Computers, Materials & Continua
DOI: 10.32604/cmc.2022.023545
Article

Tech Science Press

Multiple Forgery Detection in Video Using Convolution Neural Network

Vinay Kumar[1,*], Vineet Kansal[2] and Manish Gaur[2]

[1]Centre for Advanced Studies, Dr. A P J Abdul Kalam Technical University, Lucknow, India
[2]Institute of Engineering and Technology, Dr. A P J Abdul Kalam Technical University, Lucknow, India
*Corresponding Author: Vinay Kumar. Email: vinay.kumar@cas.res.in
Received: 12 September 2021; Accepted: 16 November 2021

Abstract: With the growth of digital media data manipulation in today's era due to the availability of readily handy tampering software, the authenticity of records is at high risk, especially in video. There is a dire need to detect such problem and do the necessary actions. In this work, we propose an approach to detect the interframe video forgery utilizing the deep features obtained from the parallel deep neural network model and thorough analytical computations. The proposed approach only uses the deep features extracted from the CNN model and then applies the conventional mathematical approach to these features to find the forgery in the video. This work calculates the correlation coefficient from the deep features of the adjacent frames rather than calculating directly from the frames. We divide the procedure of forgery detection into two phases–video forgery detection and video forgery classification. In video forgery detection, this approach detect input video is original or tampered. If the video is not original, then the video is checked in the next phase, which is video forgery classification. In the video forgery classification, method review the forged video for insertion forgery, deletion forgery, and also again check for originality. The proposed work is generalized and it is tested on two different datasets. The experimental results of our proposed model show that our approach can detect the forgery with the accuracy of 91% on VIFFD dataset, 90% in TDTV dataset and classify the type of forgery–insertion and deletion with the accuracy of 82% on VIFFD dataset, 86% on TDTV dataset. This work can helps in the analysis of original and tempered video in various domain.

Keywords: Digital forensic; forgery detection; video authentication; video interframe forgery; video processing; deep learning

1 Introduction

Nowadays, the use of the Internet has become the main component of everyone's life as digital technology is evolving very rapidly every day. As the growth in digital technology [1], the demand for multimedia supported devices like webcam, desktops, laptops, smartphones, etc. are raised. In the current time, the primary means for communication is multimedia devices through digital images

and videos with the help of social media platforms. With the awareness in people about Internet and multimedia devices in various court cases, these digital images and videos play a vital role as a piece of chief evidence for a crime or incidents. In the court, now video captured with the help of CCTV camera can be used as a shred of evidence. But the challenge is to validate the integrity of the evidence otherwise manipulated image or video can be used as a piece of false evidence and cause a severe legal problem [2]. The easy availability of manipulating tools and software for digital photos and videos at a meagre cost increases the risk of false evidence in the current days.

Due to the above problem in the field of multimedia like images and videos, researchers are carrying forward the research to investigate the integrity and reliability of multimedia data effectively. And the research approaches in the multimedia integrity checking [3] and authentication is mainly focused on digital forensics that deals with the digital evidence investigation and recovery that helps to review cyber-crimes.

There usually are a few techniques that help in predicting the originality or authenticity of video content, which denominated as video forgery detection techniques [4]. These video forgery detection techniques come under one of the two fundamental types of forensic approach–active and passive [5]. In the active forensic approach, specific equipment and embedded software is the essential requirement, and using this approach quality of the video is also demoted. Active forensics applies where digital watermarks or digital signatures present in the video, so any forgery in such video detects with the help of previous information about the source. Also, active forensic has some limitation as it is not applicable when forgery happens ere digital watermarking, and digital signatures are embedded in the video. As an active forensic approach has a limited scope and to overcome the problems present in an active forensic approach, the other passive forensic method is considered. The passive forensic approach helps in the investigation of video for its authenticity using the special temporal or/and spatial aspects present in it. The passive forensic approach can discover any tempering conducted at inter-frame or intra-frame level.

Video forgery detection methods [6,7] are applied in the two major areas that are inter-frame video forgery detection [8] and intra-frame video forgery detection. These video forgery areas are distributed based on the tempering performed in the video illegitimately. In intra-frame video forgery techniques like copy-move, re-sampling and slicing are present to temper the original video data, and it also involves forgery at object-level. Part of intra-frame video forgery is when original content of different frames are changed like few region or regions are pasted from the same frame to other location of the same frame or from one frame to another frame. For dealing with this type of forgery, pixel noise features may be considered. On the other hand, inter-frame video forgery involves alteration of the original video within the distribution of the frame. An example of inter-frame video forgery is where frames/frame is inserted from another source to the original video or deleted from the original video or sequence of frame modified in an illicit way to change the flow of events present initially.

Currently, in forensic investigations, video footage legislates one of the significant evidence. There are three categories of tampering has been found with these pieces of evidence by adding, removing or reproducing/duplicating specific frames within the video footage illegitimately. With the outcome of these forgery video footage/evidence semantically distinct from the genuine video footage and afterwards the forged video footage can be presented as actual evidence to trick the legal investigation. These generated tempered videos can be spread through the online seer to proclaim the fraudulent news. Due to the above-addressed problem in the forensic examination of video evidence, there is a need to detect the inter-frame forgery in videos with high accuracy and effectiveness.

In today's era, deep learning [9] is a widely adopted technique to be applied with high accuracy and can deal with the larger dataset. In the area of deep learning, a CNN (convolution neural network) [10] is a part of deep neural network and used to classify the visual representation like images based on the prominent features present in them. In this work, we also use deep learning to find the essential deep features for further examination. This work focuses on the detection of inter-frame video forgery. But this work carries forward the research to detect the forgery in the video for frame insertion and frame deletion. This work uses two models. The first model is a binary detection model that detects only for the authenticity of the video that implies the video is temped or not. And the second model identifies the type of forgery present in the video that suggests whether any frame is inserted or deleted. And the contribution of this work is as follows:

- The proposed work is fast in the sense that with the help of parallel processing, this work reduces the processing time for the frame as it uses two parallel CNN to process the frames and find the noticeable features.
- The proposed work uses two different models to detect the forgery present in the video and classify the type of forgery in the video, respectively.
- The proposed work does not have any relation with the conditions such as the number of frames altered and the recording device used to record the video.
- Unlike the other works, this work does not classify or detect the forgery in the video with the help of deep learning output directly. The proposed approach only uses the deep features extracted from the CNN model then apply the conventional mathematical approach on these features to find the forgery in the video.
- The proposed method is robust and generalized as it is not restricted to a single dataset and gives comparable results for the various dataset as we tested on two datasets.

