DOI: 10.21917/ijivp.2021.0364

AN OPTIMIZED CONVOLUTION NEURAL NETWORK BASED INTER-FRAME FORGERY DETECTION MODEL - A MULTI-FEATURE EXTRACTION FRAMEWORK

Jatin Patel and Ravi Sheth

School of Information Technology, Artificial Intelligence and Cyber Security, Rashtriya Raksha University, India

Abstract

Surveillance systems are becoming pervasive throughout our daily lives, and surveillance recordings are being used as the essential evidence in criminal investigations. The authenticity of surveillance videos is tough to confirm. One of the most popular methods of video tampering is inter-frame forgery. Using an optimised deep learning methodology, a novel inter-frame forgery detection and localization model is introduced in this research work. Pre-processing, feature extraction, and forgery detection will be the three main phases of the presented design forgery detection model. In the detection model, the original video frames will be pre-processed to improve the image quality. The pre-processing phase includes the frame extraction from video, grey conversion and removal of movement frames as well. Following that, features such as SURF, PCA-HOG features, MBFDF, correlation of adjacent frames, PRG, and OFG based features is extracted. These extracted features will be subjected for forgery detection using Optimised CNN with fine-tuned weights by the hybrid approach. The suggested hybrid paradigm Mayfly Optimization espoused Black Widow Optimization (MO-BWO) is a mathematical fusion of both the Black Widow Optimization (BWO) and Mayfly Optimization Algorithms (MA). In case if the video is detected to be prone to tampers, then the corresponding location gets trapped in the localization phase. Moreover, the detection phase will portray the information regarding the type of tamper like duplication, insertion and deletion of frames. Here, the exact tamper localization is accomplished based on the PRG and OFG. Finally, the supremacy of the MO-BWO+CNN is validated over other conventional models.

Keywords:

Inter-Frame Forgery Detection, Multi-Feature Extraction, CNN Based Tamper Detection, Hybrid optimization Model, PRG And OFG Based Tamper Localization

1. INTRODUCTION

Many governmental and non - governmental organizations (e.g., highways, banks, businesses, and residences) are spotted with several security cameras. Surveillance images are indeed an invaluable source of data for such a variety of sensitive application scenarios. Anyone can alter the video content to create fraudulent clips using efficient and open editing software (e.g., Adobe Premiere, Photoshop, and Cinelerra) [1] [10]-[13]. Even a professional may be deceived by the imperceptible forged contents and minor changes. If the fake videos are widely shared on the internet, they could destroy the social media and government reputation. As a result, it is essential to improve the video forgery detection approach [14]-[17]. The information falsification of the video frame is perhaps the most common video forgery regarding the practical forgery impact. In general, there are two types of video material forgery: 1) Content splicing forgery, or forgery content injection from heterogonous streams, is where the replicated details and superimposed frames come from separate video sources [33]. The detection schemes like frame modulation detector [4] and RNN [5] are based on this ambiguity in the splicing frame, which has obtained considerable improvement. 2) Video copy-move forgery seems to be another popular video information forgery. The superimposed frames and replicated contents from the homogeneous video provide a very convincing counterfeit impression [18]-[22].

The intra-frame and inter-frame copy-move forgeries are the two major video copy-move forgeries [16]. Information copying from a series of consecutive images is repeated and pasted over to other frames to cover or mask the intended items in inter-frame forgery [33]. Several new methods have been used to determine whether a frame is susceptible to inter-frame forgery or contains pristine frames [7] [23]-[25]. Frame characteristics (e.g., DCT coefficient, residual) have been calculated using different methods. The intrinsic properties of the coding structures (e.g., MPEGx) have indeed been calculated between frames in GOPs, and the coding structures (e.g., MPEGx) have also been studied to construct GOPs. Classifiers such as the SVM [8] [12] have been used to validate whether the document is legitimate or a fraud one [8].

Optimized CNN is used for tamper detection, where the weight of CNN is fine-tuned by a new MO-BWO model, which is the conceptual combination of MO and BWO, respectively.

The rest of the paper is organized as: section 2 discusses about the literature works in inter-frame forgery detection. Section 3 shows a new approach for forgery detection in inter-frames: architectural description. Section 4 describes about pre-processing for image enhancement. Section 5 portrays about multi-feature extraction and Section 6 shows on forgery (tamper) detection using convolutional neural network. The forgery localization: prediction residual gradient and optical flow gradient is discussed in Section 7. The results acquired with the propose work is discussed in Section 8. This research work is concluded in Section 9.

Table.1. Nomenclature

Abbreviation	Description
SURF	Speeded Up Robust Features
RNN	Recurrent Neural Networks
PCA-HOG	Principal Component Analysis-Histogram Of Oriented Gradient
PRG	Prediction Residual Gradient
GOP	Groups of Pictures
MBFDF	Mode based Fast Digit Feature
SVM	Support Vector Machine
2D-CNN	2D Convolution Neural Network

DCNN	Deep Convolutional Neural Network
OFG	Optical Flow Gradient
HSV	Hue-Saturation-Value
FLANN	Fast Library for Approximate Nearest Neighbors
3FAT	Three-Stage Foreground Analysis and Tracking
BWO	Black Widow Optimization
RBF-MSVM	RBF Multi-Class Support Vector Machine
MA	Mayfly Optimization Algorithm
HoG	Histogram of Oriented Gradients
SC	Single Compressed

2. LITERATURE REVIEW

2.1 RELATED WORKS

Bakas et al. [33] have developed a new scheme towards detecting inter-frame counterfeits in surveillance footage using optical forensics. Further, they have used the compressed domain video footprints i.e., footprint variance estimation and motion vectors differentiation in recordings to identify the video forgery and to localize it. They recognized the kind of forgery that has appeared in a video through this research work. A total of forty-three genuine and seven hundred and twenty inter-frame forged videos were used in the analysis. Regardless of the nature of the group of pictures or the degree of compression in images, their experimental findings demonstrated that the proposed approach performs consistently well.

Fadl et al. [2] have suggested a new detection scheme based on spatiotemporal information and fusion for deep automatically feature extraction of 2D-CNN for inter-frame forgeries (deletion of frame, frame addition, and frame eduplication). The classification procedure was carried out using a Gaussian RBF-MSVM. The experimental findings of this work had exhibited its supremacy in inter-frame forgeries detection.

Zhong et al. [3] have implemented a video copy-move forgery identification approach that addresses inter/intra-frame forgeries at both the frame and pixel scales. Initially, a coherent moment architecture was created in order to remove multi-dimensional dense moment features from video in a systematic manner. Second, to represent each dimensional feature, an innovative feature extraction method used a 9-digit dense moment feature index to concatenate any other feature sub-map index. Finally, inter-frame best match algorithm was proposed to find the best combinations for each pixel 9-digit dense moment feature index.

Zhao et al. [4] have designed a framework for identifying inter-frame forgeries centered through resemblance analysis, a passive-blind forensics scheme for video shots. The HSV, color histogram comparison, SURF, function extraction, and FLANN matching for double-checking made up the whole process. To identify and locate tampered frames in a video clip, they measured colour histograms of H-S and S-V for each frame in the video. The counterfeit categories in the tampered locations were further verified using extraction of the SURF feature and FLANN matching. In terms of forgery recognition and localization, the experimental findings demonstrate that the proposed detection system has been both appropriate and suitable.

Kingra et al. [5] have introduced a forensic technique for identifying frame-based tampering in video files— particularly those recorded through video surveillance. Frame-based tampering that includes adding, deleting, or duplicating of the frames is generally harder to identify with a clear visual examination. In the encoded H.264 images and MPEG-2 images, the frame-based tampering was identified using OFG and PRG. As a resultant there was enhancement in the proposed work in terms of average precision in localization as well as identification of tampers.

Kaur et al. [6] have used a DCNN which is work on a highly efficient approach for exposing inter-frame forgery in images. The suggested technique has detected forgery with no need for any extra frame details to be pre-embedded. The batch normalisation of input decoders has speed up the training process. The proposed algorithm outperforms with the comparison of existing models based on simulation results obtained on the two different dataset REWIND and GRIP. The suggested algorithm efficiency was checked on YouTube video.

Bakas et al. [7] have proposed a two-step forensic technique for identifying video forgery forms such as frame addition, elimination, and replication. They distinguished outlier frames in the first step using Haralick coded frame similarity, and then performed a finer degree of detection in the second step to remove false positives and thereby the improve forgery detection accuracy was improved.

