Semiconductor Optoelectronics
ISSN: 1001-5868

VIDEO INTER-FRAME FORGERY DETECTION USING MACHINE LEARNING

Raksha Pandey[1], Alok kumar Singh Kushwaha[2]

[1]Assistant Professor, Dept. of Computer Science and Engineering,
School of Studies Engineering & Technology, Guru Ghasidas Viswavidyalaya, A Central University, Bilaspur, Chhattisgarh, India

[2]Associate Professor, Dept. of Computer Science and Engineering,
School of Studies Engineering & Technology, Guru Ghasidas Viswavidyalaya, A Central University, Bilaspur, Chhattisgarh, India

Email Id: rakshasharma10@gmail.com1 ,alokkumarsingh.jk@gmail.com2

Abstract:
Surveillance equipment is present in almost every aspect of our lives, and recordings captured by surveillance systems are often considered to be substantial evidence in forensic investigations. It is of the utmost importance to quickly resolve the issue of verifying the validity of surveillance footage. One of the most prevalent methods for manipulating with videos is known as inter- frame forgery. The Video forgery will have the effect of lowering the correlation between the frames that are near to the tampering spot. Each frame's 2-D phase congruency is assessed during the whole feature extraction procedure. Due to the importance of this specific visual component, this is done. Then compute the correlation between the two frames that are next to one another. The technique consisted of looking for and labelling any apparent anomalous spots that were found using the k-means clustering algorithm. Depending on whether or not the points were within the normal range or the abnormal range, they were divided into one of two distinct categories. The study's results show that the approach offers a high degree of accuracy in target localization and identification. The approach that had been suggested had a detection accuracy of 99.17% after all of the tests had been completed to completion. This finding shows that the current method is far more precise and efficient than other methods that have been developed more recently in deep learning. In order for us to reach this conclusion, the performance of the recommended technique was compared to the performance of earlier efforts that were based on the state-of-the-art. After discussing the limitations of the experimental design that is presently being employed and analyzing the applicability of our results, we provide some suggestions for the course that future research in this field should follow. In this paper, we examine the literature and discuss the challenges of detecting video, video tampering when a passive strategy is used.

Keywords: Video forgery, Video tampering, Deep learning.

1. Introduction:

People often believe that video clips may provide more forensic proof than still images. As a consequence, because it is crucial evidence, surveillance film is often utilised in the course of an investigation. On the other hand, the digitising feature makes it simple to edit surveillance film. It is straightforward to alter a digital movie without leaving any visible traces when using video editing software like Adobe Premiere [1]. Because of this, the computer science discipline known as "digital video forensics," which investigates the issue of whether or not digital movies can be trusted, has grown to be an important and exciting area of research. investigated the advantages of such authentication for the legal and journalistic domains of the media business and proposed a system for authenticating and validating material across a range of media platforms[2]. On the other hand, it wasn't always feasible to implant watermarks into the films that were studied owing to technological restrictions. Most of the videos had this situation. As a result, the current research has given considerable weight to non-previous knowledge-based detection techniques, such as the identification of forgery evidence. This is due to the significance of the current situation. A prime example of this is knowing how to spot the telltale signs of fraud[3]. The two factors that might be utilised to distinguish between real and fake surveillance film are the source's truthfulness and the content's validity. The phrase "source authenticity" refers to the process of analysing the video to ascertain "where it originated from" and "how it came from."To analyse each of the steps of the acquisition process, several alternative methodologies have been developed. It is established whether or not the video has been manipulated in any way throughout the content authenticity testing process. Depending on the container, the video file that the camera produces will have a certain extension[4]. With the aid of the metadata, the container will specify the file's structure [5]. The codec, which is an encoded stream of bytes that makes up the video's content, is the most crucial element in deciding the overall quality of the video. For instance, a well-known container may hold a variety of various codecs. H.264 data is included in a MOV file. A video is just a collection of still photos that have been arranged into frames and joined together in a certain sequence. It is a collection of succeeding frames, often referred to as GOPs, that are organised in a three- dimensional plane and have a temporal dependency. Each each frame recorded by a single camera movement makes up a shot. One or more photos combine to form a scene that makes sense as a whole. One of the techniques used to encode a video is quantization, and it has a noticeable impact on the sequence itself. The act of deliberately editing or manipulating a digital video for fabrication purposes is referred to as "digital video forging." Its implications vary depending on the context and environment in which it is used. It has a big impact, especially in the entertainment, political, and medical fields where it's often used to tarnish popular figures, hide or make up important details, and either lie about or cover up actual occurrences[6]. Our daily lives are significantly and favourably impacted by how often we view videos on different social media platforms, including Facebook, WhatsApp, YouTube, and other news channels. "Being seen is no longer being believed," The integrity and validity of the footage that is being shown cannot be simply believed to be true [7]. In the meanwhile, methods for detecting manipulated microscopic portions inside frames are known as video intra-frame forgeries. Area duplication, area deletion, and area insertion are some of these changes. Recent advancements in video editing software have made it possible to copy and paste 3-D parts of recordings and quickly change their brightness, geometry, and other comparable attributes. It's possible for the 3-D parts to be little 3-D fragments inside subsequent frame sequences or whole consecutive frame sequences. Since these fake movies are becoming a prevalent technique used in video tampering, it could be difficult to detect them with the naked eye[8]. A lot of effort has been put into creating algorithms that can recognise this sort of fake footage. Additionally, there are frame-level changes that repeat or conceal objects in the movie. These easy adjustments may be used while changing the video's content. They would, however, produce fake movies that were difficult to view, especially with naked eyes [9]. Since the advent of contemporary computer and multimedia technology, digital video has become the dominant method of network communication due to its accessibility, mobility, and substantial information content. It has evolved into an important body of information that is used in a variety of significant situations, including the media, politics, insurance claims, defence, and legal issues, among many other important themes and subjects. On the other hand, since strong multimedia editing software is so widely accessible, even novice users are able to make fair changes to video footage. As a consequence, even for experts, it could be difficult to tell certain hoaxes apart from the real deal. This is because some of the bogus videos were produced utilising expensive multimedia editing programmes. It was because of this that everything happened[10]. As a consequence of a number of these various worries, some individuals have begun to question the veracity of the digital video recording. As a result, there is an urgent need for trustworthy forensic technology that can attest to the veracity, accuracy, and authenticity of video data. This technique may be used to maintain social order and maintain unity in communities, as well as prevent dishonest video manipulation from harming the broader population. [11].