The rest of the paper is as follows: A few literature reviews in the area of video forgery detection is described in Section 2. Section 3 explains the parallel CNN used in this work to process the frames concurrently. The proposed framework for inter-frame video forgery detection is explained in Section 4. Section 5 exhibits the experimental outcomes. And finally, this work concludes in Section 6 with a few future work directions.

2 Related Work

This section discusses a few approaches associated with video forgery detection domain. This section also reviews a summary of the present time scenario in the research field of video forgery detection.

The authors in [11] present an approach in which they excrete the trivial interference between the frames such as vibration in focus, a swift change in lighting etc. and also they detect the frame deletion point (FDP) in the video. They also consider videos with the variable motion. The authors use frame motion features and term as fluctuation features to find the frame deletion point. They use an intra-prediction exclusion method that further enhances the fluctuation feature for the appropriation of different video motions. Then the authors obtain the position of the FDP with the help of these improved features. The authors conducted the experiments for multiple motions of videos and various interfering frames present in the videos. They registered the results with true positive rate and false alarm rate to 90% and 0.3% respectively.

In the other work, the authors [12] present four approaches for copy-paste video forgery detection. They give the first approach to detect the copy-paste forgery that entails insertion and deletion of

frames in or from a video using a sensor pattern noise. The authors use a correlation of sensor pattern noise to detect the forgery in the video. Their approach can detect tempering in the video, which is captured from static or motion camera. They also state that their approach is more robust and enhanced than the already present noise-residue forensic method [13], which is the second approach they tested. The third approach authors described is the revised version of the image forensic approach [14] and termed as CFA-V, where CFA stands for Color Filter Array The artefacts from the color filter array are used to detect the forged frames of a video. And the last approach the authors proposed a mere pixel clustering method based on Hausdorff distance and named as H-DC. They use the video dataset of various encoding types such as MPEG-4, MJPEG, H.264/AVC, and MPEG-2 to perform their experiment and prove their approach. The authors perform their experiments in such an environment that mimic a scenario for forgery in real-world after diversifying variable parameter. The authors register the average accuracy in range 64.5%−82% for noise-residue method, the average accuracy in range 89.9%−98.7% for sensor pattern noise correlation and the average accuracy for CFA-V and H-DC is in the range of 83.2%−93.3% and 79.1%−90.1%, individually.

In [15] authors introduce a two-step approach to detect forgery in the video. They identify outlier using the distribution of inter-frame Haralick correlation in the first phase. In the second phase, they localize the forgery with the help of block-level examination of frames. The authors use three different datasets for their experiment. The first dataset contains 17 videos with twenty-nine frame rate collected from the SULFA (Surrey University Library for Forensic Analysis) library of video [16]. The second dataset includes 13 videos with thirty frame rate collected from the TRACE video library [17]. The last video dataset for testing they collect from the Youtube [18] with high resolutions. They demonstrate false positives are reduced and outperformed the accuracy as compared to the existing approaches using an adequate level of detection using their approach. They also state that their approach works effectively for both static and dynamic video data. They registered 97% F1 score, 98% precision, and 97%Recall when comparing their results with state-of-the-art methods. In the future work directions, they will work on the enhancement of their approach and focus on various form of video forgery.

Deep learning approaches are now widely adopted technique in various fields for the classification of multiple tasks based on several purposes. Nowadays, video forensics is also one of the areas which choose deep neural network as an attractive technique to classify the irregular alterations in the videos utilizing the vital features. The authors in [19] also use deep learning approach to detect an inter-frame forgery in the videos. They use a twelve layer configuration deep convolutional neural network (DCNN) as a deep learning method and calculate the correlation between the frames to classify the forgery in the video data. They use a video frame of 114 width and 114 height with three color RGB channels. They use two dataset Reverse Engineering of Audio-Visual Content Data (REWIND) [20] and Image Processing Research Group (GRIP) [21] to train their DCNN model. The authors registered their experimental results 98% in terms of average accuracy. They also test their approach to check the effectiveness with the compressed Youtube videos dataset, which is the part of GRIP dataset.

The authors in [22] describe a method to detect a forgery in the digital video data taken from the surveillance and mobile cameras. Their plan is applicable for inter-frame video forgery that is frame insertion, deletion and duplication. They detect the irregularity of frames in the H.264 and MPEG-2 encoded videos with the help of gradient calculation of prediction residual and optical flow. They also claim that their approach detects and localizes the tempering in the video sequences and does not depend on the numbers of frames tempered. The authors' inspections show that their approach does not perform well for videos with high brightness. The results obtained by the authors are 83% and 80% average accuracy for detecting and localizing the forgery, respectively.

There are a few approaches which detect duplication forgeries in frames and region, in [23], the author's also present two algorithms to detect frame and region duplication forgery respectively. In the first algorithm, the authors identify the frame forgery in the video for three modes by utilizing the correlation between the frame sequences in the video and obtain the mean of each video. The three modes in the first algorithm to detect frame duplication are the distribution of continuous video frames duplication at large continue position, the distribution of video frames duplication having a various length at various locations, and last, frames from different videos duplication holding similar and diverse dimensions. In the second algorithm, the authors distinguish the region duplication in videos using the relationships between the regions of two/affected frames within the likewise frame at various locations and from another frame to various sequential frame of the same video at a similar place. The authors register the average accuracy of 99.5% for the first algorithm (frame duplication forgery detection) with the dataset SULFA [16] and the average accuracy of 96.6% for the second algorithm (region duplication forgery detection) with the same dataset. The limitation of their first algorithm to detect forged frames in a video is that the authors cannot identify the smaller number of duplicated frames that estimated for the video. And the limitation for the second algorithm to determine the forged region is that their approach cannot identify the most modest part for the duplication of region forgery in the video.

3 A Sketch of Proposed CNN for Deep Feature Extraction

In this section, we explain the architecture of CNN proposed for the extraction of deep features from the frames in the video. We use VGG16 [24] as a baseline model for feature extraction. But we have modified the existing model of VGG16 for feature extractions because we are computing cross-correlation features to measure the similarity between two frames. Therefore we have used stride as 2 in convolution layers. By using stride in the VGG16 model layer output size will minimize therefore we have used only 4 convolution layers to uphold the feature output of the model. After applying the discussed changes in the VGG16 model our modified proposed CNN architecture process the frames faster as it contains a lesser number of layers to follow. The proposed parallel CNN takes every two consecutive frames as an input to extract the deep features of each frame. These deep features in the next phase are used to detect the inter-frame video forgery. The complete configuration of the proposed CNN is displayed in Tab. 1. This work extracts 8192 in depth features for each frame as an output. There are four convolutional layers, a flatten layer, and the fully connected layer present in the CNN configuration. Fig. 1 shows the graphical representation of the proposed CNN architecture.