Liu et al. [8] have proposed a 3FAT algorithm for detecting blue screen compositing. Three main stages make up the 3FAT algorithm: extraction of the foreground block, identification of forged block and tracking of the forged block. In the initial stage, they have deployed a multi-pass foreground locating approach for eliminating the foreground blocks from the images. In the second stage, they have identified the tampered foreground block by means of using the feature-comparison level fusion corresponding to the local features like contrast and luminance. A quick target search algorithm based on Compressive Tracking was used in the final stage to track the tampered block of subsequent frames. The results demonstrate that the 3FAT algorithm was most accurate and faster than other algorithms.

2.2 REVIEW

Most studies present in the literature have discussed on copy-move forgery, regardless of the spatiotemporal domain or frequency domain. However, on multi-resolution processing videos, the recent machine learning models does not achieve satisfactory performance. It should be remembered that the current approaches have four big flaws: 1) only a few techniques or methods are available which can detect the video forgeries at the pixel level. 2) The computational complexity in terms of cost of video forgery detection at the pixel level is almost unacceptably high, even though the video is indeed a few hundred frames long. 3) Most approaches only approach one kind of forgery at a time, it can be inter-frame or intra-frame forgeries; 4) Due to the homogeneous details of the source and target pasted areas, only a few statistical methods can reliably distinguish copy-move forgeries or pristine videos. Therefore, the deep learning model with optimization concept can be indulged.

3. A NEW INTER-FRAMES FORGERY DETECTION MODEL: ARCHITECTURAL DESCRIPTION

Surveillance devices are pervasive in day-to-day lifestyles, and surveillance footage is often used as key forensic evidence examinations. This investigative technique is extensively used to recognize the unlawful acts carried out by criminals in recording. In this forensic area, several varieties of testing processes for forgery detection have currently been undertaken. As technology progresses, different platform is developed to accomplish video forgery. Adobe Photoshop and Video Editor are the two multimedia tools and applications that have recently been created to tamper with or transform media files. One of the most popular methods of video tampering is inter-frame forgery. At the tampering location, the forgery would reduce the association between neighbouring frames. Surveillance videos' reliability is impossible to verify. The challenge regarding quality assurance of security video is an urgent concern.

3.1 PROPOSED MODEL

In this research work, a novel inter-frame forgery detection model is developed using optimized deep learning technique. The proposed work is developed by following three major phases: pre-processing, feature extraction and forgery detection. The Fig.1 shows the architecture of the proposed work. Initially, the collected raw video frames are pre-processed to enhance its image quality. The pre-processing phase includes the frame extraction from video, grey conversion and removal of movement frames. Then, from these pre-processed images, the multi- features like SURF, PCA-HOG feature, MBFDF, correlation of adjacent frames, PRG and OFG based features are extracted. Subsequently, these extracted features are subjected to forgery detection in optimized CNN, whose weights are fine-tuned using a hybrid model. The proposed hybrid model MO-BWO is the conceptual blending of both BWO Algorithm and MA. On detecting a tamper in the frame, the control is transferred to the automatic tamper localization phase, where the localization is made more appropriately. Here localization of tempered frame is calculated based on the prediction residual gradient and optical flow gradient.

4. PRE-PROCESSING FOR IMAGE ENHANCEMENT

4.1 NECESSITY OF PRE-PROCESSING

Preprocessing is indeed the method of preparing data for future review with the aim of fixing or compensating for systemic errors. Leading to inadequate recorded accuracy of the frames, pre-processing is needed to refine the data for future analysis.

- The foregoing reasons provide the necessity for pre-processing.
- Light propagation characteristics such as scattering and absorption induce object loss.
- When cars pass, environmental characteristics such as sun, weather, and hue become more or less noticeable.

Video captures with specificity, such as an unfamiliar rigid picture, unidentified brightness, or reduced illumination sensitivity. In this research work, the pre-processing is carried out on the collected video sequence V^{in} by following four major steps: (a) frame extraction from video sequence, (b) Read the image in the frame, (c) Grey scale conversion and (d) elimination of movement less frames.

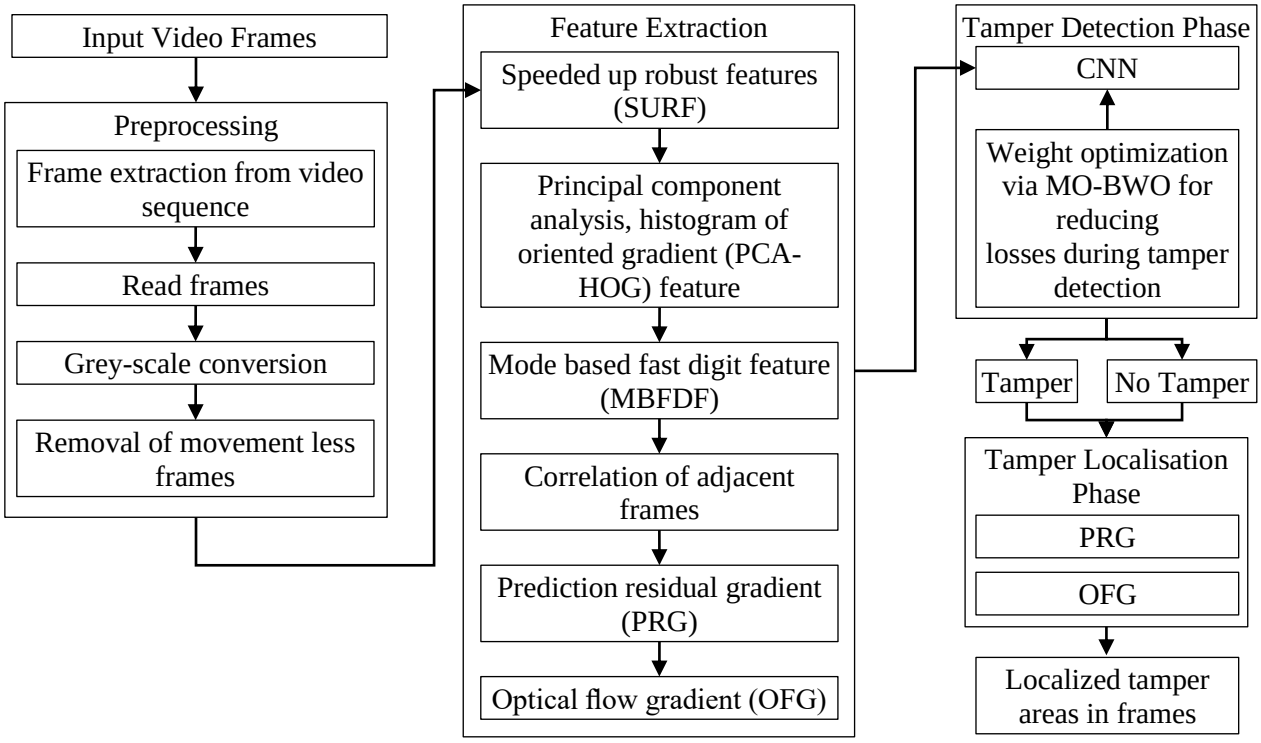


Fig.1. Proposed Inter-frame forgery detection model

All these steps are discussed comprehensively in the upcoming section. The Fig.2 shows the diagrammatic representation of pre-processing stage.

4.2 FRAME EXTRACTION FROM VIDEO

Any video or animation viewed on TV, monitor, laptop, tablet, or even at the cinema is composed of a collection of still photographs. These photographs will be repeated multiple times a second, trying to deceive the eye into believing the object moves. Capturing the photographs one at a time is labor-intensive and time intensive. Here, the frames in the video sequence are extracted using the hasFrame() function as: Fr =hasFrame(V^{in}).

4.2.1 Read Image:

From the extracted frames *Fr*, the still images are read using the read () function as *R*=readFrame (*Fr*)

4.2.2 Grey Scale Conversion:

R Image in RGB format is converted to grey scale using the *rgb2gray*() function as $Grey = rgb2gray(R)$.

4.2.3 Movement-Less Frame Removal:

The movement less frames is then removed from *Grey* by computing their histogram. The frames with threshold value below 4500 are considered to be motion-less and they are removed from *Grey*.