The method that has been created makes it simple to spot Inter-Frame Forger in video material. Figure 1 may display the system architecture of the suggested technique for identifying inter frame forger films. The suggested method included employing the YOLO face detector to find faces within video frames, as illustrated in Figure 1. The spatial-visual traits that are helpful for discrimination are extracted using the InceptionResNetV2 CNN model. These traits aid in the analysis of the visual artefacts found inside the video's frames, and the analysis's findings are put into the XGBoost classifier so it can discriminate between real and inter frame forger movies. The description that follows is thorough enough to explain the proposed strategy[12].

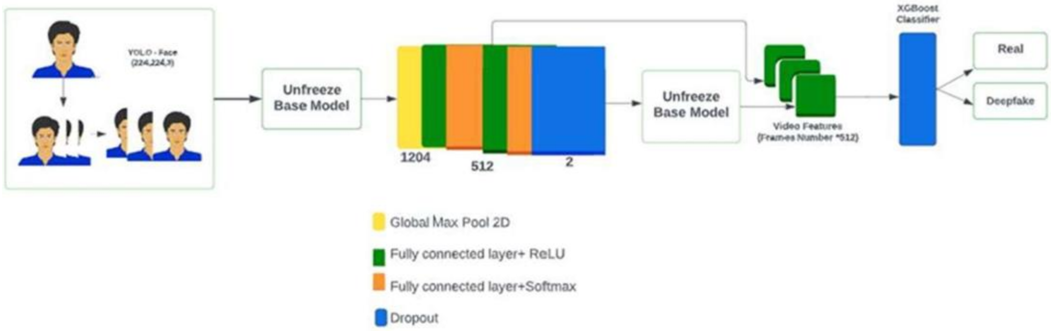

Figure 1: Method for Detecting Inter frame forgery in Videos

2. Related Work

People often believe that video clips may provide more forensic proof than still images. As a consequence, because it is crucial evidence, surveillance film is often utilised in the course of an investigation. On the other hand, the digitising feature makes it simple to edit surveillance film. It is straightforward to alter a digital movie without leaving any visible traces when using video editing software like Adobe Premiere [13]. As a consequence, the field of computer science

