu, Yonghao Wang, Hongguo Zhao, and Si Liu, "Video Steganography: A Review, *Neuro computing*", 2019, vol. 335, pp. 238-250.
- [8] Zhili Zhou Ching, Nung Yang Cheonshik and Kim Stelvio Cimat, "Introduction to the special issue on deep learning for real-time information hiding and forensics", *Journal of Real-Time Image Processing*, 2020, 17:1–5.
- [9] Aayush M., Suraj Kumar, Aditya Nigam and Saiful Islam, "VStegNET: Video Steganography Network using Spatio-Temporal features and Micro-Bottleneck", *30th British Machine Vision Conference, BMVC 2019, Cardiff, UK, 2019*, pp. 274-286.
- [10] Nandhini Subramanian, Ismahane Cheheb, Omar Elharrouss, Somaya Al-Maadeed and Ahmed Bouridane, "End-to-End Image Steganography Using Deep Convolutional Autoencoders", *IEEE Access*, 2021, Volume 9, pp. 135585-135592
- [11] Yao, Y. and Yu, N., "Motion vector modification distortion analysis-based payload allocation for video steganography", *Journal of Visual Communication and Image Representation*, 2021, vol.74.
- [12] Xinyu Weng, Yongzhi Li, Lu Chi, and Yadong Mu, "Convolutional Video Steganography with Temporal Residual Modeling, *Computer Science*", *Proceedings of the International Conference on Multimedia Retrieval*, 2019.
- [13] Ke. Niu, Jun. Li, Xiao Yuan Yang, Shuo Zhang and Bo Wang, "Hybrid Adaptive Video Steganography Scheme under Game Model", *IEEE Access*, 2019, vol.7, pp. 61523-61533.
- [14] Ramadhan J. Mstafa, Khaled M. Elleithy, and Eman Abdelfattah, "A Robust and Secure Video Steganography Method in DWT-DCT Domains Based on Multiple Object Tracking and ECC", *IEEE Access*, 2017, vol. 5, pp. 5354-5365.
- [15] Kwak, M. and Cho, Y., "A novel video steganography-based botnet communication model in telegram sns messenger", *Symmetry*, 2021, vol.13, no.1, pp.84.
- [16] Yao, Y. and Yu, N., "Motion vector modification distortion analysis-based payload allocation for video steganography", *Journal of Visual Communication and Image Representation*, 2021, vol.74.
- [17] Dalal, M. and Juneja, M., "A secure video steganography scheme using DWT based on object tracking", *Information Security Journal: A Global Perspective*, 2021, pp.1-18.
- [18] Zhao, H., Liu, Y., Wang, Y., Liu, S. and Feng, C., "A Video Steganography Method Based on Transform Block Decision for H. 265/HEVC", *IEEE Access*, 2021, vol.9, pp.55506-55521.
- [19] Zhang Y., Ye, D., Gan, J., Li, Z. and Cheng, Q., "An image steganography algorithm based on quantization index modulation resisting scaling attacks and statistical detection", *Comput. Mater. Continua*, 2018, vol.56, no.1, pp.151-167.
- [20] Mrdovic, S. and Perunicic, B., "Kerckhoffs' principle for intrusion detection. In *Networks 2008*", *The 13th International Telecommunications Network Strategy and Planning Symposium*, 2008, pp. 1-8.
- [21] Rubina Khongsit and P. Rangababu, "Scalable Discrete Tchebichef Transform for Image/Video Compression," *International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 2017.
- [22] Suresh M. and Sam I. S., "Exponential fractional cat swarm optimization for video steganography", *Multimedia Tools and Applications*, 2021, vol.80, no.9, pp.13253-13270.
- [23] Chakraborti T., Mc Cane, B., Mills S. and Pal, U., "Loop descriptor: Local optimal-oriented pattern", *IEEE Signal Processing Letters*, 2018, vol. 25, no. 5, pp. 635-639.
- [24] Sugave, S. and Jagdale, B., "Monarch-EWA: Monarch-Earthworm-Based Secure Routing Protocol in IoT", *The Computer Journal*, 2020, vol.63, no.6, pp.817-831.
- [25] Liu, M., Shi, J., Li, Z., Li, C., Zhu, J. and Liu, S., "Towards better analysis of deep convolutional neural networks," *IEEE transactions on visualization and computer graphics*, 2016, vol.23, no.1, pp.91-100.
- [26] Yu-Jun Zheng, "Water wave optimization: A new nature-inspired meta heuristic", *Computers and Operations Research*, 2015, vol.55, pp.1-11.
- [27] Gai-Ge Wang, Suash Deb, and Leandro dos Santos Coelho, "Earthworm optimization algorithm: a bio-inspired meta heuristic algorithm for global optimization problems", *IJBIC*, vol.12, no.1, pp.1-22, 2018.
- [28] Bhaladhare, P.R. and Jinwala, D.C., "A clustering approach for the-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm," *Advances in Computer Engineering*, 2014.
- [29] Shamal Salunkhe, Surendra Bhosale, "Nature Inspired Algorithm for Pixel Location Optimization in Video Steganography Using Deep RNN", *International Journal on Engineering, Science and Technology*, Volume 3, Issue 2, 2021, ISSN: 2642-4088, pp. 146-154.
- [30] CAVIAR database, "<https://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>"

- [31] Subodh Ingaleshwar, Nagraj Dharwadkar, "Water chaotic fruit fly optimization based deep convolutional neural network for image watermarking using wavelet transform", Multimedia Tools Appl. (2020).

Authors' Profiles



Shamal Salunkhe received the B. E. degree in Instrumentation and Control Engineering in 2004 from the Pune University, Pune, India and received the M. E. degree in Instrumentation and Control Engineering in 2012 from the University of Mumbai, India. She is currently pursuing the Ph. D. degree in Electrical engineering from Veermata Jijabai Technological Institute, Mumbai, India. Presently, she is working as Assistant Professor in the Department of Instrumentation Engineering, Ramrao Adik Institute of Technology, Navi Mumbai. Her research interests are deep learning, artificial intelligence, signals processing and cyber security.



Surendra Bhosale received Bachelor's degree in Electrical Engineering in 1987 from Shivaji University, Kolhapur, India, and a master's degree in Electrical Engineering from the University of Mumbai in 2001, India. Also, received Ph.D. Degree in Electrical Engineering from Veermata Jijabai Technological Institute Mumbai University, India, in 2016. He has more than 34 years of experience in teaching. Presently, He is serving as Head of the Department and Associate Professor in Electrical engineering, Veermata Jijabai Technological Institute Mumbai. His teaching and research areas include Wireless Communications and Routing algorithms, Applications of Machine Learning and Deep Learning algorithms.



Shubham V. Narkhede did bachelors degree in Electronic and Telecommunications from Ramrao Adik Institute of Technology, Navi Mumbai, India, in 2020. Currently he is pursuing master's in Data science from Stevens Institute of Technology, Hoboken, New Jersey, USA. He has two years of industrial experience as a software developer. He has research interest in Network Intrusion detection using Machine Learning and Deep learning. He is gained experience in statistical knowledge to understand data efficiently.

How to cite this paper: Shamal Salunkhe, Surendra Bhosale, Shubham V. Narkhede, "An Efficient Video Steganography for Pixel Location Optimization Using Fr-WEWO Algorithm based Deep CNN Model", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.15, No.3, pp. 14-30, 2023. DOI:10.5815/ijigsp.2023.03.02