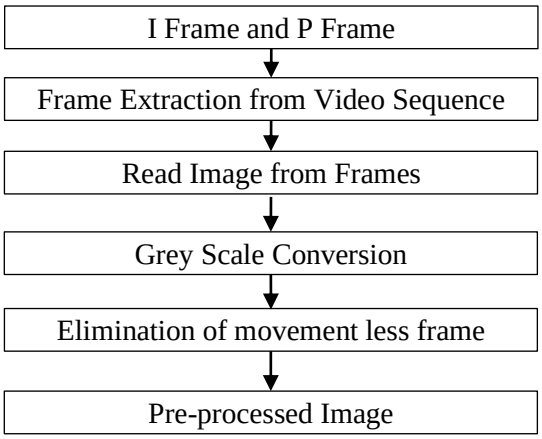


Fig.2. Pre-Processing Stage

5. MULTI-FEATURE EXTRACTION FOR TAMPER DETECTION AND LOCALIZATION

Then, from these pre-processed images, the multi- features like SURF, PCA-HOG feature, MBFDF, correlation of adjacent frames, PRG and OFG based features are extracted. The Fig.3 shows the extracted multi-features.

5.1 SURF FEATURES

The SURF method [34] is a stable and accurate technique towards representing and comparing images in a contextual, similarity invariant manner. The key attraction of the SURF methodology has always been its ability to compute operators quickly using box filters, allowing practical applications including surveillance and scene understanding. SURF is composed of two steps: 'Feature Extraction and Feature Description'. In this research work, the SURF features are extracted from V^{pre}.

- ***Feature Extraction***: The procedure towards detecting the objects of interest is based on a simple Hessian matrix approximation. To find the factor in determining of the Hessian matrix, they first perform 'convolution with a Gaussian kernel', followed by a second-order derivative'. Following Lowe success with LoG approximations (SIFT), SURF uses box filters to advance its interpretation (both convolution and second-order derivative) further more.
- ***Feature Description***: The SURF descriptor is created by following two major phases: (a) reproduceable orientation fixation in a circular region with the information around the key point and (b) alignment of the selected region orientation and SURF descriptor extraction from it. The extracted SURF feature is denoted as F^{SURF}.

5.2 PCA-HOG

HOG: HoG feature is defined as a feature descriptor for detecting V^{pre} homogeneous identical field. The steps followed in HoG feature extraction is manifested below:

- Divide V^{pre} into smaller interlinked areas (cells), and calculate a histogram of gradient directions or edge orientations for each pixel for each cell.
- On the basis of the gradient orientation, each cell is discretized into angular bins.
- Every cell pixel contributes to the weighted gradient with respect to its angular bin.
- In the spatial field, consider a group of neighbouring cells (blocks). The block histogram is constructed by representing the normalized group of histograms, and the group of these normalized groups of histograms is denoted as descriptor. This classification of cells into blocks is based on the grouping and normalisation of histograms. F^{HoG} stands for the extracted HoG features.

PCA: PCA [35] based mathematical features are provided in the following section.

Standard Deviation (SD): The average difference between the mean and the point at which the data is measured by squaring them is referred to as SD. It is mathematically defined in Eq.(1), where *U* denotes a random number and *o* denotes the sample size.

$$SD = \sqrt{\frac{1}{o}\sum_{e=1}^{o}\left(U_e - \overline{U}\right)^2} \tag{1}$$

Covariance: The quantity of variations in dimension from the mean is defined by covariance. The covariance is expressed as in Eq.(2).

$$Cov(U,T) = \frac{\sum_{e=1}^{o}\left(U_e - \overline{U}\right)\left(U_e - \overline{U}\right)}{o} \tag{2}$$

Mean: The mean of data is an arbitrary variable *U* as exposed as $\overline{U} = \frac{1}{o}\sum_{e=1}^{n} U_i$. The extracted HoG-PCA feature is denoted as $F^{HoG-PCA}$.

5.3 MBFDF

The MBFDF is used to distinguish the SC image from the corresponding DC image. The divergence ξ^2 expressed in Eq.(3) is used to determine the consistency of fitting. The observed first order distribution of coefficient is denoted by the notation $P_v(w)$, and in the v^{th} mode, the corresponding theoretical distribution is denoted by $\hat{P}_v(w)$. The extracted MBFDF feature is denoted as F^{MBFDF}.

$$\xi^2 = \sum_{u=1}^{n} \frac{\left(P_v(w) - \hat{P}_v(w)\right)^2}{\hat{P}_v(w)} \tag{3}$$

5.4 CORRELATION OF ADJACENT FRAMES

The similarity of adjacent frames is used to measure the consistency corresponding to the inter-frame content. The correlation coefficient between m^{th} and $(m+1)^{th}$ frame is represented as r_m. The 2D PC value of m^{th} frame at (G,H) location is denoted as $C_m(G,H)$.

$$r_k = \frac{\sum_a \sum_b (C_m(G,H) - \bar{C}_k).P_{m+1}(G,H) - \bar{C}_{m+1})}{\sqrt{(C_m(G,H) - \bar{C}_m)^2.(C_{m+1}(G,H) - \bar{C}_{m+1})^2}} \tag{4}$$

In which $m = 1,2,...,n-1$ and C points to the phase congruency. In addition, n denotes the overall count of frames and $\bar{C}_k$ is the average of 2D PC for m^{th} frame, which is mathematically given in Eq.(5).

$$C_m = \frac{1}{s \times h} \sum_{a,b} C_m(G,H) \tag{5}$$

where, s is the width of the frames in pixels and h corresponds to the video height. The extracted correlation coefficient feature is denoted as F^{CAF}.

5.5 PRG

In detecting the video forgeries, a key role is being played by the prediction residual principle [5]. The disparity between the original frame and the succeeding frame of the original frame is the residual prediction. This includes details about the differences between adjacent frames. Typically, a regular block-matching algorithm is deployed for predicting the next available frame. This algorithm distinguishes subsequent frames by using the corresponding information from the reference frame neighbouring blocks. The estimation residual is calculated by comparing mean square error of each pixel along with the block size 16×16 to motion shifting counterpart in the reference frame.

$$pr(i) = frame(i+1) - BM(frame(i)) \tag{6}$$

$$prg_i = pr_i + 1 - pr_i \tag{7}$$

In Eq.(6), $frame(i)$ points to the frames and $BM(\bullet)$ is the block matching function that predicts the subsequent frames by considering I or P frame as the reference frame. In addition, for $frame(i+1)$, prediction version is computed by using $BM(frame(i))$. Then, for i^{th} pair frame, the prediction residual (pr) is calculated by means of computing the difference between the frame of $(i+1)$ (i.e. $frame(i+1)$) and frame of $(i+1)$ predicted version $BM(frame(i))$. The extracted PRG feature is denoted as F^{PRG}.

5.6 OFG

The optical flow denotes the computation of the movement of brightness patterns among adjacent pixels [5]. The optical flow is used in this analysis to assess the brightness difference from one P frame to the next. At time t, the brightness of a pixel $t(x,y)$ corresponding to frame be $EHS(x,y,t)$. This is mathematically given in Eq.(8) and Eq.(9), respectively.

$$Oflow_i = \iint \left\{ \left(I_x u + I_y v + I_t\right)^2 + \alpha^2 \left[\left(\|\Delta u\|\right)^2 + \left(\|\Delta p\|\right)^2 \right] \right\} dxdy \tag{8}$$

$$Oflowg_i = Oflow_{i+1} - Oflow_i \tag{9}$$

Here the notation I_x, I_y, I_t points to the derivatives corresponding to the intensities along the direction x, y and time dimensions, respectively. The extracted OFG feature is denoted as F^{OFG}.