known as "digital video forensics," which investigates whether or not digital movies can be trusted, has become crucial and intriguing for modern research. investigated the advantages of such authentication for the legal and journalistic domains of the media business and proposed a system for authenticating and validating material across a range of media platforms. On the other hand, it wasn't always feasible to implant watermarks into the films that were studied owing to technological restrictions. Most of the videos had this situation. As a result, the current research has given considerable weight to non-previous knowledge-based detection techniques, such as the identification of forgery evidence[14]. This is due to the significance of the current situation. A prime example of this is knowing how to spot the telltale signs of fraud. The two factors that might be utilised to distinguish between real and fake surveillance film are the source's truthfulness and the content's validity. The phrase "source authenticity," which relates to the above-described procedure, refers to the analysis of the video to ascertain "where it originated from" and "how it came from." To analyse each of the steps of the acquisition process, several alternative methodologies have been developed. It is established whether or not the video has been manipulated in any way throughout the content authenticity testing process. Depending on the container, the video file that the camera produces will have a certain extension[15]. With the aid of the metadata, the container will specify the file's structure. The codec, which is an encoded stream of bytes that makes up the video's content, is the most crucial element in deciding the overall quality of the video. For instance, a well-known container may hold a variety of various codecs. H.264 data is included in a MOV file. A video is just a collection of still photos that have been arranged into frames and joined together in a certain sequence[16]. It is a collection of succeeding frames, often referred to as GOPs, that are organised in a three-dimensional plane and have a temporal dependency. Each each frame recorded by a single camera movement makes up a shot. One or more photos combine to form a scene that makes sense as a whole. One of the techniques used to encode a video is quantization, and it has a noticeable impact on the sequence itself. The act of deliberately editing or manipulating a digital video for fabrication purposes is referred to as "digital video forging." Its implications vary depending on the context and environment in which it is used[17]. It has a big impact, especially in the entertainment, political, and medical fields where it's often used to tarnish popular figures, hide or make up important details, and either lie about or cover up actual occurrences. Our daily lives are significantly and favourably impacted by how often we view videos on different social media platforms, including Facebook, WhatsApp, YouTube, and other news channels. "Being seen is no longer being believed," The integrity and validity of the footage that is being shown cannot be simply believed to be true [18]. In the meanwhile, methods for detecting manipulated microscopic portions inside frames are known as video intra-frame forgeries. Area duplication, area deletion, and area insertion are some of these changes. Recent advancements in video editing software have made it possible to copy and paste 3-D parts of recordings and quickly change their brightness, geometry, and other comparable attributes. It's possible for the 3-D parts to be little 3-D fragments inside subsequent frame sequences or whole consecutive frame sequences. Since these fake movies are becoming a prevalent technique used in video tampering, it could be difficult to detect them with the naked eye. A lot of effort has been put into creating algorithms that can recognise this sort of fake footage. Additionally, there are frame-level changes that repeat or conceal objects in the movie. These easy adjustments may be used while changing the video's content[19]. They would, however, produce fake movies that were difficult to view, especially with naked eyes. Since the advent of contemporary computer and multimedia technology, digital video has become the dominant form of network communication due to its accessibility, mobility, and substantial information content. It has evolved into a crucial piece of information in a variety of crucial circumstances, including the media, politics, insurance claims, defence, and legal matters, among many other crucial issues and themes. However, some amateurs are able to alter video footage with relative ease because to the accessibility of robust multimedia editing software, and it may be challenging for experts to tell certain fake movies from the real thing[20]. This is because some of the impostor videos were produced utilising sophisticated multimedia editing tools. This is the cause of what happened. As a consequence of a number of these various issues, some individuals have begun to question the veracity of the digital video recording. As a result, there is an urgent need for trustworthy forensic technology that can attest to the veracity, accuracy, and authenticity of video data. In terms of avoiding deceptive video manipulation from harming the general population and maintaining social order and peace, this technique has a lot of real-world applications[21].

The method that has been created makes it simple to spot inter frame forger in video material. The system architecture of the suggested technique for identifying inter frame forger films. The suggested method included employing the YOLO face detector to find faces within video frames. The spatial-visual traits that are helpful for discrimination are extracted using the InceptionResNetV2 CNN model. These traits aid in the analysis of the visual artefacts found inside the video's frames, and the analysis's findings are put into the XGBoost classifier so it can discriminate between real and inter frame forger movies. The description that follows is thorough enough to explain the proposed strategy[22].

3. Proposed Method

We suggested a strategy for identifying Inter frame forger movies that uses spatiotemporal information. This is due to the fact that several independent trials have shown the model's ability to recognise the spatiotemporal elements that are crucial to the dataset used in the Deep Fake Detection Challenge (DFDC). This is the cause of this situation. There is neither a general nor a particular standard to follow when judging the size of the provided image alone. A larger input size might be used, but doing so would need more processing power. The medical centre often uses the following picture sizes: 100 by 100, 128 by 128, 256 by 256, 299 by 299, and 300 by 300. The processing speed and the size of the input will thus constantly be in conflict with one another. 240 pictures were chosen for this research because they are even, which makes cropping and scaling processes easier to complete, and because they are large enough to allow for the identification of all pertinent properties. An LSTM layer and a time-distributed layer were applied after using a CNN model to recast the Inter frame forger detection task as a binary classification challenge. Finally, a time-distributed layer was used to assess the model's correctness. The output of the LSTM layer is transmitted into the thick layers, as shown in Figure 2. The authors hope to be able to distinguish between authentic recordings and copies, some of which may include a signal that is too faint to be picked up without an amplifier. The writers' goal is to do this. The spatio-temporal model, on the other hand, will be able to spot even the tiniest changes in the data.

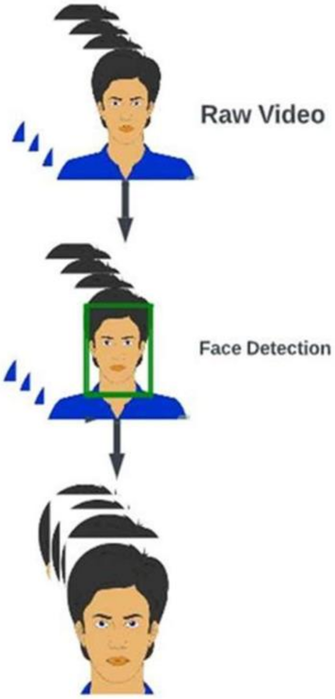

Figure 2: Proposed Method

4. Dataset Pre-processing

The whole DFDC dataset, which has over 470 terabytes of data, was utilised for this experiment. In order to balance the quantity of processing resources available with the need for more frames, the patio-temporal technique has been used, and as a consequence, 30 frames per video have been considered. Each video had an average of 300 frames, it was revealed after looking through the collection. It was discovered that this was the case. The face and its immediate surroundings are the main regions that inter frame forger techniques target. As a consequence, as shown in Figure 3, the face, which was chosen as the region of interest, was evaluated before being retrieved from the film. Frame-by-frame face retrieval has been used to manipulate video, which has led to low-level artefacts from face manipulation further emerging as temporal distortions with discrepancies across frames. To get the intended result, this has been done. A face extractor that could operate rapidly and provide trustworthy results was required due to the massive

amount of data. A decent middle ground between the two possibilities was provided by the Mobile Net SSD solution. In order to identify any malformations in that region, it was required to add an extra margin of 35% all the way around the faces.

