An efficient novel dual deep network architecture for video forgery detection

Chandrakala, Mungamuri Sasikala
Department of Computer Science Engineering, Godutai College of Education for Women, Kalaburagi, India

Article Info

Article history:

Received Dec 23, 2022
Revised Sep 5, 2023
Accepted Sep 17, 2023

Keywords:

Deep learning
Digital video
Dual deep network
VCMFD
Video forgery

ABSTRACT

The technique of video copy-move forgery (CMF) is commonly employed in various industries; digital videography is regularly used as the foundation for vital graphic evidence that may be modified using the aforementioned method. Recently in the past few decades, forgery in digital images is detected via machine intellect. The second issue includes continuous allocation of parallel frames having relevant backgrounds erroneously results in false implications, detected as CMF regions third include as the CMF is divided into inter-frame or intra-frame forgeries to detect video copy is not possible by most of the existing methods. Thus, this research presents the dual deep network (DDN) for efficient and effective video copy-move forgery detection (VCMFD); DDN comprises two networks; the first detection network (DetNet1) extracts the general deep features and second detection network (DetNet2) extracts the custom deep features; both the network are interconnected as the output of DetNet1 is given to DetNet2. Furthermore, a novel algorithm is introduced for forged frame detection and optimization of the falsely detected frame. DDN is evaluated considering the two benchmark datasets REWIND and video tampering dataset (VTD) considering different metrics; furthermore, evaluation is carried through comparing the recent existing model. DDN outperforms the existing model in terms of various metrics.

Corresponding Author:

Chandrakala
Department of Computer Science Engineering, Godutai College of Education for Women
Kalaburagi, Karnataka, India
Email: chandrakalavpatil15@gmail.com

1. INTRODUCTION

Digital videography processing software like Photoshop, Adobe Premiere, and Final Cut Pro used for rapid growth and development of widespread images and video-processing software, results in tampering with the original video without retaining any obvious traces. The malicious tampering of the videos results in serious legal and social issues. By considering an example, the tampered videos and images may serve as evidence to present in the court, which may deviate the truth from the public in news reports. As multimedia content is growing extensively, it makes it a tedious task to detect tampered video content caused by human insight, because video manipulation is common these days, academics have recently concentrated their efforts on video forensics. This is because video data is being upgraded quickly [1]. To this, several potential alterations are applied such as deleting the frame, inserting the frame, and compressing the video. Necessarily the digital forensic techniques are distinguished into the active and passive approaches. However, most of these passive forensic methods are allotted for analyzing still images [2]. Recently, the research focus is provided on video forensics, because video tampering becomes easy with each passing day. Among these, copy-move forgery (CMF) is extended to hide particular objects in the same video in contradiction with similar techniques. Hence, the frames are retrieved from similar video sequences based on their operational functions, which is convenient for operating and complex to distinguish [3]. Regional forgery and frame cloning are two categories used to classify video-copy motion forgeries based on various operational domains. Similarly, to the image copy-move mechanism, regional copy-move causes alterations for specific portions of frame images seen in the more mature images. Table 1 shows the original image and CMF image.

Video copy-move forgery (VCMF) produces homogenous information and intricate modifications without different forgery traces. The VCMF is differentiated into three types: interframe, intra-frame, and hybrid known as inter/intra frame. Video intra-frame forgery is comparable to CMF modification, which pastes the items copied in a similar frame. Inter-frame movie forgery copies and pastes the contents of the objects in concurrent frames in the same video. The items in the movie are further divided into additive and occlusive classes in the context of modification. The items targeted are added up in additive forging. Occlusive forging summarizes the background information, covering the target material as a result. The video content is plagiarized in line with their clip examples [4].

VCMF is classified into two categories: the first category includes frame cloning and regional forgery; Similar to copy-image estimated by a mature image, regional-CMF modifies specific portions of the frame. The imperceptibility and difficulty estimation frame CMF enhance the pasting and cloning of subsequent frames in a frame known as CMF, which results in ineffective colour changes, shooting parameters, and illuminating conditions [5]. Leads to an anomaly in the parameter distribution, which leads to the correlation of original and duplicated frames. Various methods are designed to detect frames and CMF is classified into two groups i.e. video-based and image-based. The algorithms used in the image feature exploit and extract each frame to detect correlation; this includes the detection of categories of grey values, image texture detection, noise features and colour modes. Different types of feature extraction techniques are applied to identify films using their distinctive motion features, when the video is combined with coding features, the copy-move operation creates a disadvantage [6]; moreover Figure 1 shows the CMF illustration.

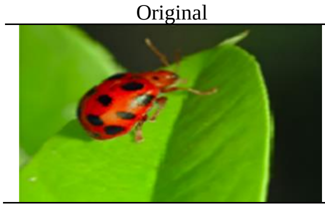
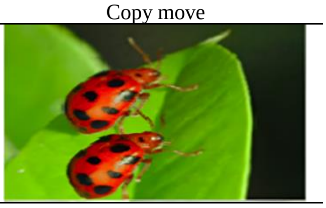
Video tampering is increasing each day; however, a few digital videography contents have been discovered, this occurrence has worn-out public interest in digitalized content videography clips. The main aim of video tampering detection here ensures authenticating the potential modifications and forgeries i.e needlessly checking whether a specified clip is tampered with or not. The forged area within a frame and its adjacent frame indicates position of frame insertion, replacement, ordering, and deletion of a tampered video. Various approaches are proposed that authenticate and localize tampering necessarily in the images [5], [6]. However, these techniques are not applied directly to the videos for the following reasons: i) due to the presence of the enormous amount of data the storage transmission is compressed before the videos are encoded into video frames, ii) the techniques reported here apply to video sequences that generate a huge amount of computational complexity, and iii) the temporal tampering mechanism like that as insertion, deletion, duplication, and data shuffling in a video is not responsible for the detection of applicability of any image forgery detection mechanism.

There exist various techniques depicted through the literature surveys, particularly for detection as well as localization of video tampering. VCMF requires an exceptional mechanism that relevantly changes complicated modifications that are classified into two types one is inter-frame and the other is intra-frame, intra-frame forgery involves simultaneous activities by pasting each copied object from one frame into the same frame, as opposed to inter-frame forgery, which copies and pastes the object from one frame to another in a subsequent manner. The main aim of the VCMF mechanism results in confusing the frames by the addition of a few objects termed additive modification. Consequently, this is called a modification that aims at hiding a few objects. It is a complex task, which is cautiously constructed inter/interframe forgery by the above-stated machine learning techniques achieved by constant statistical measures. It is necessary because the relevant objects copied and background of the frame pasted is shot under specific surveillance camera, these techniques exhibit similar statistical applications and are hence in differentiable [7]–[10]. CMF mechanism seems to be the most challenging problem to tackle in the field of video forensics. Consequently, the proposed detection algorithms shift towards video copy-move forgeries, which leaves a strong impact on the current methodologies [11].