All these extracted features are together represented as the extracted OFG feature is denoted as

$$F = F^{OFG} + F^{PRG} + F^{CAF} + F^{MBFDF} + F^{HoG-PCA} + F^{SURF}.$$

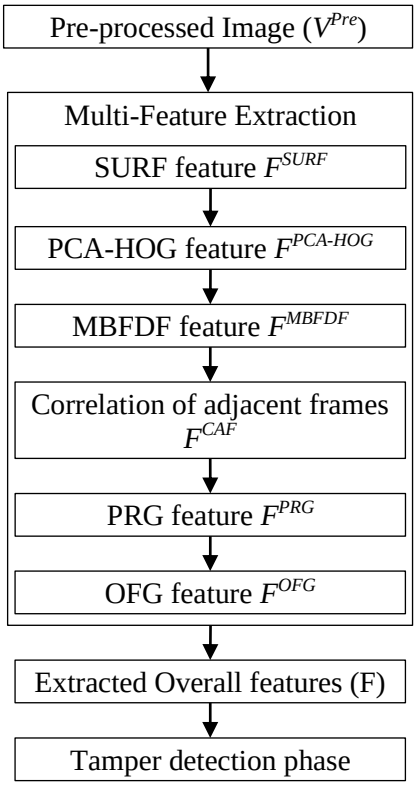


Fig.3. Multi-Feature extraction

6. FORGERY (TAMPER) DETECTION USING OPTIMIZED CONVOLUTIONAL NEURAL NETWORK

6.1 CNN

The extracted features F are used to train the CNN to detect inter frame forgery. The CNN [30]-[32] is a well-known deep learning model that belongs to the artificial neural network class.

It functions on the concept of local connectivity, which separates it from the neural network. The Fig.5 depicts the general CNN architecture. Convolutional layer, pooling layer, and fully-connected layers are the three main layers of the CNN. The neurons of the current layer are bound to the neurons of the preceding stage in the convolution layer, and this type of interconnection is known as the neuron receptive field.

The Fig.5 shows the diagrammatic representation of tamper detection phase. Numerous kernels are united to form complete function maps. Eq.(10) is used to calculate at position (*a*,*b*) in the k^{th} function map lying on the l^{th} layer, and it is represented using the notation

$$Z^{l}_{a,b,k} \cdot Z^{l}_{a,b,k} = Weight^{l^T}_{k} . F^{l}_{a,b} + bias^{l}_{k} \quad (10)$$

where, $Weight^{l}_{k}$ is the vector that corresponds to the weight function in k^{th} attribute map on the l^{th} layer. This weight function is fine-tuned using the newly introduced MO-BWO model. Furthermore, the notation $bias^{l}_{k}$ denotes the bias function of the l^{th} layer l^{th} attribute map. $F^{l}_{a,b}$ represent the input direction in the k^{th} function map on the l^{th} layer. The activation value for $Z^{l}_{a,b,k}$ is then $act^{l}_{a,b,k}$, which is determined using Eq.(11). The activation function $act^{l}_{a,b,k}$ resides on the k^{th} feature map corresponding to the l^{th} layer.

$$act^{l}_{a,b,k} = act\left(Z^{l}_{a,b,k}\right) \quad (11)$$

The activation function $act(\bullet)$ belongs to the k^{th} feature map, which corresponds to the l^{th} layer. The activation function, in general, adds nonlinearities to CNN and helps in the detection of nonlinear characteristics. By decreasing the function map resolution, the pooling layer $pool(\bullet)$ helps in achieving the shift-in-variance. The pooling function $I^{l}_{a,b,k}$ is then evaluated as per Eq.(12) for the activation function $act^{l}_{:,:,k}$.

The nearby neighbourhood is denoted as $\Re_{a,b}$ around the location (*a*,*b*). The output layer is the final layer of CNN, and it contains the softmax feature for performing precise final identification of presence/ absence of tampers in frames (targets). In order to accomplish the optimal goal, CNN loss function must be reduced. The loss function (*Loss*) is seen mathematically in Eq.(13)

$$I^{l}_{a,b,k} = pool\left(act^{l}_{a,b,k}\right), \forall (u,v) \in \Re_{a,b} \quad (12)$$

$$Loss = \frac{1}{M}\sum_{m=1}^{M} l\left(\theta; I^{(m)}, O^{(m)}\right) \quad (13)$$

$$Obj = \min(Loss) \quad (14)$$

Here, the count of input-output relations $\left\{\left(F^{(m)}, I^{(m)}\right); m \in [1,\cdots,M]\right\}$ is denoted as *M*. The θ (overall parameter) and $F^{(m)}$ is the m^{th} input data. In addition, the detected target labels are denoted as $I^{(m)}$ and the output of CNN is denoted as $O^{(m)}$. In order to achieve the objective defined in Eq.(14), the weight function *Weight* of CNN is fine-tuned via new MO-BWO model. The solution fed as input to MO-BWO is depicted in Fig.4.

Solution Encoding

The outcome tells whether there is temper/ no-tamper in the video frames. It also tells about the type of tampers like duplication, deletion or insertion. Once, a tamper is found to be available in a frame, its location is identified via an automatic tamper localization phase.

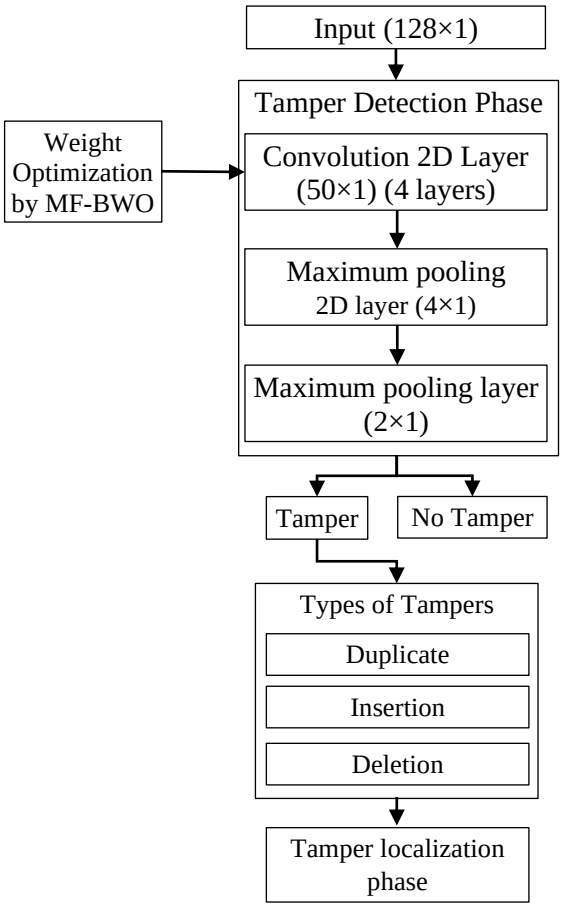


Fig.4. Tempered Detection Frame work

6.2 MO-BWO

Hybrid optimization algorithms combine the advantages of many optimization algorithms to achieve rapid convergence. The convergence behaviour of the hybrid algorithm is said to be superior to that of traditional algorithms. Mayflies in swarms were divided into male and female individuals for the MO [27] algorithm. Since the male mayflies have always been powerful, they perform better in optimization. The BWO [26] was motivated by black widow spiders' unusual mating behaviour. The BWO algorithm achieves exceptional convergence speed and escapes from the issue of local optima in the exploration and exploitation stages. Therefore, the MO and BWO are blended together to form MO-BWO model. The steps followed in the proposed MO-BWO model is manifested below:

Step 1: The population P of the search agent is initialized. The male mayfly $X = X_1, X_2, ..., X_d$ and female mayfly $Y = Y_1, Y_2, ..., Y_c$ is initialized. Here *d*,*c* points to the overall count of male and female mayflies. The velocity of male mayfly and female mayfly Q_{MA} and Q_{FE} is

initialized. In addition, $Weight \rightarrow Z$ and the current iteration is itr and the maximal iteration is Max_{itr}.

Step 2: Compute the fitness of the search agents using Eq.(14)

Step 3: The global best position G_{best} is explored.

Step 4: While $itr < Max_{itr}$ do

Step 5: The velocities as well as the solutions of the female as well as males are updated

Step 6: The next position $X_i(t+1)$ of i^{th} mayfly is updated using the velocity $Q_i(t+1)$ as per Eq.(15)

$$X_i(t+1) = X_i(t) + Q_i(t+1) \tag{15}$$

Step 7: The velocity of i^{th} mayfly at j^{th} direction is V_{ij}, which is updated as per Eq.(15). Here the notation a_1, a_2 points to the positive attraction constants. In addition, $pbest_{ij}$ and $gbest_{ij}$ is the personal best position and the global best position at time t. If the current fitness is better than the past fitness $fit(X(t)) > fit(X(t\text{-}1))$, then update the male mayflies velocities take place as per Eq.(16).

$$\begin{aligned} Q_{ij}(t+1) = Q_{ij}(t) + a_{1.} re^{-\beta r_p^2} \left(pbest_{ij} - X_{ij}(t) \right) \\ + a_2 re^{-\beta r_g^2} \left(gbest_{ij} - X_{ij}(t) \right) \end{aligned} \tag{16}$$

Step 8: The velocity of female mayflies is updated using Eq.(17)

$$\mathbf{Q}_{ij}(t+1) = \begin{matrix} \mathbf{Q}_{ij}(t) + a_{1.} re^{-\beta r_{mf}^2} & if\ fit(Y_i) > fit(X_i) \\ V_{ij}(t) + fl.r & if\ fit(Y_i) \le fit(X_i) \end{matrix} \tag{17}$$

Here the notation β is the visibility coefficient r_{mf} points to the Cartesian distance amongst the female and male mayflies. The random walk coefficient is fl and r is a random number between the range -1 to 1.