Table 1: The configuration of CNN for deep feature extraction in video forgery

S.No.	Type of layer	Function activation	Shape of output	Kernel size	Number of filters	Number of strides	Weights
0	Input	-	(None, 250, 250, 1)	-	-	-	-
1	Conv2D	ReLU	(None, 124, 124, 16)	3	16	2	160
2	Conv2D	ReLU	(None, 61, 61, 16)	3	16	2	2320
3	MaxPool		(None, 20, 20, 16)	3			-
4	Conv2D	ReLU	(None, 18, 18, 32)	3	32		4640

(Continued)

Table 1: Continued

S.No.	Type of layer	Function activation	Shape of output	Kernel size	Number of filters	Number of strides	Weights
5	Conv2D	ReLU	(None, 16, 16, 32)	3	32		9248
6	Flatten	-	(None, 8192)	-	-	-	0

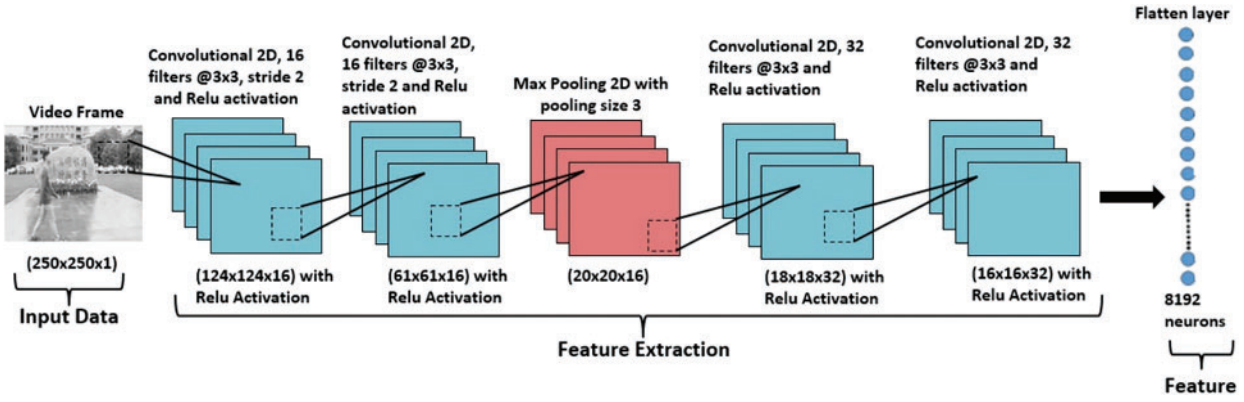

Figure 1: CNN architecture diagram

This work uses these deep features from the nearby frames to find the correlation coefficient. If the time interim of the adjacent frames is small, then the correlation is high, else vice-versa.

4 Proposed Methodology

The complete working architecture of the proposed work is displayed in Fig. 2. This section discusses the proposed methodology of the work into phases which consist of four basic working blocks which are as follows:

1) Inter frame relation block
2) Parallel deep neural network block
3) Adaptive thresholds block
4) Forgery detection block

4.1 Inter Frame Relation Block

The proposed model uses preprocessing technique where we perform the frame extraction operation. Which uses the OpenCV library [25] to extract the frames from the input video as OpenCV supports maximum video format. Hence approach is generic for as many video formats and compression techniques as OpenCV support. In the initial phase, this work takes a video as an input, and supplies into the inter frame relation block. In this block, input video is preprocessed to extract the RGB frames from the video and converted into the grayscale representation. To reduce the storage space necessities approx three order RGB to grayscale conversion [26] is performed. Finally, in this block, each frame is transformed into an adjacent matrix of order m × n (250 × 250) as an output and input for the next block that is parallel deep neural network block.

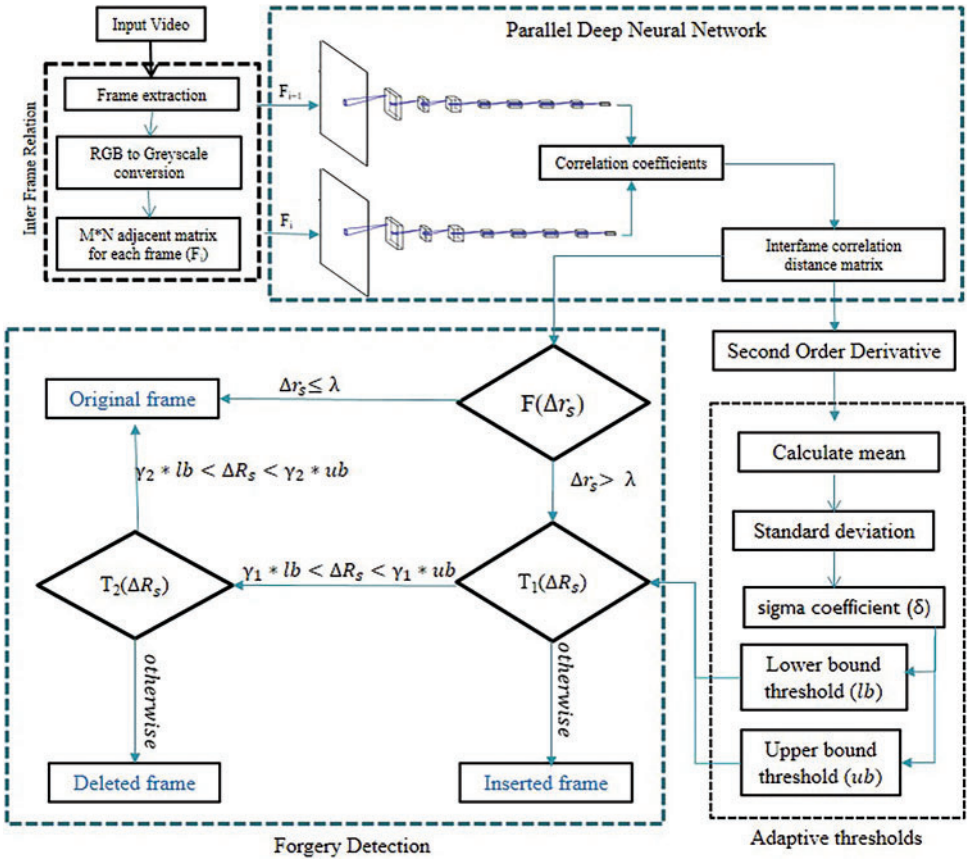

Figure 2: Proposed framework architecture

4.2 Inter Frame Relation Block

In this block, CNN model is used parallel to process the two consecutive frames and produce an output as deep features. The configuration of the CNN model to extract the deep features is explained in Section 3. These deep features from the two continuous frames help to find the correlation coefficients between them. After calculating correlation coefficients, this work computes the difference of the adjacent correlation coefficient (Δr_s) of frames as the inter-frame correlation distance matrix. The output of this block is passed to the forgery detection block to calculate the elimination function $F(\Delta r_s)$ and to find the second-order derivative which helps to obtain the upper bound and lower bound threshold in the adaptive thresholds blocks. There are two primary procedures performed in this block which are defined as follows:

- First, determine the correlation coefficient using the deep features extracted within two neighbouring frames coming from the parallel CNN models. When comparing with the Pearson's correlation coefficient [27] formula for the frames, this work uses the single dimension as a deep feature list rather than two dimensions (row and column) in the frames.
- Second, after finding the correlation coefficient within the two deep feature list for the adjacent frames, this work computes the inter-frame correlation distance (Δr_s). Eq. (1) defines the formula to calculate the Δr_s from the correlation coefficients calculated in the previous step. The Eq. (1) helps to get the content variability in the video sequence. Whenever any tampering

is done in the video sequences, the variability affected and does not remain constant, on the other side in the original video where no tampering is performed; it remains stable.

$$\Delta r_s = \begin{Bmatrix} |r_s - r_{s+1}|, & s \geq 2 \\ 0, & s = 1 \end{Bmatrix} \tag{1}$$

After calculating the correlation coefficient difference (Δr_s) this work calculate the second order derivative of the Δr_s sequences obtained from the input video. The second order derivative helps to highlight the portions of rapid changes in the video sequences. Let the sequence of Δr_s is obtained from the input video is $\Delta r_1, \Delta r_2, \ldots, \Delta r_n$. Now this work get the series for second order derivative (ΔR_s) from the sequence of Δr_s. The formula for calculating ΔR_s is present in the Eq. (2).

$$\Delta R_s = \frac{\partial^2 \Delta r}{\partial x^2} = \Delta r_{x+1} + \Delta r_{x-1} - 2\Delta r_x \tag{2}$$

where $2 \geq x \leq (n-1)$, n is the total points in the Δr_s. Δr_x is the current point where processing is being done, and Δr_{x+1}, Δr_{x-1} are the neighboring points.

4.3 Adaptive Thresholds Block

In the adaptive thresholds block the calculated pieces of information like mean, standard deviation, and sigma coefficient (δ) are used to find the supporting values (upper bound and lower bound thresholds) to detect the inter frame forgery. We calculate the mean, standard deviation and use the least distance score to eliminate the original and common consecutive frames. And we also calculate distance and second-order derivative so that every feature represents up to 5 consecutive frames to reduce the effect frame duration. At the same time, we do not use a fixed window size, which can be generic for frame duration effects in the video. Thus, if forgery present at some location in high frame rate videos and the window size is small, with this respect, the forgery will not be detected as the standard deviation of that window size considered all forged frames into original frames. In high frame rate videos, consecutive frames have common features. Suppose there is any insertion forgery in the video. In that case, the mean of the second-order derivative of the correlation coefficient difference will lead to near zero. Still, values near insertion frames if far from zero, which can be detected using standard deviation by γ_1 & γ_2. If we use fixed window size, then mean can be calculated for insertion frames; therefore, mean and insertion frames values are close to each other. All the values lie between y1. Therefore, it makes it harder to find the insertion type forgery in the video having a fixed small window size. The output of the adaptive thresholds block is supplied to the forgery detection block to check the condition of two threshold functions $T_1(\Delta R_s)$ & $T_2(\Delta R_s)$.

The major objective of this block is to find the upper and lower bound threshold values. These value with few intrinsic value are checked to detect the forgery in the video. There are few steps to calculate the upper bound threshold (ub) and the lower bound threshold (lb) and the steps are as states:

- First, we calculate the mean value [28] of the second order derivative ΔR_s, which is obtained from the correlation coefficient difference in the previous block. Mean helps to find the behavior of a few abnormally large or abnormally low values. The Eq. (3) shows the formula for calculating the mean $\overline{\Delta R_s}$.

$$\overline{\Delta R_s} = \frac{\sum \Delta R_s}{n} \tag{3}$$

where n is the number of values present in the ΔR_s list.

- Second, we find the standard deviation (σ_s) for the second order derivative of the correlation coefficient distance. It helps to show data deviation from the mean value. The Eq. (4) shows the formula for calculating the σ_s.

$$\sigma_S = \sqrt{\frac{\sum \left(\Delta R_s - \overline{\Delta R_s}\right)^2}{n-1}} \tag{4}$$

where n is the number of values present in the ΔR_s list.

- In the next step, we calculate the sigma coefficient (δ) value for the ΔR_s series with the help of the mean $\overline{\Delta R_s}$ and standard deviation σ_s by using the traditional method used by the community.
- At last to find the upper bound threshold ub and lower bound threshold lb we need to find the specific value of the sigma coefficient δ as shown in the Eq. (5) to calculate the value of ub and lb. So we have tested for the various value of sigma coefficient and plot the graph to see the optimal value that can be used to find the ub and lb. The value is taken for δ is 2.6 as we can see in Fig. 3, it gives the highest result in the value used.

$$(\text{ub}, \text{lb}) = \overline{\Delta R_s} \pm (\delta * \sigma s) \tag{5}$$

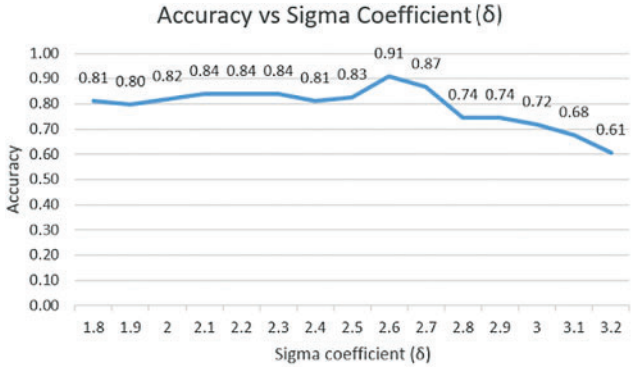

Figure 3: Sigma coefficient *vs.* accuracy graph plot

4.4 Forgery Detection Block

After finding all the values from the previous blocks, forgery detection block looks for elimination function $F(\Delta r_s)$ and two threshold functions $T_1(\Delta r_s)$ & $T_2(\Delta r_s)$ to find the category of the forgery such as frame insertion and frame deletion based on the conditions. Forgery detection block is the main block where we detect the tampering in the input video. First, we detect whether the input video is forged/tampered or the original video, so we call this phase as video forgery detection phase. In the second phase, we identify the type of forgery like frames inserted or deleted in or from the original video. We name this second phase to video forgery classification.