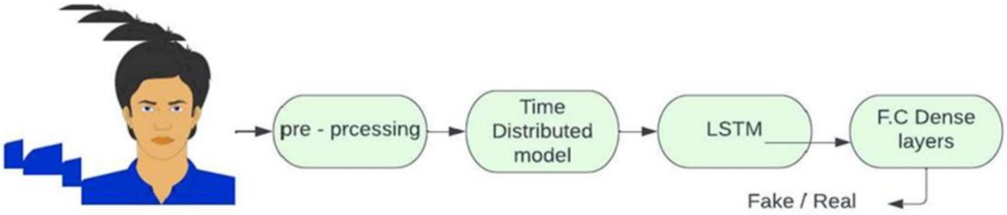

Figure 3: Dataset Pre-processing

5. Methodology

The method that we examine is based on a two-step procedure that comprises the extraction and categorization of forensics-based data as separate but connected activities. A modified form of deep convolutional neural networks is used in this strategy. GoogLeNet and ResNet serve as the basis during the classification phase; however, two additional filters serve as the foundation during the feature extraction phase. (CNN).

6. Forensics-Based Filters

The filters that we used in this research were created with the goal of giving the participants output maps that they could assess for aesthetic quality. The names Q4 and Cobalt, respectively, have been allocated to these filters. The Q4 filter is used to analyse the breakdown of the picture using the discrete cosine transform (DCT). Regardless of whether the conversation is about an I frame, a P frame, or a B frame, it applies to each and every video frame. Each of these image blocks is then modified using the two-dimensional discrete cosine transform in order to combine them into a block that is the same size as the blocks that were first segmented from the frame after each frame has been divided into N N blocks. (DCT). This is done so that the coefficients may be recognised based on how often they show up in the image. N often equates to 8. Larger coefficients, as opposed to the starting coefficient (0, 0), which represents information with a low frequency, reveal information with a higher frequency. We further analyse the brightness data included inside the Y channel after the JPEG compression procedure in the YCbCr colour space has been finished.

The second kind of filter we use in this specific situation is the cobalt filter. In this study, the analysis and comparison of the original video and a modified version that was re-quantized using MPEG-4 at a different quality level (and a correspondingly different bit rate). It is quite likely that this segment of the movie underwent MPEG-4 quantization at a different level than the rest of it if there is a (small) chunk of the original video that was made from a different stream. This is due to the fact that only one level of quantization was used to encode the majority of the film. This is done in order to achieve different levels of compression with a wide variety of different quantization levels. It's possible that a global strategy won't be able to locate this region even after looking for a lot of quantizations. The following simple theoretical framework can be used to understand how the cobalt filter works: In order to show the changes that have occurred as a result of the process, we create an error film after requantizing the video and figuring out the values for each pixel. According to the theory, there shouldn't be nearly as much inaccuracy in the requantized video as there was in the original if we use the exact same parameters as we did for the first video [23]. This is the result that is obtained when the video is requantized using the same parameters as the original video. How accurate the video is will be determined by how much it deviates from the original material, and the opposite is also true. In order to produce cobalt, researchers looked at a process known as the "compare-to-worst" approach. This strategy specifies that the lowest quality will be used for comparison if constant quality encoding is applied. This technique states that the lowest bit rate will be used for comparison if constant bit rate encoding is employed. Both of these conclusions are predicated on the idea that the encoding won't be changed in any way while the procedure is being done. If the quantization history of the source video is not homogenous, a dramatically contrasted film of mistakes is produced [24].

7. Filter Output Classification

Images with an RGB colour space are produced when the two filters are used together. We concluded that the problem should be solved as a visual classification task since the filter maps were created from the start with the intention of being visually inspected by a human professional. This directly led us to the conclusion that the issue should be resolved by some kind of visual classification work. As a result, we were able to combine the maps with convolutional neural networks that had been previously trained to differentiate between different types of pictures. We modify two separate instances of each of GoogleNet [18] and ResNet [5] to make them acceptable for the needs of our inquiry. Both of these networks have been pre-trained on the ImageNet classification assignment. These two networks are the result of Google's development. The 224 × 224 pixel default input size of the CNNs is taken into account when scaling the image outputs from the filtering technique. This makes sure the results seem as excellent as they can within the conditions. The filters that we use were created with human visual interpretation in mind, in contrast to other forensics-based techniques, such those in which rescaling could cause the loss of sensitive information. Rescaling shouldn't be problematic as a consequence, just as it shouldn't be problematic in any other categorization task [25].