Step 9: **Mating of Mayflies:** The mating mechanism between two mayflies is described by the crossover operator. Instead of the traditional arithmetic crossover followed in MO, we introduce a new blend crossover technique to generate solutions with higher convergence. The two parents X and Y together under crossover and they are commonly referred as Z. This mechanism is given in Eq.(18) and Eq.(19), respectively.

$$Z_1^i = \min(Z_i^1 - Z_i^2) - \alpha.d_i \tag{18}$$

$$Z_2^i = \min(Z_i^1 - Z_i^2) - \alpha.d_i \tag{19}$$

Here, Z_1^i and Z_2^i are two parents, who take part in crossover to generate the new solutions. In addition, Z_i^1 and Z_i^2 are the i^{th} element of Z_1^i and Z_2^i, respectively. In addition, the value of α=0.5 is set. These newly acquired solutions will be updated with Procreate phases of the BWO

Procreate: Since the combinations become self-contained, those individuals continue to mate to reproduce the younger breed in tandem, just as they do in nature. Each pair mates with its own web independently. To replicate, an array called alphaα is being generated that would be as long as the widow array and contains random numbers. Then offspring is generated while using the following equation, in which Z_1^i and Z_2^i are parents, and W_1^i and W_2^i are offspring. This is demonstrated in Eq.(20) and Eq.(21).

$$W_1^i = \alpha \times Z_1^i + (1-\alpha).Z_2^i \tag{20}$$

$$W_2^i = \alpha \times Z_2^i + (1-\alpha).Z_1^i \tag{21}$$

Step 10: Return the best solution from P

7. FORGERY LOCALIZATION BASED ON PREDICTION RESIDUAL GRADIENT AND OPTICAL FLOW GRADIENT

7.1 LOCALIZATION PROCEDURE

By taking the account of variable duration GOP of the pattern, forgery has been automatically localized. The localization method given below seems to be the suggested solution for any GOP pattern.

Step 1: Locate the position upon its x-axis that represents the PRG or OFG value and has 2 consecutive peak points, $Loc(i), Loc(i+1)$

Step 2: Multiplication of $Loc(i+1)$ by g, where g is the difference between two consecutive P frames or it can be difference between two consecutive I frames and P frames.

Step 3: This value defines the exact location of frame where frame forgery begins or stop like insertion/deletion or frame replication. It likely that the received value defines the exact forgery position within the GOP. As a result, the precise location of the tampering is discovered inside the GOP duration spectrum. As a resultant, the specific model of the forged frame ((frame insertion, deletion or duplication) is localized. The Fig.6 shows the diagrammatic representation of tamper localization approach.

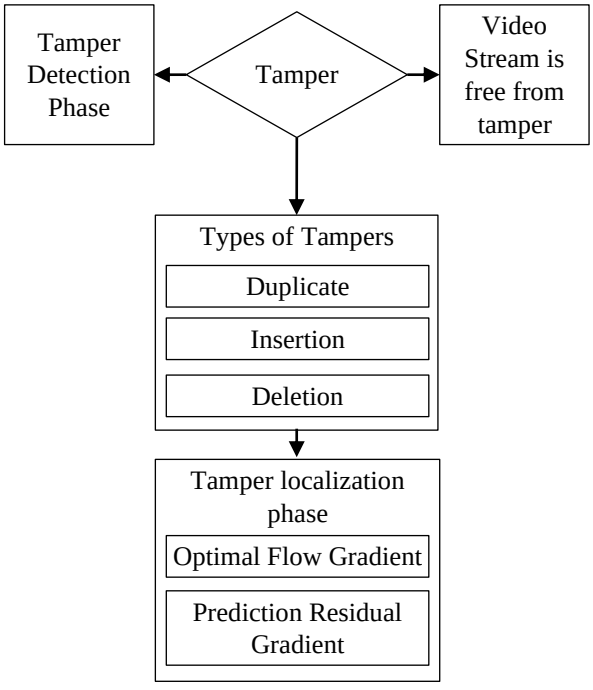


Fig.5. Tamper Localization Phase

8. RESULT SAND DISCUSSIONS

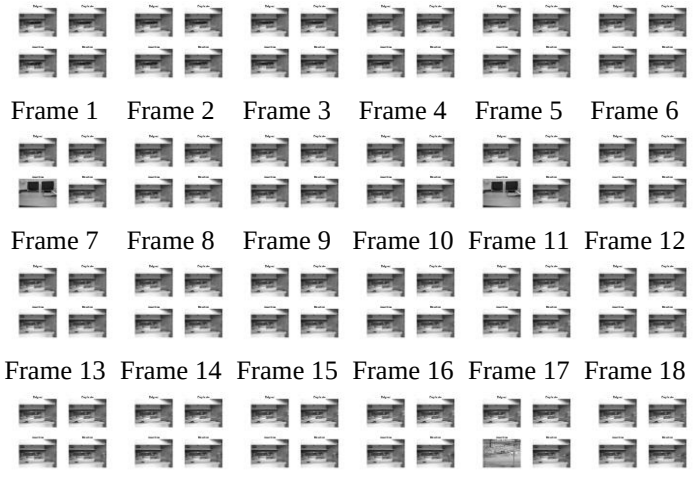
8.1 EXPERIMENTAL SETUP

The proposed inter-frame video detection model was implemented in MATLAB. The results acquired are evaluated in terms of convergence analysis and performance analysis (positive, negative and others). In order to validate the supremacy of the MO-BWO, a comparative evaluation is made between the proposed as well as existing works. The verification is accomplished with the data collected from: https://drive.google.com/file/d/0Bw7zEDZoqsHGWHBrcWpVLW5Yck0/view [Access Date: 2021-03-27]. The database includes 9 videos (video1, video2, video3, video 4, viedo5, video6, video7, video8 and video9). The count of frames in original, duplicate, insertion and deletion are provided in Table.1.

Table.1. Count of Video Frames

Video type	Original	Duplicate	Insertion	Deletion
Video1	157	157	165	149
Video2	152	152	162	142
Video3	295	295	303	287
Video4	294	294	305	283
Video5	96	96	106	86
Video6	35	35	45	25
Video7	100	100	115	85
Video8	43	43	52	34
Video9	92	92	104	80

An example collected sample inter-frame surveillance videos, where we have 36 frames (frame1 to frame 36) inclusive of original and 3 types of forgeries performed in the frames. The proposed work (MO-BWO+CNN) is compared over the existing models like the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO+CNN, OFG+PRG based localization [5], respectively. This evaluation is undergone by varying the count of iterations from 0, 5, 10, 25, respectively. The effect of forgery is demonstrated in Fig.8.