- Video forgery detection phase: In this phase, we define an elimination function $F(\Delta r_s)$ which takes input as correlation coefficient distance Δr_s using the least distance score which eliminate the most original frame from the input video. Hence distance should be negligible or near to zero value because zero distance represent that both are the same frame. Therefore the value of λ is taken near to zero that is 0.1. Finally, if the correlation coefficient distance is less than or equal to the value of λ then the frame is original else forgery is detected in the frame and start the process for the classification in the next phase. The formula to calculate the elimination

function is shown in Eq. (6).

$$F(\Delta r_s) = \begin{Bmatrix} \boldsymbol{T_1}\ (\boldsymbol{\Delta R_s}), & \Delta \boldsymbol{r_s} > \lambda \\ \boldsymbol{O}(\boldsymbol{s}), & \Delta \boldsymbol{r_s} \leq \lambda \end{Bmatrix} \tag{6}$$

where O(s) indicates the original frame if the correlation coefficient distance is less than the taken value of λ.

- Video forgery classification phase: In this phase, we set the two threshold functions $T_1(\Delta R_s)$ and $T_2(\Delta R_s)$ based on the relation and condition of second order derivative of correlation coefficient distance with the upper and lower bound threshold and threshold control parameters (γ_1 & γ_2). Eqs. (7) and (8) explain the threshold functions with their conditions to classify the forgery whether frame is inserted or deleted in or from the original video.

$$T_1(\Delta R_s) = \begin{Bmatrix} \boldsymbol{T_2}\ (\boldsymbol{\Delta R_s}), & \boldsymbol{\gamma_1} * lb < \Delta\, \boldsymbol{R_s} < \boldsymbol{\gamma_1} * ub \\ \boldsymbol{I}(\boldsymbol{s}), & \boldsymbol{otherwise} \end{Bmatrix} \tag{7}$$

$$T_2(\Delta R_s) = \begin{Bmatrix} \boldsymbol{O}(\boldsymbol{s}), & \boldsymbol{\gamma_2} * lb < \Delta \boldsymbol{R_s} < \boldsymbol{\gamma_2} * ub \\ \boldsymbol{D}(\boldsymbol{s}), & \boldsymbol{otherwise} \end{Bmatrix} \tag{8}$$

Here γ_1 and γ_2 are the threshold control parameter which is used as 1.4, 1 respectively to distinguish the frames into the inserted frame I(s), deleted frame D(s), and original frame O(s).

Eq. (7) explain if the value of second order derivative computed from the correlation coefficient distance ΔR_s is lying between the upper bound and lower bound threshold with threshold control parameter product γ_1, then the second threshold function is checked given in the Eq. (8). Else insertion forgery is present in the input video. The Eq. (8) depicts that if the value of ΔR_s lies between the product of γ_2 with the upper and lower threshold, then the frames are original no forgery is present else deletion of frame forgery is detected. The value for γ_2 is selected as 1. On the other hand boundaries of threshold parameter γ_1 discriminate the forgery type such as deletion and insertion. So the value of γ_1 is computed empirically as shown in the Fig. 4, and 1.4 value is selected for γ_1 as we find the best results for this value in both dataset. Accuracy achieved on multiple forgery is 82% on VIFFD dataset and 86% on TDTV dataset.

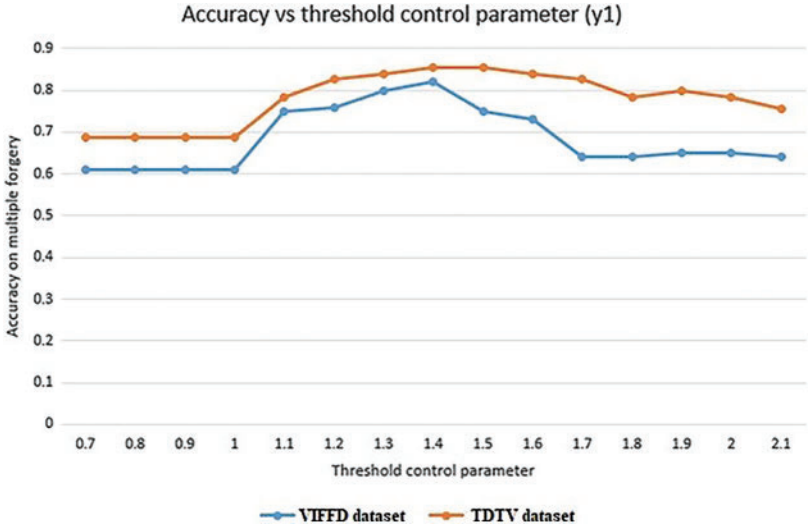

Figure 4: Accuracy on multiple forgery *vs.* threshold control parameter graph

4.5 Dataset

This work selects the VIFFD dataset [29,30] and TDTV dataset [31] for the evaluation of the performance and validation of the proposed work. This work test videos of different frame rate for both the dataset. This work considers that there is a high chance of no significant change in consecutive frames while having a higher frame rate. Therefore, we define an elimination function $F(\Delta r_s)$, which takes input as correlation coefficient distance and Δr_s, eliminating the most original frame (the common consecutive frames) from the input video using the least distance score. Tab. 2 shows the details about the dataset used in this work.

Table 2: Dataset used for experiment

Dataset	Number of videos	Resolution	Length (Average)	Size of dataset	Bitrate	Fps
VIFFD dataset	90 (30 original 60 forged)	$720 * 404$	10.1 s	4.47 GB	4318 kbps	25 fps
TDTV dataset	156 (16 original 140 forged)	$320 * 240$ or 640×360	12.0 s	5.03 GB	3439 kbps	30 fps

4.5.1 VIFFD Dataset

There are in total 90 videos are present in the dataset for the validation purpose. The ratio of the dataset is 1:2 that consists of 30 original and 60 forged video. Also, further classification of the forged video is in the proportion of 1:1 having 30 videos in each category deleted and inserted forgery.

4.5.2 TDTV Dataset

There are in total 156 videos are present in the dataset for the validation purpose. dataset is labelled into original and forged video, where 16 videos are original and 140 videos are forged. Also, further classification of the forged video is in the proportion of 1:1 having 70 videos in each category deleted and inserted forgery.