The initial round of evaluations of the proposed method were based on within-dataset research using five-fold cross-validation. We used both the merged version of the two NIST Challenge datasets (Dev1 and Dev2) as well as each dataset separately for our experiments. Table 1 (shown below) contains a summary of the results.

Dataset	Testing	Filter-DCNN	Accuracy	MAP	MP020
Dev 1	Dev 1	cobalt-gnet	**0.6833**	0.7614	**-**
Dev 2	Dev 2	cobalt-gnetq4-guet	0.8791 **0.8843**	**0.9568** 0.9472	**0.8200** 0.7900
DevI+Dev2		cobalt-gnetq4-guet	**0.8509** 0.8408	0.9257 **0.9369**	0.**9100** **0.9200**

Dev1	Dev2	frameDifference-gnet frameDifference-resnet	0.6942 **0.7190**	**0.8553** 0.8286	**0.9000** 0.8500
	FVC	q4-resnet	**0.6029**	**0.6947**	0.7000
Dev2	Dev1	q4-gnet	**0.6500**	0.7191	**0.7000**
	FVC	q4-gnet rawKeyframes-gnet	**0.6177** 0.5147	**0.6558** 0.6208	0.7000 **0.7000**
Dev1 + Dev2	FVC	q4-gnet	**0.6471**	**0.7114**	0.7000

Table 1: Filter Output Classification

The findings show that Dev1 provides a difficulty that is by far the most challenging of all the filters and models that were taken into account. When using Dev2, the accuracy score may be anywhere between 0.79 and 0.88, however it often falls between 0.58 and 0.68.Patterns that have been previously identified in the data are shown by the Mean Average Precision. It is important to remember that the MP@20 measure is not appropriate in this case since just a small number of samples were used in the cross-validation of Dev1. In light of this, it is essential that this information be taken into account (the test set would always contain less than 20 items). Thanks to the merging of the two datasets, we now have the biggest cross-validation dataset collection that is even remotely achievable; based on this collection, we may be able to deliver the most precise future projections. Although by a very tiny margin, the combined impacts of Dev1 and Dev2 provide results that are noticeably better than those created by Dev2 alone. Even if Dev2 is inferior to MP@20, one reason for this discrepancy may be because Dev2 is considerably smaller. The results as a whole provide cause for cautious optimism since the Mean Average Precision for the Dev1 + Dev2 set was determined to be 0.94. The majority of the time, Google Net seems to perform better than ResNet. The two filters seem to be about equal in terms of performance, with Cobalt sometimes beating Q4 and vice versa.

8. Clustering Results of Original Video

The movie was recorded as a series of sixteen still images, each with an initial width and height of 480 pixels and 640 pixels, respectively. The figure 6a displays these stills. Figure 6c displays the r1 clustering's findings. Figure 6c, which was created by applying Equation to the aforementioned data, displays the clustering findings. represents the S1 cluster, identifies the cluster centroids, and is representative of the S1 cluster. The locations of each cluster's geographic centres were discovered to be, respectively, 1.0389 and 1.0118. If the video is real, we find that both centroids' values were pretty close to one if the video is real.Original video frames and clustering results. (1) The curve of r1; (2) clustering results. 0 2 4 6 8 10 12 14 0 0.2 0.4 0.6 0.8 1 1.2 1.4 frame number the value of r1 r1-original 1 2 3 4 5 6 7 8 9 0 2 4 6 8 10 12 14 S1 S2 centroid' Marks the cluster centres and indicates that the S2 cluster is being represented. The centroids of the two clusters were discovered to be, respectively, 1.0389 and 1.0118. In the event that the video is authentic, we find that both centroids' values were extremely close to 1.