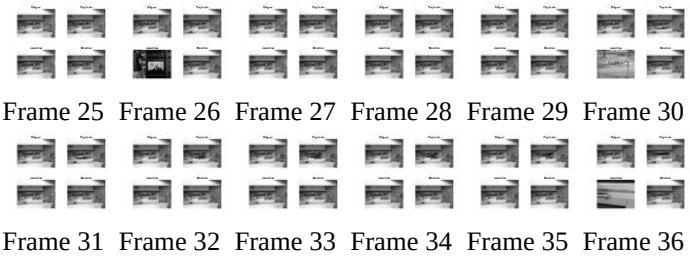


Fig.6. Collected Video-sequence with original, deleted, inserted and duplicated frames

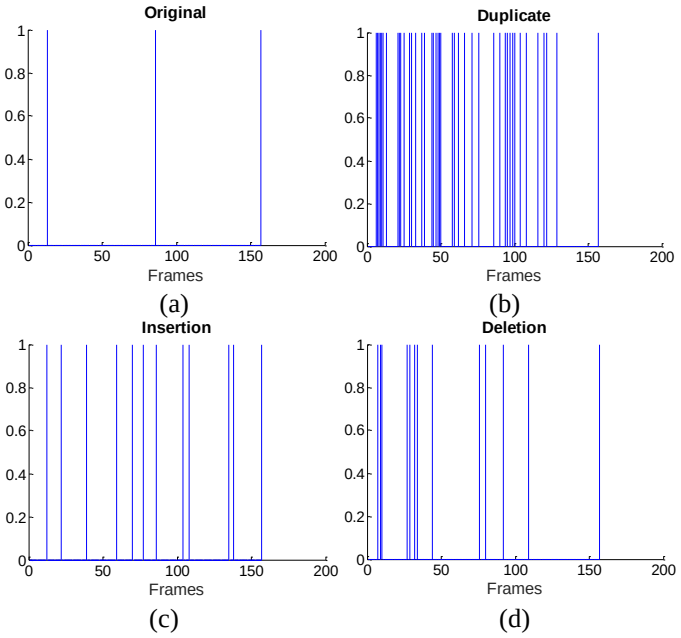


Fig.7. Influence of forgery on video frames

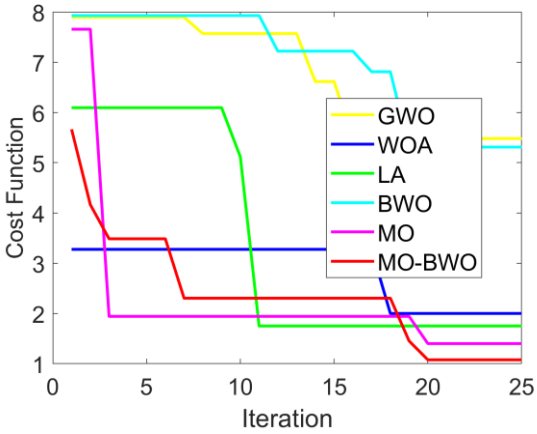
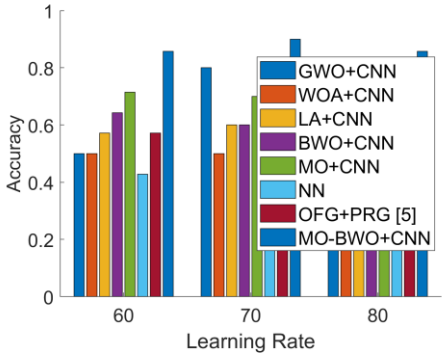


Fig.8. Convergence Analysis

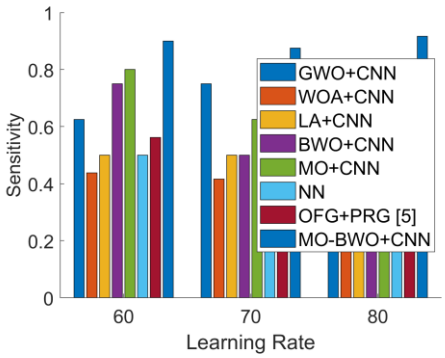
8.2 CONVERGENCE ANALYSIS

The convergence of the proposed hybrid model is validated over the individual optimization models like GWO, WOA, LA, BW and MA, respectively with respect to minimum loss as in Eq.(14). This evaluation is undergone by varying the count of iterations from 0, 5, 10, 25, respectively. The outcome acquired is recorded graphically and it is exhibited in Fig.9. On analyzing the

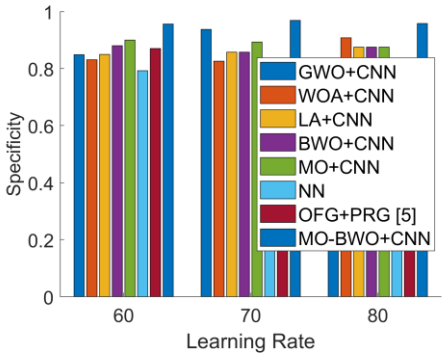
acquired results, the cost function of both the proposed as well as existing models seems to be higher at the lowest count of iterations (i.e. in between 0 to 5th iteration). But also, in this range the cost function of the MO-BWO was found to be lower than MA, BW, GWO and LA, respectively. Then, as the count of iterations increases, a downfall of convergence is recorded in both the proposed as well as existing models. At the 20th iteration, the cost function of the MO-BWO is lower (~1), which is 83.3%, 54.5%, 50%, 84.6% and 33.3% better than the recorded cost functions of GWO, WOA, LA, BW and MA, respectively. Interestingly, from the outcomes, the cost function of the MO-BWO was found to be much lower even at the highest count of iterations, which clearly portrays its efficiency in lessening the detection losses for massive video frames. Moreover, this reduction in the cost function is owing to the newly introduced model is good in generating the global solutions without getting trapped into the local optima.



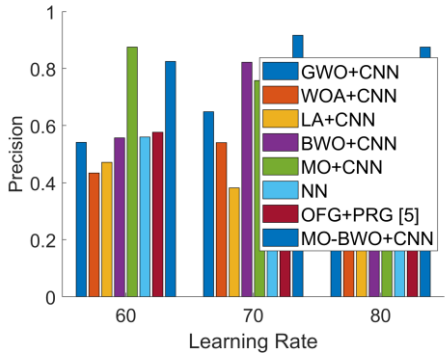
(a)



(b)

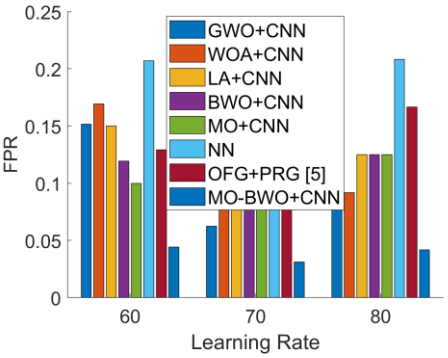


(c)

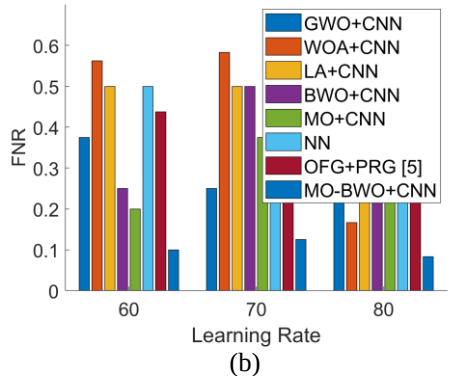


(d)

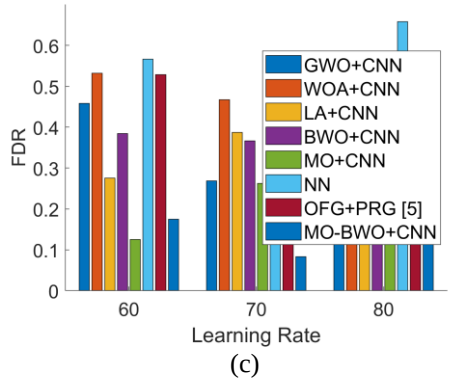
Fig.9. Positive Performance of MO-BWO+CNN in terms of (a) Accuracy, (b) Sensitivity, (c) Specificity and (d) Precision



(a)



(b)



(c)

Fig.10. Negative Performance analysis of MO-BWO+CNN in terms of (a) FPR, (b) FNR, (c) FDR