5 Experimental Results and Analysis of Proposed Method

In this section, experiments are performed into two phases, as explained in subsection 4-D. This work test the proposed approach for the original and forged videos. All the experiments are performed with NVIDIA DGX-2 server, 2 petaFLOPS performance, 2x Intel Xenon Platinum CPUs, 16 NVIDIA Tesla V100 32GB HBM2, 1.5 TB DDR4 RAM, 30TB SSDs, 8 x 100 Gb/sec network speed. To validate the proposed methods, we use various metrics such as accuracy, precision, Recall, and F1-Score. For the evaluation of the experiments, the confusion matrix is presented in both the phases. In the video forgery detection phase original video is labelled as a negative class and forged video as the positive class. In the video forgery classification phase, where forgery is divided into two categories like insertion or deletion forgery multi-class classification is done. Formulas for the different metrics [32] are presented in the like precision, recall, f1score, accuracy.

Where TP stands for true positive means, forged video is predicted as forged video. FP attains for false-positive, which refers to the original video is classified as forged video. FN stands for false-negative that is a forged video is detected as the original video. And the TN stands for true negative means the original video is predicted as the original video.

- Video forgery detection from the forged and original video: The first experiment is performed by taking a random sample video from the original video dataset. Fig. 5 shows the frames extracted from the original video. One can observe in Fig. 5 the movement of the person is from right to left. Fig. 6 shows the graph plot of relatively constant correlation coefficients of original frames from the left and in the center second order derivative of correlation coefficient distance which is near to zero value from the first frame to 10 frames. There is a deviation in the last frames means forgery possibility is present. In the rightmost part, where threshold parameters with an upper and lower bound threshold are compared then all the values lie within the boundaries. So this work concludes that the video is original. The same operation is performed on the forged videos when the video is predicted as the tampered video, then it goes to the next phase to detect the type of forgeries such as insertion and deletion forgery. Authors have considered only frame insertion and deletion type of editing/forgery in this experiment. The experimental results for the video forgery detection phase are presented in Tabs. 3 and 4. This work achieves the accuracy of 91% in VIFFD dataset and 90% in TDTV dataset for the detection of video forgery.
- Video forgery classification into insertion or deletion of frames for the input video: For the next experiment, we choose a forged video from the dataset where insertion frame forgery is present. Fig. 7 shows the frames extracted from the sample forged video. We can observe from Fig. 7 that a player wearing a red t-shirt is playing basketball in the sample video. But in the left side frame, another player wearing the black t-shirt can be seen that frame is the inserted frame in the video. Fig. 8 shows the graph plot of correlation coefficients of forged frames from the left and in the center second order derivative of correlation coefficient distance which is near to zero value from the first frame to near 24th frame. There is a large deviation from the 24 to 40 frame number that signifies that forgery possibility is present. In the rightmost part, where threshold parameters with an upper and lower bound threshold are compared then all the values lie within the boundaries except the frames 24 to 40. In the range of 24 to 40 frames highly deviation is found and exceeded from the $\gamma 1$ threshold control parameter and lower bound and satisfy the Eq. (7) this concludes the insertion forgery present in the input video. Fig. 9 shows a few frames of the sample video. We can observe from Fig. 9 that the momentum of the person who is riding on cycle is tiny in the first 0.4 s but in the right the last frame of 0.8 s it is a huge difference. So the possibility is that the frames are deleted from the sample video. Fig. 10 shows the graph of the correlation coefficients from the forged frames in the left. In the middle the second order derivative of correlation coefficient distance which has zero value from the first frame till to 12 frame number that means frames are highly coupled no chance of forgery in this range. But there is a large deviation from the 12th to 16th frames that implies that forgery opportunity is present. In the rightmost part, where threshold parameters with an upper and lower bound threshold are matched then all the values lie within the boundaries except the frames 12 to 16. In this range frames are highly deviated and exceeded from the $\gamma 2$ threshold control parameter and lower bound and satisfy the Eq. (8) this concludes the deletion forgery present in the input video. After doing all the experiments for original, insertion, and deletion forgery and validating proposed methodology the video forgery classification phase achieves the accuracy of 82% in VIFFD dataset and 86% in TDTV dataset. The confusion matrix and experimental results for this phase are presented in Tabs. 5 and 6 respectively.

Figure 5: The sample frames of the original video

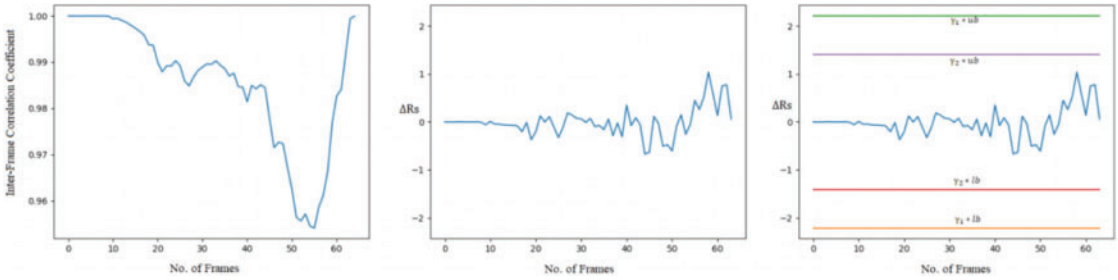

Figure 6: Representation of correlation coefficients, second order derivative of correlation coefficients distance, and with dual threshold of original video from left to right

Table 3: Confusion matrix of original and forged video detection

	VIFFD Dataset			TDTV Dataset		
Original	**27**	**3**		Original	**12**	**4**
Forged	**4**	**56**		Forged	**11**	**129**
	Original	Forged			Original	Forged

Table 4: Experimental results for original and forged video detection

	VIFFD Dataset					TDTV Dataset			
	Precision (%)	Recall (%)	F1score (%)	Support		Precision (%)	Recall (%)	F1score (%)	Support
Original	87	90	88	30	Original	52	75	61	16
Forged	95	93	94	60	Forged	97	92	94	140
Accuracy			91			Accuracy		90	