Pixel embedded correlation directed approach based on the applicability generally suffers from potential computational load termed as high computational complexity. In comparison with numerous videos, the majority of specific data ensures maximum effort on a large number of videos in comparison to the still images. Techniques based on image features result in unstable performance estimation incorporating additive noise, secondary compression, and post-processing of all threats to textual noise and pixel grey values. The constraints for finding sensitive parameters consider robustness into account for the existing approaches. Few techniques have been restricted in detecting videos in a specific format, the tampered frames, and ways of tampering for manipulation in various ways that restrict the applicability in video forensics. This method explicitly implies that a CMF detection mechanism resulting in excess demand necessitates three basic types of functionalities termed as a low computational complication, increased accuracy with robust pertinence. In this paper, a new approach is recreated by incorporating these three techniques and designing a unique technique for detecting the CMF mechanism.

Video copy-move forgery detection (VCMFD) is a major challenging task due to various obstacles including the requirement of video information, homogenous forgery sources, rich forgery objects and diverse types of forgery; these issues create challenges such as high false positives in forgery video detection, low trade-off efficiency and effectiveness. Hence, motivated by the challenges, this research work adopts the deep learning domain and provides the solution for the same; further research contribution is given as follows: i) this research work proposed dual deep network (DDN) for efficient and effective video forgery detection; DDN comprises two networks, first detection network (DetNet1) is utilized for general feature extraction whereas second detection network (DetNet2) is developed for custom and deep feature extraction; ii) DetNet1 and DetNet2 both are integrated models as the output of DetNet1 is given to DetNet2; iii) furthermore, the proposed research also develops algorithms for frame detection, frame matching and optimization of false detection; and iv) DDN is evaluated considering the REWIND and video tampering dataset (VTD) dataset considering different metrics like accuracy, precision, recall, and F1-score; further comparative analysis is carried out with various existing models.

Table 1. Original and copy move

Original	Copy move
	

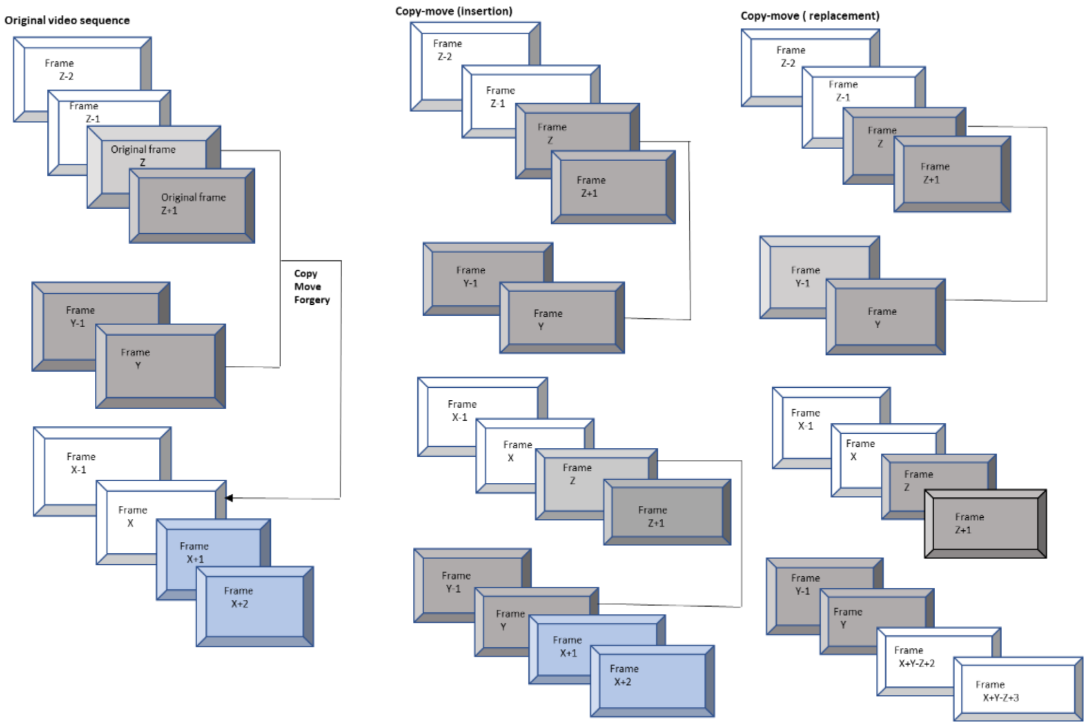


Figure 1. Copy move forgery

This research s organized as follows: the first section of the research starts with a background of video forgery and different video forgery along with highlights of CMF detection. The second section conducts a brief survey of the various existing model along with its shortcoming. Furthermore, the third section presents the proposed methodology along with a mathematical model and algorithm; at the last performance, an evaluation is carried out along with a comparative analysis to prove the model's efficiency.

2. RELATED WORK

Research carried out on various existing systems for the process that focuses on CMF detection is depicted in the form of examining the copy-move process's unintended consequences also described as feature correlation among the duplicated frames and original frame. Moreover, these are carried out through frame replacement or frame insertion. This section focuses on a review of various existing VCMFD. The prevailing effective existing systems are known as VCMFD techniques, such as dense moment feature index and best match (DFMI-BM) [12], exponential Fourier moments (EFMs) [13], PatchMatch-2D (PM-2D) [14], and PM-2D (fast) [14], are meticulously created and share the common concepts. Extracting the robust features by incorporating invariant features for several geometric and post-processing tasks for the section of forgery objects, serves as the critical approach to detecting the effectiveness of the approach VCMFD. In recent years the VCMFD method has applicability to existing methods for block extraction for invariant moments.

These invariant moments (such as the Polar complex exponential transform (PCET) [15] for the DFMI-BM, the Zernike [16] for the PM-2D and PM-2D (fast), and the EFMs have faultless invariances for rotation and mirror but lack scaling capabilities. These methods fall short of addressing scaled forgery techniques resulting in large-scale exponential transformations through factors ranging at least from 150% to 50%. Various algorithms match effective features, including the batch algorithm proposed by the effective DFMI-BM approach. PatchMatch is an algorithm proposed by PM-2D, whereas a fast match is an algorithm proposed by EFMs that looks for a potential block between matching pairs. Filtering and morphology are the post-processing techniques represented as the implementation. Summarizing the VCMFD methods are not capable of resisting scaling attacks as well as matching each step based on block approaches. The block features are determined in every pair, this particular process yields inefficient experimental findings. However, dense neural network (DNN) is studied in-depth and successful in an application to pattern classification and recognition with each aspect. The primitive DNN models, such as DenseNet [17], are not entirely competent when it comes to fraud detection because of the various forging kinds and complex backdrop contents. CMFD schemes are a few copy-image forgery detection approaches. Techniques like end-to-end Dense-InceptionNet (E-DIN) [18], a serial CMFD approach [19], and dual-order attentive generative adversarial network (DOA-GAN) [20] enhance the DNN detecting capabilities. The DenseNet, InceptionNet, VGG16, and VGG19 networks are essentially used for feature extraction in all three methods.