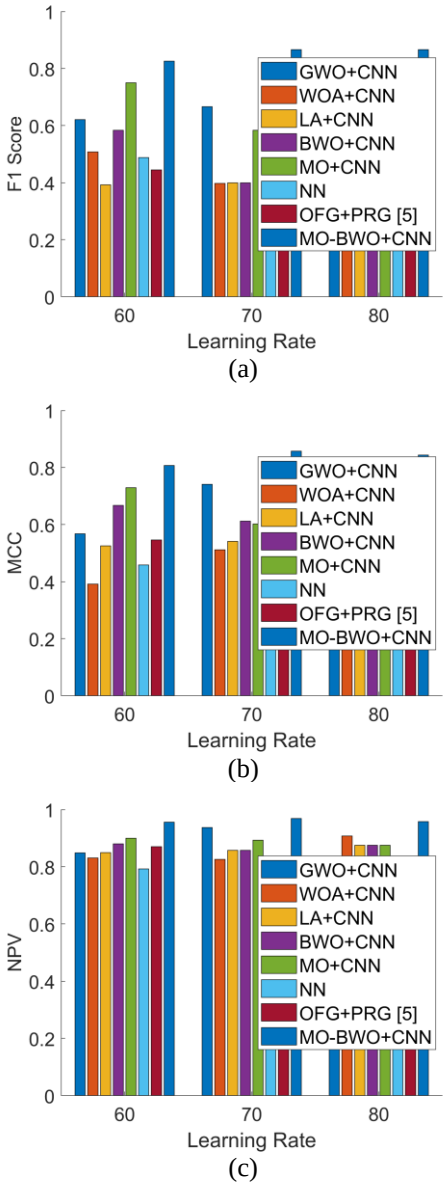


Fig.11. Other Performance analysis of MO-BWO+CNN in terms of (a) F1-Score, (b) MCC and (c) NPV

8.3 PERFORMANCE ANALYSIS

The performance of the MO-BWO model is validated over the existing models (both algorithmically as well as classification based) like GWO+CNN, WOA+CNN, LA+CNN, BW+CNN, MA+CNN, NN and OFG+PRG based tamper detection and localization, respectively. All these evaluations are undergone by varying the learning rate from 60, 70 and 80, respectively.

The evaluation is made in terms of positive measures (accuracy, specificity, specificity and precision), negative measures (FPR, FNR and FDR), and other measures (F1-Score, MCC and NPV), respectively.

The results acquired in terms of positive performance are shown in Fig.10. It is crucial to maintain the detection accuracy of the MO-BWO+CNN at its highest range to exhibit its supremacy over the existing models. On observing the accuracy measure, the MO-BWO+CNN seems to have achieved the highest value for every variation in the learning arte. At learning rate=80, the accuracy of the MO-BWO+CNN is (~) 85%, which is 50%, 23.5%, 34.1%, 32.9%. 31.82%, 76.4% and 5.8% better than the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO+CNN, OFG+PRG based tamper detection and localization at learning percentage=70.

In addition, the sensitivity, specificity as well as precision of the MO-BWO+CNN are higher for every variation in the learning rate. The sensitivity of the MO-BWO+CNN is above 85% for all the three variation in the learning rate, while the sensitivity of the existing work is below 78%. In addition, the specificity as well as precision of the MO-BWO+CNN is above 96% and 83%, respectively which is the most optimal score. The results acquired in terms of negative performance are shown in Fig.11.

In addition, the FPR of the MO-BWO+CNN at learning percentage=80 is 75%, 50%, 58.3%, 57.6%, 58.2%, 80% and 65% better than the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO+CNN, OFG+PRG based tamper detection and localization, respectively. In addition, the FNR, of the MO-BWO+CNN has achieved the most favourable value (least value), which is approximately less than 0.1.

Table.2. Overall Performance Evaluation: proposed over conventional models

Measures	**GWO+CNN**	**WOA+CNN**	**LA+CNN**	**BWO+CNN**	**MO+CNN**	**NN**	**OFG+PRG based tamper detection and localization**	**MO-BWO+CNN**
Accuracy	0.5	0.5	0.57143	0.64286	0.71429	0.42857	0.57143	0.85714
Sensitivity	0.625	0.4375	0.5	0.75	0.8	0.5	0.5625	0.9
Specificity	0.84829	0.83077	0.85	0.88077	0.9	0.79293	0.87083	0.95577
Precision	0.54167	0.43428	0.47176	0.55724	0.875	0.55961	0.5768	0.825
FPR	0.15171	0.16923	0.15	0.11923	0.1	0.20707	0.12917	0.044231
F1-Score	0.62079	0.50764	0.39286	0.58333	0.75	0.48854	0.44444	0.82639
MCC	0.56751	0.3914	0.52583	0.66744	0.72996	0.45881	0.54626	0.80714
FNR	0.375	0.5625	0.5	0.25	0.2	0.5	0.4375	0.1
NPV	0.84829	0.83077	0.85	0.88077	0.9	0.79293	0.87083	0.95577
FDR	0.45833	0.53213	0.27554	0.38459	0.125	0.56638	0.52888	0.175

In addition, the FDR of the MO-BWO+CNN is found to be lower than that of the existing works. In addition, the other performance measures like F1-Score, MCC and NPV of the MO-BWO+CNN had achieved the better than (highest value), which is the most favourable one, which is evident from Fig.11. Therefore, from the evaluation, a clear conclusion can be derived that the MO-BWO+CNN is much more suitable for inter-frame forgery detection.

8.4 OVERALL PERFORMANCE ANALYSIS

The overall performance of the MO-BWO+CNN is evaluated and the corresponding results acquired are tabulated in Table.2. The overall accuracy of the MO-BWO+CNN is 0.85714, which is 41.6%, 41.6%, 33.3%, 24.9%, 16.6%, 50% and 33.3% better than the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO+CNN, NN, OFG+PRG based tamper detection and localization, respectively. In addition, the MO-BWO+CNN has achieved the highest specificity, Sensitivity as well as Precision. The Specificity of the MO-BWO+CNN is 0.95577, which is the highest value when compared to GWO+CNN = 0.84829, WOA+CNN = 0.83077, LA+CNN= 0.85, BWO+CNN= 0.88077, MO+CNN= 0.9, NN= 0.79293 and OFG+PRG based tamper detection and localization = 0.87083. On the other hand, the negative measures like FPR, FNR and FDR seems to have recorded the least value, which is identified to be the most favourable one. The other measures like F1-Score, MCC and NPV has been recorded to be the best value in case of the MO-BWO+CNN. The F1-Score of the MO-BWO+CNN is 0.82639, which seems to be 24.8%, 38.5%, 52.4%, 29.4%, 9.2%, 40.8% and 46.2% better than the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO+CNN, NN and OFG+PRG based tamper detection and localization, respectively. The MO-BWO+CNN has indeed achieved the maximal NPV value as 0.95577, while the NPV of GWO+CNN= 0.84829, WOA+CNN= 0.83077, LA+CNN= 0.85, BWO+CNN= 0.8807, MO+CNN= 0.9, NN= 0.792, OFG+PRG based tamper detection and localization = 0.8708. Therefore, from the evaluation, a clear conclusion can be derived that the MO-BWO+CNN has achieved the most favourable outcomes, while compared to the existing models. All these improvements are owing to the generation of the new solutions by blend crossover, which is embedded with the potential of generating optimal solutions. Therefore, a clear conclusion can be derived that the propose work is much appropriate for inter-frame forgery detection.

9. CONCLUSION

In this paper, a novel inter-frame forgery detection and localization model was constructed. The forgery detection model included the Pre-processing, feature extraction and forgery detection phase. The original video frames were pre-processed at first to improve the image quality. The pre-processing phase includes the frame extraction from video, grey conversion and removal of movement frames. Following that, features such as SURF, PCA-HOG features, MBFDF, correlation of adjacent frames, PRG, and OFG based features were extracted from these pre-processed images. These extracted features were subjected to forgery detection, which was constructed using an optimised CNN. In order to enhance the detection accuracy of CNN, its weights were fine-tuned using the newly introduced MO-BWO model. The CNN also talks about the type of tamper in the video. In case if the video is detected to be prone to tampers, then the control is transferred to the automatic localization phase. In the automatic localization of forgery phase, the prediction residual gradient and optical flow gradient are used to define the exact localization of the tampers. The performance of the proposed inter-frame forgery detection framework is validated over the existing models (both algorithmically as well as classification based) like GWO+CNN, WOA+CNN, LA+CNN, BW+CNN, MA+CNN, NN and OFG+PRG based tamper detection and localization, respectively. The overall accuracy of the MO-BWO+CNN is 0.85714, which is 41.6%, 41.6%, 33.3%, 24.9%, 16.6%, 50% and 33.3% better than the existing models like GWO+CNN, WOA+CNN, LA+CNN, BWO+CNN, MO CNN, NN, OFG+PRG based tamper detection and localization, respectively.