Figure 7: The sample frames of the insertion forgery video

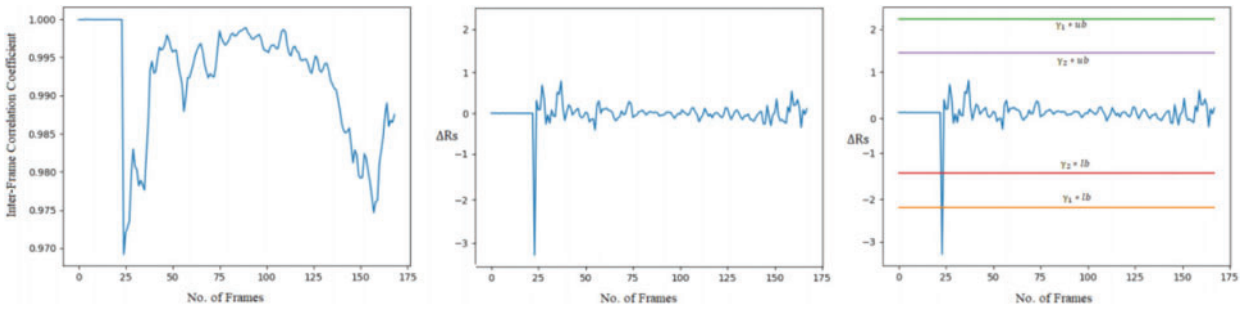

Figure 8: Representation of correlation coefficients, second order derivative of correlation coefficients distance, and with dual threshold of insertion forged video from left to right

Figure 9: The sample frames of the deletion forgery video in VIFFD dataset

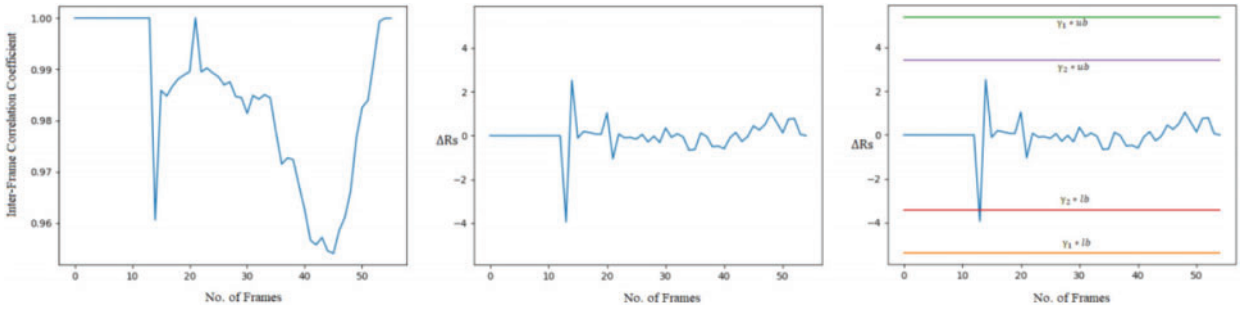

Figure 10: Representation of correlation coefficients, second order derivative of correlation coefficients distance, and with dual threshold of deleted forged video from left to right

Table 5: Confusion matrix of multiple forgery detection

VIFFD Dataset				TDTV Dataset			
Original	**27**	**3**	**0**	Original	**12**	**4**	**0**
Deleted	**3**	**21**	**6**	Deleted	**11**	**59**	**0**
Inserted	**1**	**3**	**26**	Inserted	**0**	**6**	**64**
	Original	Deleted	Inserted		Original	Deleted	Inserted

Table 6: Experimental results for video forgery classification phase

VIFFD Dataset					TDTV Dataset				
	Precision (%)	Recall (%)	F1score (%)	Support		Precision (%)	Recall (%)	F1score (%)	Support
Original	87	90	88	30	Original	52	75	61	16
Deletion	78	70	74	30	Deletion	86	84	85	70
Insertion	81	87	84	30	Insertion	100	91	95	70
Accuracy			82		Accuracy			86	

Tab. 7 represents the results for original, forged, frame deletion, and frame insertion forgery detection on the VIFFD video dataset with the various metrics. The total accuracy of 90% is achieved for the original video, maximum accuracy of 93% is obtained for the forged video type. The accuracy of 70% and 87% is achieved for the frame deletion and frame insertion forgery for the VIFFD dataset using the proposed methods. All the result are compared with TDTV dataset. The total accuracy of 75% is achieved for the original video, this is because there are very few samples available of that video type which may result in an unbalanced dataset. If the dataset is balanced then the proposed model can achieve higher accuracy, maximum accuracy of 92% is obtained for the forged video type. The accuracy of 84% and 91% is achieved for the frame deletion and frame insertion forgery for the TDTV dataset using the proposed methods.

Table 7: Final combined experimental result

VIFFD Dataset					
Video type	Precision (%)	Recall (%)	F1score (%)	Support	Accuracy (%)
Original	87	90	88	30	90
Forged	95	93	94	60	93
Frame deletion	78	70	74	30	70
Frame insertion	81	87	84	30	87
TDTV Dataset					
Video type	Precision (%)	Recall (%)	F1score (%)	Support	Accuracy (%)
Original	52	75	61	16	75
Forged	97	92	94	140	92
Frame deletion	86	84	85	70	84
Frame insertion	100	91	95	70	91

Table 8: Comparison with state-of-art

Forgery type	Dataset	Method	Precision(%)	Recall(%)	F1score(%)
Insertion	SULFA + LASIESTA [33] + IVYLAB [35]	fadl et al. [34]	93.51	90	91.72
	VTD	bakas et al. [36]	91.79	93.72	92.74
	VIFFD	Proposed	81	87	84
	TDTV	Proposed	100	91	95
Deletion	SULFA + LASIESTA [33] + IVYLAB [35]	fadl et al. [34]	78.51	79.16	78.33
	VTD	bakas et al. [36]	88.64	84.98	86.77
	VIFFD	Proposed	78	70	74
	TDTV	Proposed	86	84	85

In this Tab. 8 we have compare the proposed method with the recent state-of-art in various terms for insertion and deletion type forgery detection. The Fadl et al. [34] proposed the universal image quality index based video authentication. The Bakas et al. [36] proposed prediction footprint variations pattern technique. The work is compared with TDTV dataset as its combines many dataset which is very similar to Bakas dataset. This work is based on single-shot videos. Therefore, any transition from one shot to another shot is considered as insertion as per the proposed algorithm. As transitions can be another type of forgery and there are various types of transitions, authors have not considered the effect of transition in the video. However, abrupt transitions should be detected as insertion forgery by our proposed method as through our test results. At the same time, the gradual transition is not considered inside this algorithm. Therefore, it might be detected as insertion or original with the proposed method.

6 Conclusion

In this work, we introduce an approach to detect and classify the inter-frame video forgery such as frame insertion and deletion. The proposed work comprises of four working blocks. The first block has the basic functionality of converting an input video to frames and matrix form for each frame. The second block operations include deep feature extraction using two parallel CNN models and calculation of correlation between the deep features and correlation difference to find the relation between the frames. Threshold parameters, upper and lower bound threshold values are obtained in the third block. These threshold values help to find the forgery in the video. The last block checks the elimination and threshold functions using threshold values obtained from the previous blocks to detect the forgery and originality for the video. The experimental analysis of forged video and original video with the proposed method shows that our approach performs well for the detection and classification of the video forgery. Our method is generic enough because our model is capable of detecting and classifying forgery for multiple dataset. In future work, we enhance the performance of our proposed work for video classification and time-based analysis would be done.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Orben and A. K. Przybylski, "The association between adolescent well-being and digital technology use," *Nature Human Behaviour*, vol. 3, no. 2, pp. 173–182, 2019.