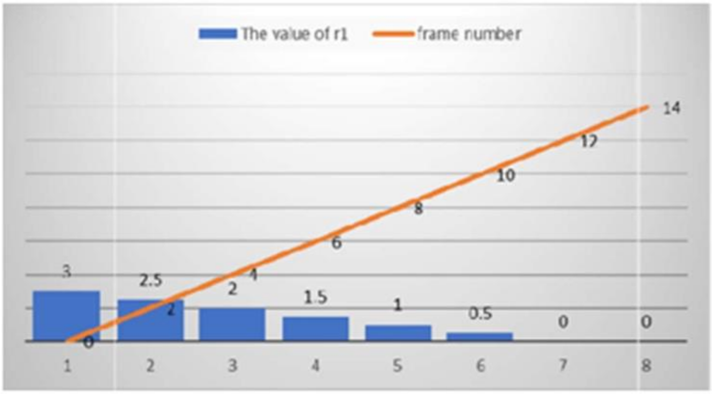

Chart 1: The curve of r1

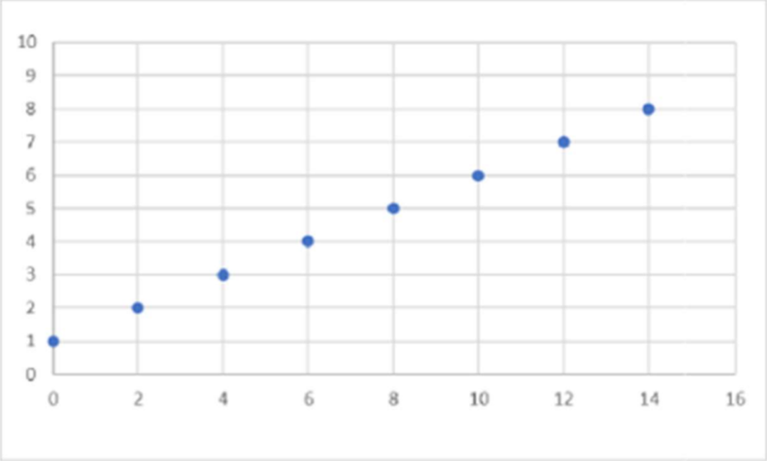

Chart 2: clustering results

9.Clustering Results of Forged Video by Frame Insertion

The video's earliest eight frames in a row are regarded as its original frames, while its most recent eight frames in a row are regarded as its insertion frames. There are two peaks that occur in the seventh and eighth frames because of the weak connection between the eighth and ninth frames. The centre of the graph is where these peaks are positioned. If we are successful in identifying the strange locations, we will be able to prove that the movie was purposefully made. At the location of the anomalous point, the manipulation is done. The clustering analysis's findings show the outliers that were discovered in Cluster S1. The centroids of the two distinct clusters were found to be 2.5144 and 1.0413, respectively. We find that the centroid of S1 does not, as previously thought, correlate to 1 due to the change in methodology.

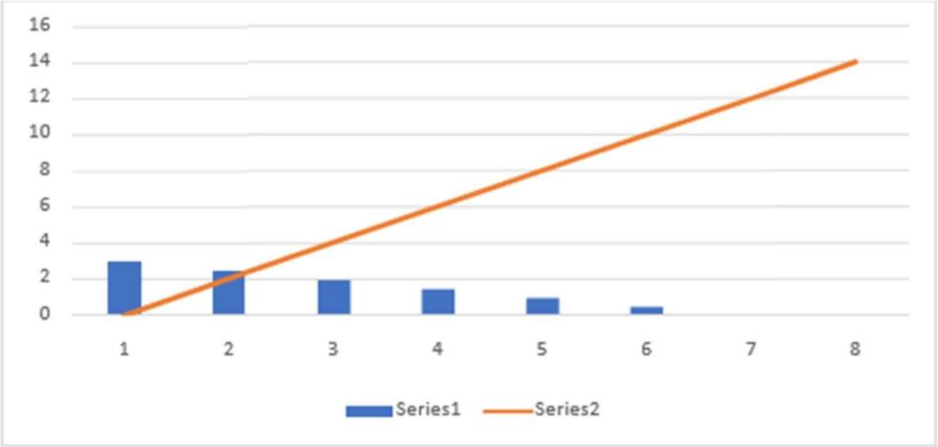

Chart 3: The curve of r1

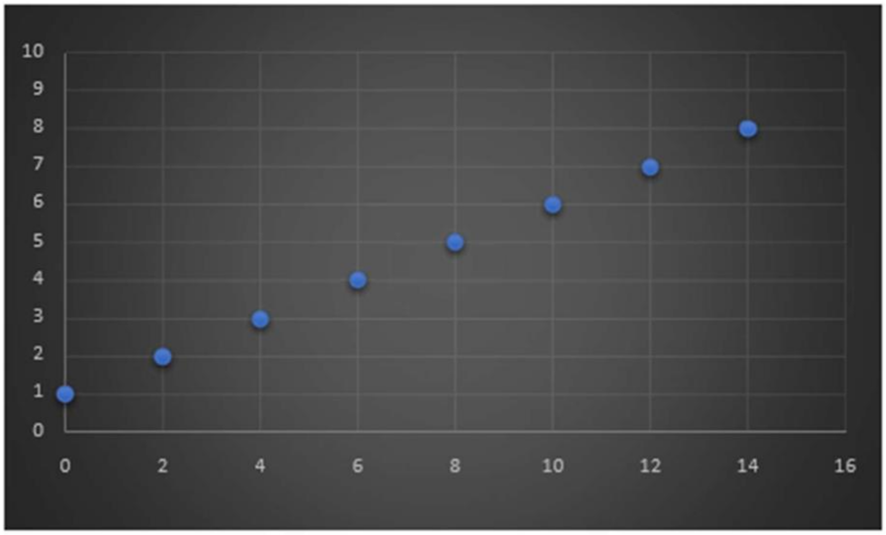

Chart 4: clustering results.

Digital movies are often saved in a compressed format since they need a lot of storage space. The effectiveness with which they may be both preserved and transmitted is improved as a result. Each each frame must be deleted and altered in order to deceive the user before any kind of tampering action can be taken. Before doing any kind of procedure, this must be completed. The damaged movie is reconstructed using the changed frames, resulting in a twofold decrease in file size, since it is almost impossible to save a video without first putting it though some sort of compression every time the video is saved. The discovery of signs of double compression in video sequences served as the foundation for the early advancements in the 1990s in the field of video inter-frame forgery detection. In order to provide the groundwork for further developments in the area, this was done. In spite of this, double compression will continue to occur whenever a video is aired, uploaded, downloaded, or even viewed. This shows that there is a chance that inter-frame falsification was not employed in the making of the recompression signs were included in it.