An image CMFD feature matching approach is the main component of these models, which are embedded in images, and it acts as a manual procedure. The E-DIN technique segments the correlation of feature matches using a second nearest-neighbor (2NN) test to determine the best match correspondingly. According to Liu *et al.* [21], a unique two-stage platform is designed specifically for the detection of copy-move fraud. The self-deep matching network's foundation is provided by the first stage. The second stage refers to the proposal SuperGlue, whereas the first stage shows the Atrous convolution-incorporating skip matching that ensures a spatial combination of and influences hierarchical features. A spatial mechanism based on self-correlation incorporates the capability to notice the appearance of relevant areas. In the second phase proposal, the superglue technique is to discard false alarmed regions and provide a remedy to incorporate incomplete regions. Furthermore, in [22] An accurate convolutional neural network (CNN) architecture-based method is suggested for the efficient detection of copy-move image tampering. The appropriate number of pooling convolutional layers is determined computationally by the suggested method. According to Zhong and Pun [18], an end-to-end-based method termed Dense-InceptionNet requires a multi-dimensional dense-feature connection known as a DNN. The first DNN model incorporates automatically based forgery snippets by matching values. The techniques for hierarchical post-processing, PFE modules are proposed to extract a multi-dimensional feature approach from a dimensional multi-scanned approach. For extracting dimensional and multi-scale information, the PFE modules are proposed. The features of each layer, which are ordered by direction, are extracted.

3. PROPOSED METHOD

Video is considered forged if the content is subjected to manipulation for the general viewer where the person’s intellect can be challenged and influenced. Forged video can mislead the general public and is quite difficult to identify especially forgery like copy move; thus, VCMFD has been one of the vital research areas utilizing various techniques like deep learning as it tends to extract the deep feature in comparison with the traditional approach. This research work adopts the deep learning domain for forgery detection where the main goal of our proposed model is CMF detection to differentiate between being original area and tampered area in a digital video. This research introduces DDN for VCMFD; the DDN model detects the tampered area and the original area. DDN comprises two detection networks i.e. DetNet1 and DetNet2; First detection network is responsible for general feature extraction and the second network i.e. DetNet2 is utilized for deep feature extraction. Moreover, the proposed workflow is presented in Figure 2.

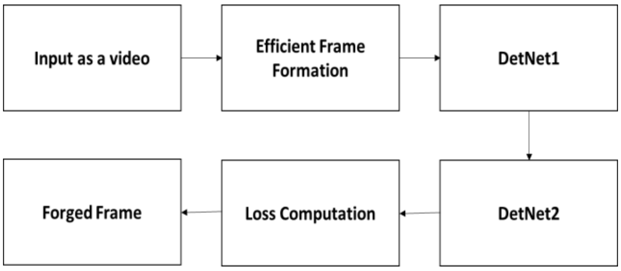


Figure 2. Dual deep network workflow

3.1. Efficient frame formation

In a collection video of S frames, in the first phase, the extraction of individual frames results in computing the optical flow of two parallel frames x to $x + 1(x = 1,2, \ldots \ldots \mathrm{S} - 1)$. A matrix is computed resulting in two directions like $\mathrm{oa_x}$ the matrix in the x direction and $\mathrm{oa_y}$ the matrix in the y direction. They are summed up to determine the sequences of the sums computed consisting of $\mathrm{S} - 1$ values. For the $\mathrm{x} - \mathrm{th}$ frame, it is possible to detect whether the frame is tampered with or the original one, a tampered area results in a sudden spike in the symmetry. The average mean is estimated by the parallel frames determined by (1):

$$\mathrm{total_{oa_x}} = \frac{1}{2\mathrm{S}}\sum_{\mathrm{a}=1}^{\mathrm{S}}(\mathrm{total_{oa_{x-a}}} + \mathrm{total_{oa_{x+a}}}) \tag{1}$$

here S is the size for finding the parallel frames. The shift of to α_x determine change of the $\mathrm{x} - \mathrm{th}$ frame is given by (2):

$$\alpha_\mathrm{x} = \frac{\mathrm{total_{oa_x}}}{\mathrm{total_{oa_x}}} \tag{2}$$

consider α_x larger than the threshold_A results in a spike in $\mathrm{total_{oa_x}}$, the tampered parallel frames as $(\mathrm{x} - 1)$th, $(\mathrm{x} + 1)$th frames are detected to find the tampered area. The $\mathrm{x} - \mathrm{th}$ frame is detected based on the symmetric center, which determines the CMF for $\mathrm{total_{oa_x}}$ where the x th frame is satisfied.

$$\mathrm{total_{oa_{x+a}}} \approx \mathrm{total_{oa_{x-a-1}}}\ \mathrm{a} = 0,1, \ldots \ldots \ldots \mathrm{s} \tag{3}$$

This determines that the frames have accurate $\mathrm{total_{oa_x}}$ before and after computation during symmetric centre and tampered frames. In the Algorithm 1, hence the $\mathrm{x} - \mathrm{th}$ frame is a probable tampered area detection process.

Algorithm 1. Probable tampered area detection

```
Input: $\mathrm{total_{oa_x}}(1 \leq \mathrm{x} \leq \mathrm{x} - 1)$, frame_size S, spike,
threshold_A
$\mathrm{W} = \{\}, \mathrm{U} = \{\}$
for $\mathrm{x} = 1;\ \mathrm{x} < \mathrm{S};\ \mathrm{x}++$ do
Compute $\mathrm{total_{oa_x}}$ from equation ()
Compute $\alpha_\mathrm{x}$ from equation ()
      If $\alpha_\mathrm{x} >$ threshold_A then
            Add $\mathrm{x}, \mathrm{x} + 1, \mathrm{x} - 1$ into W
      end if
            If $\frac{7}{10} < \frac{\mathrm{total_{oa_x}}}{\mathrm{total_{oa_x}}} < \frac{10}{7}$ (a = 0, 1, …… T)
                  add a into U
            end if
end for
Output: tampered area detection: spike set $W$ and symmetric centre $U$
```

3.2. DetNet1

The number of network features is decreased in the pooling layer henceforth it results in a reduction of spatial resolution. To enhance the features generated results in the high-resolution feature maps neural network that extracts the features as shown in Figure 3. The CMF detection mechanism separates the original area and the tampered area.

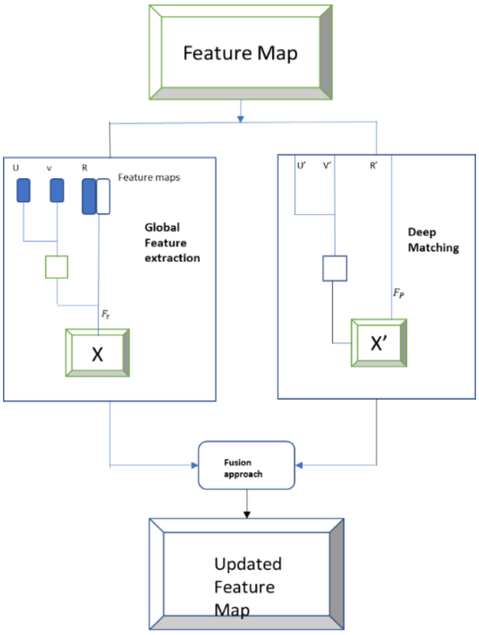


Figure 3. DetNet architecture

3.2.1. Global feature extraction through dilated convolution

Upon application of the self-attention methodology, a broad spectrum of information is embedded in the features. These features enhance the neural network features, below-given matrix maps the features of AM are computed as stated in (4):