[2] R. D. Singh and N. Aggarwal, "Video content authentication techniques: A comprehensive survey," *Multimedia Systems*, vol. 24, no. 2, pp. 240, 2018.

[3] D. Liu, J. Shen, P. Vijayakumar, A. Wang and T. Zhou, "Efficient data integrity auditing with corrupted data recovery for edge computing in enterprise multimedia security," *Multimedia Tools and Applications*, vol. 79, no. 15, pp. 10851–10870, 2020.

[4] Y. Liu and T. Huang, "Exposing video inter-frame forgery by zernike opponent chromaticity moments and coarseness analysis," *Multimedia Systems*, vol. 23, no. 2, pp. 223–238, 2017.

[5] K. Sitara and B. M. Mehtre, "Digital video tampering detection: An overview of passive techniques," *Digital Investigation*, vol. 18, pp. 8–22, 2016.

[6] V. Kumar, A. Singh and M. Gaur, "A Comprehensive Analysis on Video Forgery Detection Techniques," in *Proc. of the Int. Conference on Innovative Computing & Communications (ICICC)*, India. SSRN, 2020.

[7] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "Mesonet: A compact facial video forgery detection network," in *2018 IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Hong Kong, IEEE, pp. 1–7, 2018.

[8] J. Chao, X. Jiang and T. Sun, "A novel video inter-frame forgery model detection scheme based on optical flow consistency," in *Int. Workshop on Digital Watermarking*, China, Springer, pp. 267–281, 2012.

[9] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, "Deep learning," volume 1 MIT Press Cambridge, Cambridge, 2016.

[10] Z. Ding, M. Zhu, V. W. Tam, G. Yi and C. N. Tran, "A system dynamics-based environmental benefit assessment model of construction waste reduction management at the design and construction stages," *Journal of Cleaner Production*, vol. 176, pp. 676–692, 2018.

[11] 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, 2016.

[12] R. D. Singh and N. Aggarwal, "Detection and localization of copy-paste forgeries in digital videos," *Forensic Science International*, vol. 281, pp. 75–91, 2017.

[13] C. C. Hsu, T. Y. Hung, C. W. Lin and C. T. Hsu, "Video forgery detection using correlation of noise residue," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, Australia, IEEE, pp. 170–174, 2008.

[14] P. Ferrara, T. Bianchi, A. De Rosa and A. Piva, "Image forgery localization via fine-grained analysis of CFA artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1566–1577, 2012.

[15] J. Bakas, R. Naskar and R. 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, no. 4, pp. 4905–4935, 2019.

[16] G. Qadir, S. Yahaya and A. T. Ho, "Surrey university library for forensic analysis (SULFA) of video content," *IET Conference on Image Processing (IPR 2012)*, London, pp. 1–6, 2012. http://dx.doi.org/10.1049/cp.2012.0422.

[17] A. Pulipaka, P. Seeling, M. Reisslein and L. J. Karam, "Traffic and statistical multiplexing characterization of 3-D video representation formats," *IEEE Transactions on Broadcasting*, vol. 59, no. 2, pp. 382–389, 2013.

[18] Use-IP Ltd, "Hikvision 4 K DS-2CD4A85F-I sample footage (Day and night)," 2021, [Online]. Available: https://www.youtube.com/watch?v=66Ob1aJedHc&t=14s, 2016.

[19] H. Kaur and N. Jindal, "Deep convolutional neural network for graphics forgery detection in video," *Wireless Personal Communications*, vol. 112, pp. 1–19, 2020.

[20] P. Bestagini, S. Milani, M. Tagliasacchi and S. Tubaro, "Local tampering detection in video sequences," *2013 IEEE 15th Int. Workshop on Multimedia Signal Processing (MMSP)*, Pula, Italy, 2013.
[21] D. Cozzolino, P. Giovanni and V. Luisa, "Efficient dense-field copy–move forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, 2015.
[22] S. Kingra, N. Aggarwal and R. D. Singh, "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.
[23] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11527–11562, 2019.
[24] H. Qassim, A. Verma and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conf. (CCWC)*, Las Vegas, IEEE, pp. 169–175, 2018.
[25] G. Bradski Reading and Writing Images and Video, 2021, [Online]. Available: https://docs.opencv.org/2.4/modules/highgui/doc/reading_and_writing_images_and_video.html#videocapture, 2021.
[26] T. Kumar and K. Verma, "A theory based on conversion of rgb image to gray image," *International Journal of Computer Applications*, vol. 7, no. 2, pp. 7–10, 2010.
[27] J. Benesty, J. Chen, Y. Huang and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, Springer, pp. 1–4, 2009.
[28] S. C. Tsiang, "The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money," *The American Economic Review*, vol. 62, no. 3, pp. 354–371, 1972.
[29] X. H. Nguyen, "VIFFD-A dataset for detecting video inter-frame forgeries," 2020.
[30] X. H. Nguyen, Y. Hu, M. A. Amin, G. H. Khan and D. -T. Truong, "Detecting video inter-frame forgeries based on convolutional neural network model," *International Journal of Image, Graphics and Signal Processing*, vol. 12, no. 3, pp. 1, 2020.
[31] H. D. Panchal and H. B. Shah, "Video tampering dataset development in temporal domain for video forgery authentication," *Multimedia Tools and Applications*, vol. 79, no. 33, pp. 24553–24577, 2020.
[32] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv preprint arXiv: 2010.16061, 2020.
[33] C. Cuevas, E. M. Yáñez and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: Lasiesta," *Computer Vision and Image Understanding*, vol. 152, pp. 103–117, 2016.
[34] S. Fadl, Q. Han and Q. Li, "Surveillance video authentication using universal image quality index of temporal average," in *Int. Workshop on Digital Watermarking*, Korea, Springer, pp. 337–350, 2018.
[35] H. Sohn, W. D. Neve and Y. M. Ro, "Privacy protection in video surveillance systems: Analysis of subband-adaptive scrambling in jpeg xr," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 170–177, 2011.
[36] J. Bakas, R. Naskar and S. Bakshi, "Detection and localization of inter-frame forgeries in videos based on macroblock variation and motion vector analysis," *Computers & Electrical Engineering*, vol. 89, pp. 106929, 2021.