REFERENCES

[1] A. Ullah, K. Muhammad, T. Hussain, S.W. Baik and V.H.C. De Albuquerque, "Event-Oriented 3D Convolutional Features Selection and Hash Codes Generation Using PCA for Video Retrieval", *IEEE Access*, Vol. 8, pp. 196529-196540, 2020.

[2] Sondos Fadl and Qi Han Qiong Li, "CNN Spatiotemporal Features and Fusion for Surveillance Video Forgery Detection", *Signal Processing: Image Communication*, Vol. 90, pp. 1-14, 2020.

[3] Jun Liu Zhong, Chi Man Pun and Yan Fen Ganb, "Dense Moment Feature Index and Best Match Algorithms for Video Copy-Move Forgery Detection", *Information Sciences*, Vol. 23, No. 2, pp. 1-13, 2020.

[4] Dong Ning Zhao and Ren-Kui Wang, "Inter-Frame Passive-Blind Forgery Detection for Video Shot based on Similarity Analysis", *Multimedia Tools and Applications*, Vol. 77, pp. 25389-25408, 2018.

[5] Staffy Kingra and Naveen Aggarwal, "Inter-Frame Forgery Detection in H.264 Videos using Motion and Brightness Gradients", *Multimedia Tools and Applications*, Vol. 76, No. 24, pp. 25767-25786, 2017.

[6] Harpreet Kaur and Neeru Jindal, "Deep Convolutional Neural Network for Graphics Forgery Detection in Video", *Wireless Personal Communications*, Vol. 112, pp. 1763-1781, 2020.

[7] Jamimamul Bakas, Buchira Naskar and Rahul Dixit, "Detection and Localization of Inter-Frame Video Forgeries based on Inconsistency in Correlation Distribution between Haralick Coded Frames", *Multimedia Tools and Applications*, Vol. 78, pp. 4905-4935, 2018.

[8] Yuqing Liu and Tianqiang Huang, "A Novel Video Forgery Detection Algorithm for Blue Screen Compositing based on 3-Stage Foreground Analysis and Tracking", *Multimedia Tools and Applications*, Vol. 77, No. 12, pp. 1-22, 2017.

[9] S. Jia, Z. Xu, H. Wang, C. Feng and T. Wang, "Coarse-to-Fine Copy-Move Forgery Detection for Video Forensics", *IEEE Access*, Vol. 6, pp. 25323-25335, 2018.

[10] L. Su, H. Luo and S. Wang, "A Novel Forgery Detection Algorithm for Video Foreground Removal", *IEEE Access*, Vol. 7, pp. 109719-109728, 2019.

[11] L. Su, C. Li, Y. Lai and J. Yang, "A Fast Forgery Detection Algorithm Based on Exponential-Fourier Moments for Video Region Duplication", *IEEE Transactions on Multimedia*, Vol. 20, No. 4, pp. 825-840, 2018.

[12] L. D. Amiano, D. Cozzolino, G. Poggi and L. Verdoliva, "A PatchMatch-Based Dense-Field Algorithm for Video Copy-Move Detection and Localization", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 29, No. 3, pp. 669-682, 2019.

[13] C. Feng, Z. Xu, S. Jia, W. Zhang and Y. Xu, "Motion-Adaptive Frame Deletion Detection for Digital Video Forensics", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 27, No. 12, pp. 2543-2554, 2017.

[14] F. Khelifi and A. Bouridane, "Perceptual Video Hashing for Content Identification and Authentication", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 29, No. 1, pp. 50-67, 2019.

[15] Q. Liu, "An Improved Approach to Exposing JPEG Seam Carving under Recompression", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 29, No. 7, pp. 1907-1918, 2019.

[16] S. Ghimire, J.Y. Choi and B. Lee, "Using Blockchain for Improved Video Integrity Verification", *IEEE Transactions on Multimedia*, Vol. 22, No. 1, pp. 108-121, 2020.

[17] Y. Zheng, Y. Cao and C. Chang, "A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification", *IEEE Transactions on Information Forensics and Security*, Vol. 15, pp. 620-634, 2020.

[18] E.A. Armas Vega, E. Gonzalez Fernandez, A.L. Sandoval Orozco and L.J. Garcia Villalba, "Passive Image Forgery Detection Based on the Demosaicing Algorithm and JPEG Compression", *IEEE Access*, Vol. 8, pp. 11815-11823, 2020.

[19] X. Ding, G. Yang, R. Li, L. Zhang, Y. Li and X. Sun, "Identification of Motion-Compensated Frame Rate Up-Conversion Based on Residual Signals", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 7, pp. 1497-1512, 2018.

[20] A. Chintha, "Recurrent Convolutional Structures for Audio Spoof and Video Deepfake Detection", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14, No. 5, pp. 1024-1037, 2020.

[21] M. Saddique, K. Asghar, U.I. Bajwa, M. Hussain, H.A. Aboalsamh and Z. Habib, "Classification of Authentic and Tampered Video Using Motion Residual and Parasitic Layers", *IEEE Access*, Vol. 8, pp. 56782-56797, 2020.

[22] H. Chen, W. Wang, J. Zhang and Q. Zhang, "EchoFace: Acoustic Sensor-Based Media Attack Detection for Face Authentication", *IEEE Internet of Things Journal*, Vol. 7, No. 3, pp. 2152-2159, 2020.

[23] J. Hou and H. Lee, "Detection of Hue Modification using Photo Response Nonuniformity", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 27, No. 8, pp. 1826-1832, 2017.

[24] E.T. Grogan, "Techniques to Detect Modified Video", *SMPTE Motion Imaging Journal*, Vol. 127, No. 1, pp. 61-67, 2018.

[25] I.J. Yu, S.H. Nam, W. Ahn, M.J.Kwon and H.K. Lee, "Manipulation Classification for JPEG Images using Multi-Domain Features", *IEEE Access*, Vol. 8, pp. 210837-210854, 2020.

[26] Vahideh Hayyolalam, Ali Asghar and Pourhaji Kazem, "Black Widow Optimization Algorithm: A Novel Meta-Heuristic Approach for Solving Engineering Optimization Problems", *Engineering Applications of Artificial Intelligence*, Vol. 87, pp. 1-19, 2020.

[27] Seyedali Mirjalili, "Moth-Flame Optimization Algorithm: A Novel Nature-Inspired Heuristic Paradigm", *Knowledge-Based Systems*, Vol. 89, pp. 228-249, 2015.

[28] M. Marsaline Beno, I.R. Valarmathi, S.M. Swamy and B.R. Rajakumar, "Threshold Prediction for Segmenting Tumour from Brain MRI Scans", *International Journal of Imaging Systems and Technology*, Vol. 24, No. 2, pp. 129-137, 2014.

[29] B.R. Rajakumar, "Static and Adaptive Mutation Techniques for Genetic Algorithm: A Systematic Comparative Analysis", *International Journal of Computational Science and Engineering*, Vol. 8, No. 2, pp. 180-193, 2013.

[30] Malige Gangappa, C. Kiran Mai and P. Sammulal, "Enhanced Crow Search Optimization Algorithm and Hybrid NN-CNN Classifiers for Classification of Land Cover Images", *Multimedia Research*, Vol. 2, No. 3, pp. 12-22, 2019.

[31] Raviraj Vishwambhar Darekar and Ashwinikumar Panjabrao Dhande, "Emotion Recognition from Speech Signals using DCNN with Hybrid GA-GWO Algorithm", *Multimedia Research*, Vol. 2, No. 4, pp. 12-22, 2019.

[32] G. Gokulkumari, "Classification of Brain tumor using Manta Ray Foraging Optimization-based DeepCNN Classifier", *Multimedia Research*, Vol. 3, No. 4, pp. 1-12, 2020.

[33] M.R. Puttaswamy, "Improved Deer Hunting Optimization Algorithm for Video based Salient Object Detection", *Multimedia Research*, Vol. 3, No. 3, pp. 1-16, 2020.

[34] H.J. Bouchech, S. Foufou and M. Abidi, "Strengthening Surf Descriptor with Discriminant Image Filter Learning: Application to Face Recognition", *Proceedings of International Conference on Microelectronics*, pp. 136-139,2014.

[35] Jamimamul Bakas, Ruchira Naskar and Sambit Bakshi, "Detection and Localization of Inter-Frame Forgeries in Videos based on Macroblock Variation and Motion Vector Analysis", *Computers and Electrical Engineering*, Vol. 89, pp. 1-23, 2020.