$$\mathrm{AM}_{\mathrm{xy}} = \frac{\exp[\,\mathrm{U_x}*\mathrm{V_y}]}{\sum_{\mathrm{x}=1}^{\mathrm{S}}\exp[\,\mathrm{U_x}*\mathrm{V_y}]}` \tag{4}$$

in the attention module $\mathrm{AM}_{\mathrm{xy}}$ determines the impact of $\mathrm{x}-\mathrm{th}$ pixel,$\mathrm{y}-\mathrm{th}$ pixel, U and V are feature maps after convolution, normalization and rectified linear unit (ReLU), the self-attention feature maps $\mathrm{F_P}$ (5):

$$\mathrm{F_P} = \beta(\mathrm{R} * \mathrm{H}) + \mathrm{A} \tag{5}$$

β is the learning constraint initialised with a value of 0. R is the feature map extracted after each convolution. Whereas $\mathrm{F_P}$ and $\mathrm{F_t}$ is determined by Figure 2. This transfer results in information loss independent of the weights associated with each other. The $\mathrm{F_P}$ and $\mathrm{F_t}$ values are fused along each other ensuring a relationship between the features at various positions. The CMF detection module captures the context information that represents the convolution features. This is given as (6):

$$\mathrm{F} = \tau * \mathrm{F_P} + \mu * \mathrm{F_t} \tag{6}$$

here, τ and μ are the parameters associated with the Gaussian distribution, that are learned during the training process.

3.2.2. Estimation of correlation

To estimate the correlation features of the main issue encountered here, the forged frames are generated, in correspondence to this the original area in the frame is also found and, the tampered area is mapped from the original area which helps in allocating the similar area. L^3, L^4 and L^5 estimate the mapped features. The similarity measure of $\mathrm{T}_{\mathrm{a,b}}^{\mathrm{v}}$ between the $\mathrm{a}-\mathrm{th}$ patch, $\mathrm{L}_{\mathrm{a}}^{\mathrm{v}}$ and the mapped feature of b_th patch $\mathrm{L}_{\mathrm{b}}^{\mathrm{v}}$ is determined as given by (7):

$$\mathrm{T}_{\mathrm{a,b}}^{\mathrm{v}} = (\mathrm{L}_{\mathrm{a}}^{\mathrm{v}})^{\mathrm{X}}(\mathrm{L}_{\mathrm{b}}^{\mathrm{v}}) \tag{7}$$

the irrelevant information is not considered, a sorting technique is used here that selects the index corresponding to $\mathrm{index}_{\mathrm{v}}(\mathrm{X})$ and further mathematical formulation of it is given as (8):

$$\mathrm{index_v(X) = Peak\_X\_index(T^v, X)} \quad (8)$$

Peak_X_index denotes the peak value, and T^v is considered as the similarity measure of mapped feature L^v. The mapped features have a similar dimension but different channels. The mapped features have the same dimensions but different paths, the matching process L_{total} is given as (9):

$$\mathrm{L_{total} = (L^3\big(index_3(X)\big), L^4(X), L^5(index_5(X))} \quad (9)$$

the tampered region is necessarily scaled in the CMF given, as it is essential to utilise the correlation mechanism.

3.2.3. DetNet2

The existing methods are capable of only detecting the tampered area and not the fine-tuning of the model, which affects the model largely affects the detection. DetNet2 comprises five components; the first component includes the input layer, a down-sampling layer, an up-sampling layer, a bridge layer, and an output layer. Moreover, the input layer comprises 64 filters along with activation function and batch normalization; furthermore, bilinear interpolation is utilized for up-sampling and average pooling for down sampling. The skip connection layer is introduced after up sampling; also another activation function is added. Figure 4 displays the DetNet2 architecture.

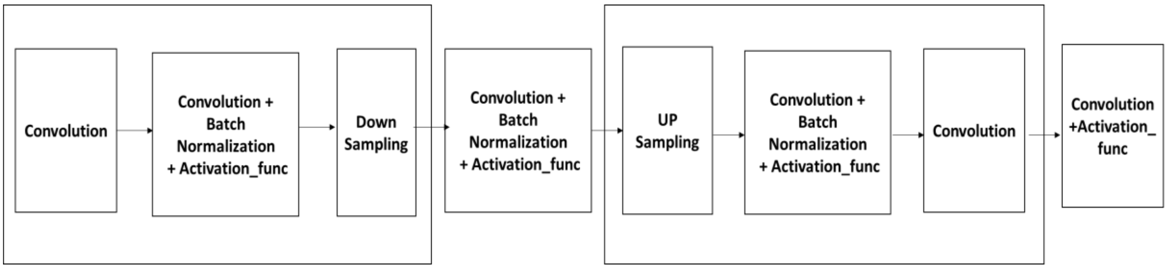

Convolution
Convolution + Batch Normalization + Activation_func
Down Sampling
Convolution + Batch Normalization + Activation_func
UP Sampling
Convolution + Batch Normalization + Activation_func
Convolution
Convolution +Activation_ func

Figure 4. DetNet2 architecture

3.3. Frame matching algorithm

The algorithm presented in Algorithm 2 is to match the frame after duplication, determine the tampered area, and estimate the correlation coefficient. It is essential to sample the input for reducing the number of pixels for computational purposes. The efficiency of the computation is enhanced to find the distribution coefficient. The procedure for frame matching is given by Algorithm 2.

Algorithm 2. Algorithm frame matching

```
Input: oa sequence $oa_x(1 \le x \le S)$, Spike set W, symmetry centre U, $threshold\_A_{D1}$, $threshold\_A_{D2}$
$FD = \{\}$
for each frame $x \in U$ do
     for $y = 1; y < S; y++$ do
          Calculate $CC(x,y)$ from equation ()
     end for
obtain $Max\_CC_{x,y}$
$cc(x,y)_{max,}$ x€y
if $cc(x,y)_{max,} \ge threshold\_A_{D1}$ then
     add$(x,y),(x+1,y+1)$ into FD
end if
end for
for each frame number $x \in G$ do
     $a = 0$
     while cc $(x+a, x-a-1) > threshold\_A_{D2}$ &
          cc $(x+a+1, x-a-2) \ge threshold\_A_{D2}$
              do
              $a = a + 2$
     end while
     add$(x-a-1, x+a+1)$ into FD
end for
Output: Frame set duplication FD
```

3.4. False detection reduction

Furthermore, Algorithm 3 presents the optimization process of false frame detection that comprises various phases, in the case of the first phase the tampered frame is detected by an abnormal spike by the correlation coefficients to determine the maximum among the correlation coefficient used. Among the correlation coefficients threshold_A_{D1} has significantly higher value in determining the similar frame set. For each tampered frame x is detected as n the local symmetric center. The while loop is iterated multiple times for the copy-moved frames. The output of the given algorithm is given by the initiation and end of the tampered frames.

Algorithm 3. False frame detection optimization

```
Input: tampered frame set $DF, threshold\_A_{D2}$
$Min\_tampered frames\ \ WS = 10$
for each frame set $(x, y)\ \in\ DF$ do
    If $(|y - x| < H)||\ (cc(x-1, y-1), (cc(x+1, y+1) <\ threshold\_A_{D2})||$
        $threshold\_A_{D2})\ ||\ (cc(x-1, y-1), cc(x+1, y+1) >$
        $threshold\_A_{D2}$ do
        Remove$(x, y)$ from $DF$ do
    end if
end for
Select $(x_m, y_m), (x_n, y_n) \in\ DF\ x_m < y_m$, and $x_m - y_m = |x_n - y_n|$
Output $\{(x_m, x_{m+1,.......}x_n, x_{n+1}\}$ as Tampered frame
$\{y_m, y_{m+1}, \dots \dots y_{n,}y_{n+1}\}$as genuine frame
For each frame set $(a, x, y)\ \in\ DF$ do
    If $|a - y| < 2\ WS$ do
        Remove $(a, x, y)$ from $DF$
        else output $\{x+1, \dots, y-1, y\}$ as tampered frames,
            $\{a, a+1, \dots \dots . . x-1\}$ as genuine frame
    end if
end for
Output: tampered along with original_frame sequences
```

The copy-move forgeries necessarily result in the abnormal behavior of the sum of sequences. The tampered area is not the only factor necessary for determining the spikes when a tampered area is detected. Many other factors are also responsible for the rise in spikes or local symmetric centres. In the correlation phase, the parallel frames with high similarity may result in false detection.

3.5. Loss computation network

While training the module the cross-entropy function value minimizes the constraint set in the network. Forgery detection is essential for classification. The cross-entropy function value is calculated as (10):

$$L_{cel} = -\sum_{m,n} P(x, y) \log\big(X(x, y)\big) + (1 - P(x, y))\log(1 - X(x, y))) \qquad (10)$$

were, $P(x, y) \in \{0,1\}$ denotes the pixel value of (x, y) and X, also denoted as the tampered area. The loss is considered in each pixel and the relationship between the adjacent pixels is considered between the boundary of the tampered area and the original area. To ensure the structural information the summation of all the losses is given as L_{lf}:

$$L_{if} = L_{cel} + L_{sl} + L_{iu} \qquad (11)$$

here L_{cel}, L_{sl}, L_{iu} where L_{cel}, determines its ability for segmentation purposes at each pixel level and assists the model to meet on all pixels, L_{sl} determines the similarity loss and L_{iu} determines the loss encountered by performing intersection over the union. L_{cel} loss determines the total loss encountered by each pixel. Whereas L_{sl} loss is responsible for fine-tuning the network that focuses more on the tampered area. L_{sl} loss is determined by (12):

$$L_{sl} = 1 - \frac{(2i_P i_X + \mu_1)(\tau_{PX} + \mu_2)}{(i_P^2 + i_X^2 + \mu_1)(\tau_P^2 + \tau_X^2 + \mu_2)} \qquad (12)$$

here, $i_P i_X$ indicates the average mean of P and X, τ_P and μ_2 is the standard deviation and covariance matrix:

$$L_{iu} = \frac{\sum_{x=1}^{K} \sum_{y=1}^{R} P(x,y) X(x,y)}{\sum_{x=1}^{K} \sum_{y=1}^{R} (X(x,y) + P(x,y) - X(x,y) + P(x,y)} \qquad (13)$$

L_{iu} loss is estimated during the training process to detect the object and segmentation. These three losses are combined to generate necessarily a hybrid loss as depicted in (11).

4. PERFORMANCE EVALUATION

This section of the research evaluates the proposed model; moreover, evaluation is carried out on the ideal system configuration of Windows 10 packed with 16 GB of RAM along with 4 GB of Cuda-enabled Nvidia graphics. Furthermore, the model is designed considering the deep learning architecture with the help of various libraries using python as a programming language. This section evaluates the proposed model considering the different metrics; also, the efficiency of the model is proved through comparative analysis with the state-of-art technique and existing model.

4.1. Dataset details

VCMF is one complex manipulation, which is carried out with relatively complex manipulation; thus, designing the dataset for the same is quite complicated. This research considers two distinctive datasets namely REWIND [23] and VTD [24]. This two-benchmark dataset comprises various CMF i.e. inter-frame and intra frame, which has been discussed later.

4.2. Metrics evaluation

4.2.1. Accuracy

Accuracy is metric which is described as how a model performs across various classes; here it tends to predict the forgery frame and is computed as (14).

$$\text{Accuracy} = \frac{\text{true}_{\text{neg}}+\text{true}_{\text{pos}}}{\text{true}_{\text{pos}}+\text{false}_{\text{neg}}+\text{false}_{\text{pos}}+\text{true}_{\text{neg}}} \tag{14}$$

4.2.2. Precision

Precision is defined as the collective ratio among the correctly classified forged frame and positive samples observed, given as (15).

$$\text{precision} = \frac{\text{true\_pos}}{\text{false}_{\text{pos}}+\text{true\_pos}} \tag{15}$$

4.2.3. Recall

The recall is defined as the collective ratio among the number of the positive samples classified correctly to the completely positive numbers and given as (16).

$$\text{Recall} = \frac{\text{true}_{\text{pos}}}{\text{false}_{\text{neg}}+\text{true\_pos}} \tag{16}$$

4.2.4. F1-score

F1-score integrates the precision along with the recall of classifier into the particular metric through computation of harmonic mean and it is computed as (17):

$$\text{F1} - \text{score} = \frac{2\text{True}_{\text{positive}}}{\text{False}_{\text{negative}}+\text{False}_{\text{positive}}+2\ \text{True}_{\text{positive}}} \tag{17}$$

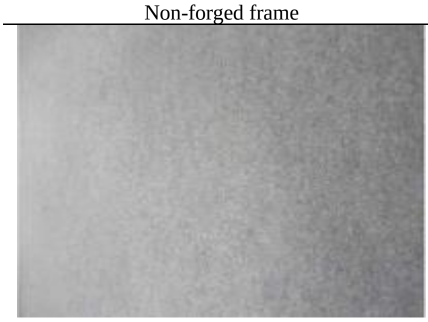
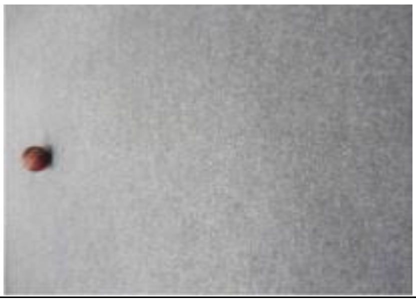
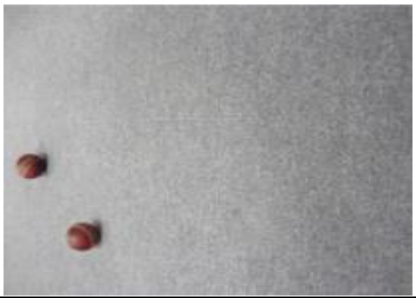
4.3. Dataset 1 evaluation

REWIND dataset is one of the benchmark datasets where there are 10 distinctive genuine videos along with 40 derivative inter-frame forgeries and 10 forged videos; moreover, each sequence has a frame rate of 30 fps This dataset is designed for video-based CMFD. Furthermore, evaluation is carried out on considering the Detection accuracy, false positive and F1-score with existing comparison model E-DIN [18], serial-CMFD [19], PM-2D (fast) [14], PM-2D [14], DFMI-BM [12], and existing model novel-VCMFD [4]. Table 2 presents the sample frame of the non-forged frame and forged frame.

4.3.1. Detection accuracy

Figure 5 shows the number of frames detected correctly in a given video; the x-axis presents the number of various methodologies and y-axis presents the forged videos. In the case of the E-DIN mechanism, 5 videos were detected correctly whereas, in the case of serial-CMFD, PM-2D (fast), and PM-2D observes 6, 9 and 9 videos were detected as forged respectively. Similarly, DFMI-BM also detects 9 videos as forged videos. Moreover, the existing model detects 10 videos as forged so as the proposed model.

Table 2. Sample non-forged frame and forged frame

Non-forged frame	Forged_frame
	
	

4.3.2. Falsely positive comparison

Figure 6 presents a false positive comparison; y-axis presents false positive and x-axis presents methodologies; despite detecting the video as forged, it is also important to detect the correct frame as an incorrect frame leads to misconception; Figure 6 shows the comparison of the falsely detected frame. Moreover, serial-CMFD, E-DIN, PM-2D (fast), and PM-2D detect 5, 3, 3 and 2 videos incorrectly out of 10. VCMD i.e. existing model fails in 1 whereas the proposed model fails in none.

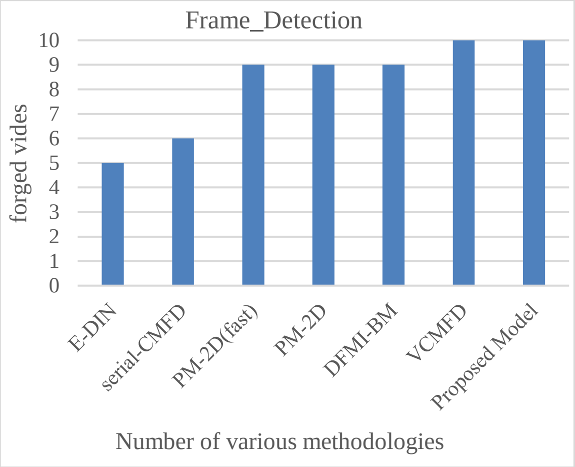


Figure 5. Frame detection rate

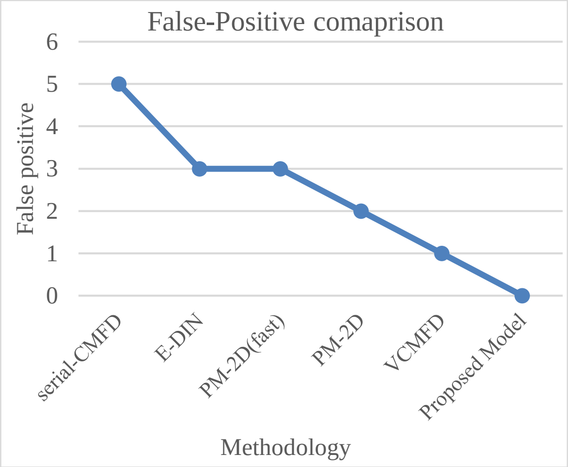


Figure 6. False positive

Figure 7 shows the F1-score comparison on dataset 1; E-DIN and serial-CMFD observe very low F1-score of 16% and 19%, whereas other methodologies like PM-2D (fast), PM-2D, and DFMI-BM observe above-average F1-score of 79%, 84%, and 86%. Similarly, the existing model observes 87% whereas the proposed model observes a 95% F1-score.

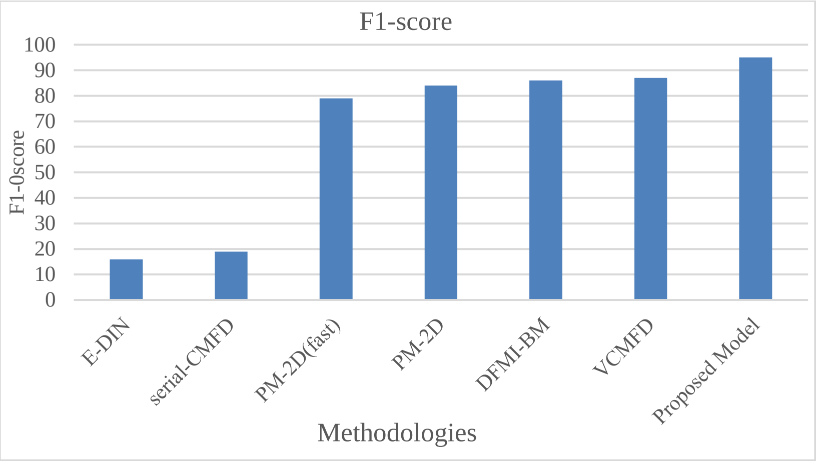


Figure 7. F1-score comparison

4.4. Dataset 2 evaluation

VTD dataset [24] is another public forensic library for different types of forgery including CMF; moreover, this dataset is modified in the year 2019. Each of the videos comprises the quality of 720p. Table 3 presents the sample forged frame and non-forged frame.

Table 3. Sample forged and non-forged frames from the VTD dataset

Non_forged frame	Forged_frame

DDN model is evaluated considering the accuracy, precision, recall, and F1-score with comparing with various existing model like fast and robust [25], histogram of oriented gradients (HOG) and compression [26], adaptive over segmentation [27], spatio-temporal context [28], inter-frame mechanism [29], local binary patterns (LBP)-detection [30], discrete Radon polar complex exponential transform (DRPCET) [31], fast and effective [32], and existing model i.e. video forgery detection using the histogram of second order gradients (VFDHSOG) [33].

Figure 8 shows the accuracy comparison of the various existing model considering the various model; method like fast and robust-CMFD achieves an average accuracy of 69.7%, and other models like HOG and compression, adaptive over-segmentation and spatiotemporal context achieves good accuracy of 88.3%, 91.4%, and 93.1%. Similarly, inter-frame achieves the accuracy of 96.3; in comparison to all these

VFDHSOG achieves the accuracy of 92.6 and the proposed model DDN achieves the accuracy of 98.3%. Figure 9 shows the recall comparison of the various existing model; model like LBP-detection, DRPCET and VFDHSOG model achieves a recall value of 82.7 %, 92.7%, and 93.2% respectively. Similarly, fast and effective-CMFD achieves a recall value of 95.8 whereas dual deep network-proposed system (DDN-PS) achieves 97.2%.

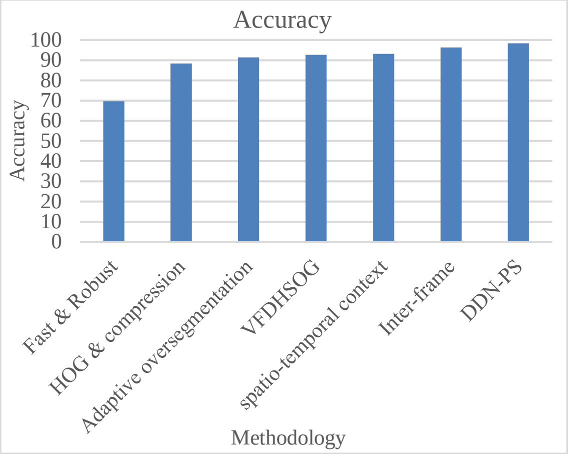


Figure 8. Accuracy comparison

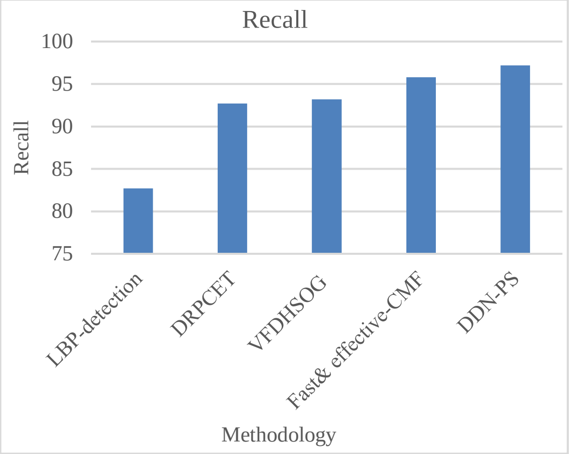


Figure 9. Recall comparison

Figure 10 shows the precision comparison of a various existing model; model like LBP-detection, DRPCET and the fast and effective model achieves a recall value of 89.5%, 94.5%, and 94.4% respectively. Similarly, VFDHSOG achieves a recall value of 95.4 whereas DDN-PS achieves 97.3%. Figure 11 shows the F1-score comparison of the various existing model; models like LBP-detection, DRPCET, and VFDHSOG model achieves recall value of 88.1%, 93.6%, and 94.28% respectively. Similarly, fast and effective CMFD achieves a recall value of 95.2 whereas DDN-PS achieves 98.6%.

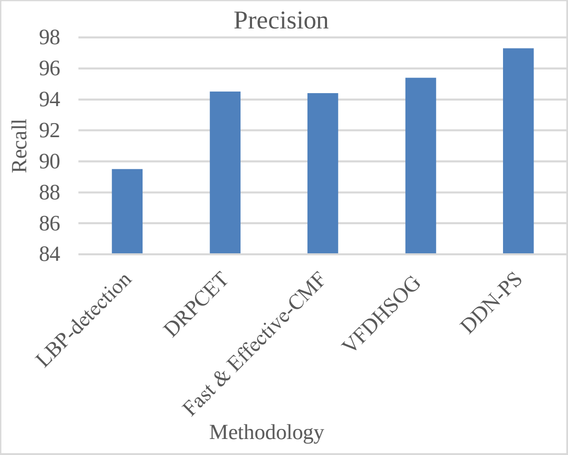


Figure 10. Precision comparison

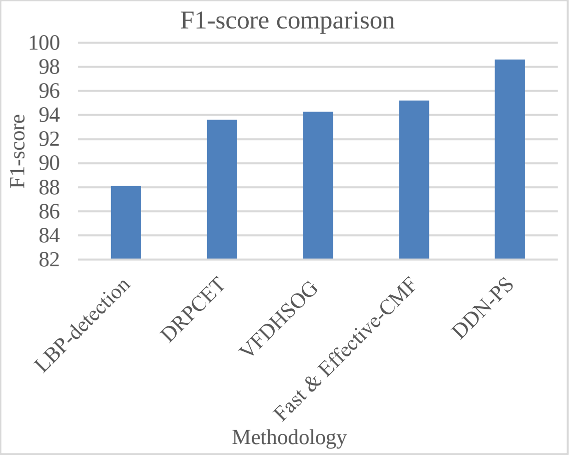


Figure 11. F1-score comparison

4.5. Comparative analysis and detection

This section discusses the improvisation of DDN over the existing model considering various parameters; considering the dataset 1 evaluation, all 10 videos were detected correctly. Furthermore, the existing model false positive is 1 out of 10 whereas DDN-PS false positive is 0. Furthermore, considering dataset 2, DDN achieves accuracy improvisation of 2%, recall improvisation of 1.45%, precision improvisation of 1.97%, and F1-score of 3.50% with the best performance model.

5. CONCLUSION

Tampering the digital videography, which serves as a reference in the court, is uncertain and stays still in its early stages and reliability in the field of digital video forensics. Various models for video editing that as Adobe's (Premier and After Effect), GNU Gimp, Premier, and Vegas are freely available which tamper with the video content. Various techniques are proposed here in the past literature survey that detect tampered video content; however, these models suffer from limitations. Thus, this research develops DDN for video forgery detection; DDN comprises two networks for general feature detection and a deep custom feature to distinguish between the original frame and tempered frame. DDN is an end-to-end approach for forgery detection where the output of DetNet1 is integrated to DetNet2 and optimality is carried out; also, three algorithms for probably tampered detection algorithm, frame matching and reducing false detection are introduced for efficient and effective forgery detection. DDN is evaluated considering the two benchmark datasets i.e. REWIND and VTD dataset considering the various metrics; comparative analysis shows that DDN outperforms the other existing model with marginal improvisation as DDN achieves lower false positive, higher detection accuracy for REWIND dataset and higher value of precision, recall, accuracy and F1-score for dataset 2. The future work would focus on enhancing the ability of the system to deal with tampered videos in the context of large static scenes and careful modification. Further, the aim focus should be a generation of a more comprehensive approach based on a large-scale video forgery approach, which serves as the basis for future work.

REFERENCES

[1] S. Teerakanok and T. Uehara, "Copy-move forgery detection: a state-of-the-art technical review and analysis," *IEEE Access*, vol. 7, pp. 40550–40568, 2019, doi: 10.1109/ACCESS.2019.2907316.

[2] K. H. Rhee, "Generation of novelty ground truth image using image classification and semantic segmentation for copy-move forgery detection," *IEEE Access*, vol. 10, pp. 2783–2796, 2022, doi: 10.1109/ACCESS.2021.3136781.

[3] Abhishek and N. Jindal, "Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3571–3599, 2021, doi: 10.1007/s11042-020-09816-3.

[4] J.-L. Zhong, Y.-F. Gan, C.-M. Vong, J.-X. Yang, J.-H. Zhao, and J.-H. Luo, "Effective and efficient pixel-level detection for diverse video copy-move forgery types," *Pattern Recognition*, vol. 122, p. 108286, Feb. 2022, doi: 10.1016/j.patcog.2021.108286.

[5] S. Tyagi and D. Yadav, "A detailed analysis of image and video forgery detection techniques," *The Visual Computer*, vol. 39, no. 3, pp. 813–833, Mar. 2023, doi: 10.1007/s00371-021-02347-4.

[6] J. Hu, X. Liao, J. Liang, W. Zhou, and Z. Qin, "FInfer: frame inference-based deepfake detection for high-visual-quality videos," *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, vol. 36, pp. 780–789, 2022, doi: 10.1609/aaai.v36i1.19978.

[7] S. Fadl, Q. Han, and Q. Li, "CNN spatiotemporal features and fusion for surveillance video forgery detection," *Signal Processing: Image Communication*, vol. 90, p. 116066, Jan. 2021, doi: 10.1016/j.image.2020.116066.

[8] N. Akhtar, M. Saddique, K. Asghar, U. I. Bajwa, M. Hussain, and Z. Habib, "Digital video tampering detection and localization: review, representations, challenges and algorithm," *Mathematics*, vol. 10, no. 2, 2022, doi: 10.3390/math10020168.

[9] H. Pu, T. Huang, B. Weng, F. Ye, and C. Zhao, "Overcome the brightness and JITTER noises in video inter-frame tampering detection," *Sensors*, vol. 21, no. 12, p. 3953, Jun. 2021, doi: 10.3390/s21123953.

[10] M. Aloraini, M. Sharifzadeh, and D. Schonfeld, "Sequential and patch analyses for object removal video forgery detection and localization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 917–930, Mar. 2021, doi: 10.1109/TCSVT.2020.2993004.

[11] S. T. Nabi, M. Kumar, P. Singh, N. Aggarwal, and K. Kumar, "A comprehensive survey of image and video forgery techniques: variants, challenges, and future directions," *Multimedia Systems*, vol. 28, no. 3, pp. 939–992, Jun. 2022, doi: 10.1007/s00530-021-00873-8.

[12] J.-L. Zhong, C.-M. Pun, and Y.-F. Gan, "Dense moment feature index and best match algorithms for video copy-move forgery detection," *Information Sciences*, vol. 537, pp. 184–202, Oct. 2020, doi: 10.1016/j.ins.2020.05.134.

[13] X. Bi and C.-M. Pun, "Fast copy-move forgery detection using local bidirectional coherency error refinement," *Pattern Recognition*, vol. 81, pp. 161–175, Sep. 2018, doi: 10.1016/j.patcog.2018.03.028.

[14] 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, doi: 10.1109/TCSVT.2018.2804768.

[15] J.-L. Zhong and C.-M. Pun, "Two-pass hashing feature representation and searching method for copy-move forgery detection," *Information Sciences*, vol. 512, pp. 675–692, Feb. 2020, doi: 10.1016/j.ins.2019.09.085.

[16] S. J. Ryu, M. Kirchner, M. J. Lee, and H. K. Lee, "Rotation invariant localization of duplicated image regions based on zernike moments," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 8, pp. 1355–1370, 2013, doi: 10.1109/TIFS.2013.2272377.

[17] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected con- volutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[18] J. L. Zhong and C. M. Pun, "An end-to-end Dense-InceptionNet for image copy-move forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2134–2146, 2020, doi: 10.1109/TIFS.2019.2957693.

[19] B. Chen, W. Tan, G. Coatrieux, Y. Zheng, and Y. Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," *IEEE Transactions on Multimedia*, vol. 23, pp. 3506–3517, 2021, doi: 10.1109/TMM.2020.3026868.

[20] A. Islam, C. Long, A. Basharat, and A. Hoogs, "DOA-GAN: dual-order attentive gen- erative adversarial network for image copy-move forgery detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4676–4685.

[21] Y. Liu, C. Xia, X. Zhu, and S. Xu, “Two-stage copy-move forgery detection with self deep matching and proposal SuperGlue,” *IEEE Transactions on Image Processing*, vol. 31, pp. 541–555, 2022, doi: 10.1109/TIP.2021.3132828.

[22] K. M. Hosny, A. M. Mortda, M. M. Fouda, and N. A. Lashin, “An efficient CNN model to detect copy-move image forgery,” *IEEE Access*, vol. 10, pp. 48622–48632, 2022, doi: 10.1109/ACCESS.2022.3172273.

[23] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, “Local tampering detection in video sequences,” in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, Sep. 2013, vol. 15, pp. 488–493, doi: 10.1109/MMSP.2013.6659337.

[24] O. I. Al-Sanjary, A. A. Ahmed, and G. Sulong, “Development of a video tampering dataset for forensic investigation,” *Forensic Science International*, vol. 266, pp. 565–572, 2016, doi: 10.1016/j.forsciint.2016.07.013.

[25] S. Fadl, Q. Han, and L. Qiong, “Exposing video inter-frame forgery via histogram of oriented gradients and motion energy image,” *Multidimensional Systems and Signal Processing*, vol. 31, no. 4, pp. 1365–1384, Feb. 2020, doi: 10.1007/s11045-020-00711-6.

[26] V. K. Singh, P. Chakraborty, and R. C. Tripathi, “Detection and localization of duplicated frames in doctored video,” *Advances in Intelligent Systems and Computing*, vol. 736, pp. 661–669, 2018, doi: 10.1007/978-3-319-76348-4_64.

[27] L. Su and C. Li, “A novel passive forgery detection algorithm for video region duplication,” *Multidimensional Systems and Signal Processing*, vol. 29, no. 3, pp. 1173–1190, 2018, doi: 10.1007/s11045-017-0496-6.

[28] C. M. Pun, X. C. Yuan, and X. L. Bi, “Image forgery detection using adaptive oversegmentation and feature point matching,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1705–1716, Apr. 2015, doi: 10.1109/TIFS.2015.2423261.

[29] A. V. Subramanyam and S. Emmanuel, “Video forgery detection using HOG features and compression properties,” *2012 IEEE 14th International Workshop on Multimedia Signal Processing, MMSP 2012 - Proceedings*, pp. 89–94, 2012, doi: 10.1109/MMSP.2012.6343421.

[30] J. W. Wang, G. J. Liu, Z. Zhang, Y. W. Dai, and Z. Q. Wang, “Fast and robust forensics for image region-duplication forgery,” *Zidonghua Xuebao/ Acta Automatica Sinica*, vol. 35, no. 12, pp. 1488–1495, 2009, doi: 10.3724/SP.J.1004.2009.01488.

[31] Y. Li and J. Zhou, “Fast and effective image copy-move forgery detection via hierarchical feature point matching,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1307–1322, 2019, doi: 10.1109/TIFS.2018.2876837.

[32] J. Zhong, Y. Gan, J. Young, and P. Lin, “Copy move forgery image detection via discrete radon and polar complex exponential transform-based moment invariant features,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 2, 2017, doi: 10.1142/S0218001417540052.

[33] Z. Zhang, J. Hou, Q. Ma, and Z. Li, “Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames,” *Security and Communication Networks*, vol. 8, no. 2, pp. 311–320, 2015, doi: 10.1002/sec.981.

BIOGRAPHIES OF AUTHORS

Chandrakala currently, working as assistant professor in Godutai College of Education for Women, Kalaburagi with 10 years of teaching experience in computer science and engineering domain. Her areas of interest are image processing, cyber security, IoT and machine learning. Published 4 papers attained many conferences and workshops. She can be contacted at email: chandrakalavpatil15@gmail.com, chandrakala_2022@rediffmail.com, or chandrakalmail@rediffmail.com.

Mungamuri Sasikala completed B.E. in electrical engineering from Osmania University Hyderabad in the year 1985 with first class distinction. She obtained M.E. in the year 1987, in power systems from Osmania University and was awarded a gold medal for standing first in the university. She completed her Ph.D. from JNTU Hyderabad in electrical engineering in the year 2008. She working as a professor/principal in various colleges since 2008 and currently working as the principal of Godutai Engineering College for Women, Sharnbasva University, India. Kalaburagi’ since 2014. She has 3 decades (30 years) of teaching experience. She can be contacted at email: sasi_mum@rediffmail.com or sasi_mun@redifmail.com.