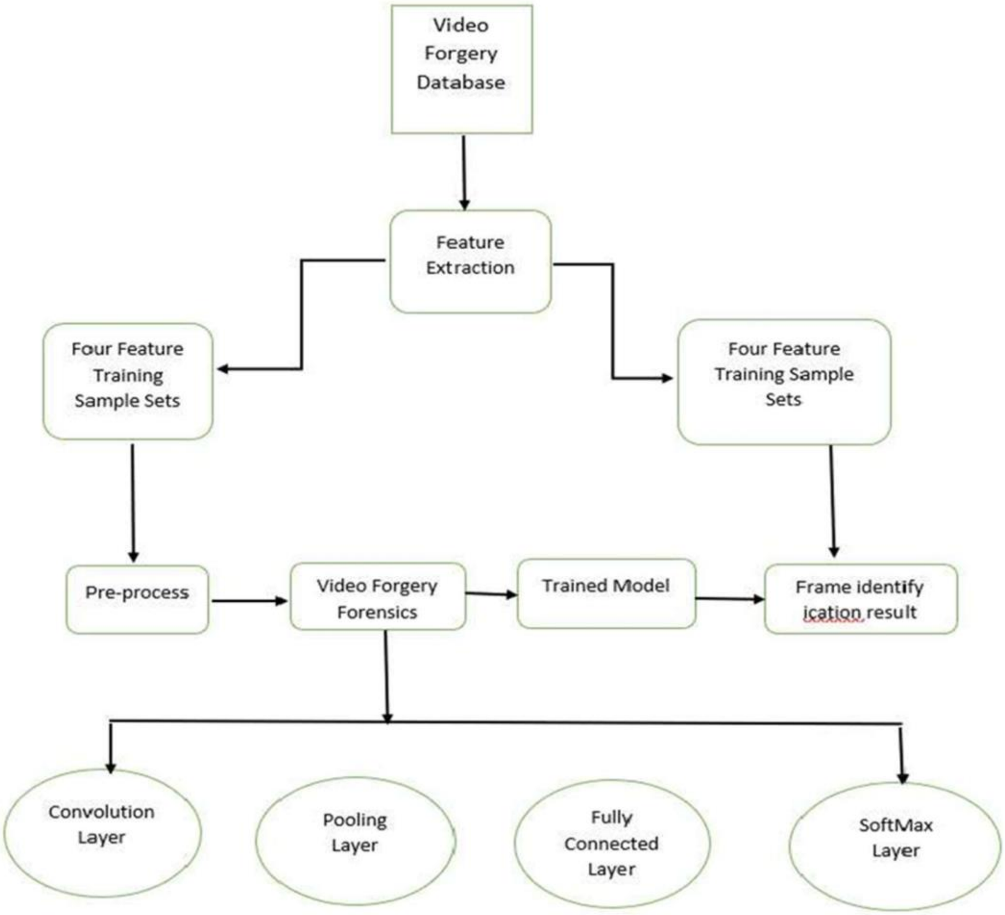

Figure 4: Video object forgery detection

10. Threshold Decision

The normal and abnormal points have been combined into two different groups using k-means clustering. The basis for detecting whether or not the video has been manipulated with is the centroid's value. Using the approach that was provided, we tested the original sub-database, and we used all 599 of the original videos to analyse the cluster centre for S1. The x-axis displays the centroid value of the cluster, while the y-axis displays the frequency at which each value occurs.

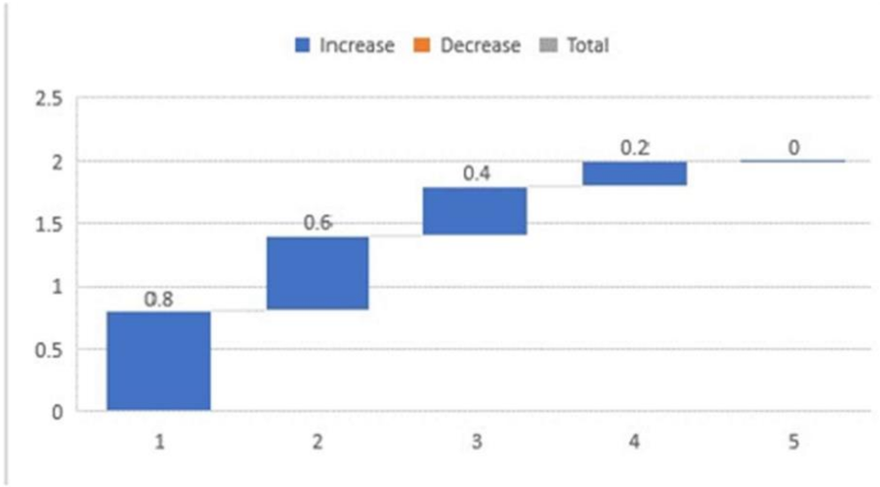

Chart 5: Threshold Decision

11. Evaluation Metrics and Method Assessment Procedure

We use six performance indexes while evaluating the strategy's effectiveness. The accuracy score (F1 score), the location precision score, the true positive rate (TPR), also known as recall, the true negative rate (TNR), and the location precision score are the six metrics. In addition, recall is another name for the true positive rate (TPR) (TNR). Accuracy is the term used to describe a detection's overall effectiveness. The accuracy and recall values are weighted to get the F1 score. It is conceivable to interpret the outcome in this way. The percentage of real videos with proper localization relative to all real videos that have been exposed as fakes is known as precision of location:

12. Test results

Negative samples were generated to run tests on after each movie in the testing dataset. The result would be video snippets. This sample data was categorised using these trained models. The highest fcon value of a video ultimately determines its veracity. If max (fcon(i)) above the threshold, I is less than 1, and T is larger than 1, else it is fake. The Threshold was found to have the value of 0.5 in each of our experiments.

We based our performance evaluation on the following standards:

In this article, the following measurements are used: False Positive (FP): It was determined that the original video was a fab ication. True Positive (TP): The fake was recognised for what it was: a fake. False Negative (FN): manufactured video declared real; True Negative (TN): original video judged authentic. sensitivity is another name for the proportion of true positives. value(TPRV); False Positive Rate value (FPRV) and Detection Accuracy Rate (DAR) as follow:

Datasets	DAR (%)	FPRV (%)	TPRV (%)
Dataset_1	**96.668**	**6.663**	**96.773**
Dataset_1	84 .163	36.667	91.102
Dataset_1	83.325	13.328	82.211

Dataset_1	95.000	13.326	98.774
Dataset and Dataset 4	**99.164**	**3.321**	**100**

Table 2: Detection Accuracy and True, False Positive Rate models on Dataset1& 4

True Positive Rate value:

$$TPRV = \frac{TP}{P} = \frac{TP}{TP + FN}$$

False Positive Rate value:

$$FPRV= \frac{FP}{N} = \frac{FP}{TN + FP}$$

Detection Accuracy Rate:

$$DAR = \frac{TP + TN}{N + P}$$

to assess the usefulness of the many factors that went into creating the training dataset. On both the individual and integrated features of the product, we have received further training. All of the studies were carried out and the results were presented using the indicated method, which is based on the pre-trained MobileNetv2 model.

The results of the suggested method, which is based on MobileNetv2, after it has been retrained on datasets made up of various properties.

We will compare the effectiveness of the model that utilises transfer learning to the effectiveness of the model that is trained from scratch in order to ascertain if the model is capable of transfer learning on the ImageNet database and whether it can detect video inter-frame forgeries. We experimented with two alternative models—MobileNetv2 and Resnet18—on the same target dataset by first training them from scratch and then retraining them using a model that had already been trained.

Comparing the results of starting from scratch on Dataset 1 and using transfer learning to train on Mobile Netv2 and ResNet 18 models.

Methods	**DAR (%)**	**FPRV(%)**	**TPRV (%)**
Mobilenetv2-Transfer learning	96.662	6.660	97.774
Mobilenetv2-Trained scratch	84.160	33.325	90.000
ResNet 18-Transfer learning	97.40	3.326	97.772
ResNet 18-Trained scratch	93.324	16.667	96.667

Table 3: Mobile Netv2 and ResNet 18 models on Dataset 2

13. Conclusion

Today, the great majority of people have smartphones, and many of them have cameras built in. This is partly due to the quickly growing hardware market, particularly the introduction of cameras, which were utilised for surveillance in a variety of settings, including traffic, residences, workplaces, schools, etc. Thanks to the invention of cameras, which were utilised for monitoring everywhere from traffic to houses to workplaces to schools, the great majority of people today now possess cameras. Cameras were formerly mostly employed for security reasons in establishments like offices, but today they are a standard feature of cellphones. The number of individuals worldwide who have access to the internet has dramatically increased over the course of the previous few years. This makes it easy to make audio and video recordings anywhere, edit them as needed, and then share them right away online. One piece of evidence in this inquiry that shouldn't be disregarded is the first video. The only processes that can currently be utilised to verify movies, however, are ones that are either very slow or very wasteful in their functioning. The goal of this study is to establish a method for spotting video inter-frame forgeries that may be used in further studies, based on the most current advancements in CNN model technology. This method's accuracy varies from 97.5% to 99.17%, and it has produced results that are both fair and encouraging. Tests performed on the same dataset show that the suggested strategy is much more efficient than the methods presently in use. The evidence that has been provided supports this statement. Our objective is to do more research and provide recommendations for a suitable CNN architecture with fewer parameters and a lower overall level of complexity. As a result, we will be able to recognise and classify the many types of video forgeries that are now in use. We demonstrated our effort to combine deep learning models for visual classification with video forensics filters, which were first created to be visually examined by experts. In the past, these filters were developed so that experts could evaluate them visually. The use of these filters was anticipated from the outset. We used two alternative deep network topologies to evaluate the performance of two forensics-based filters. We observed that the suggested method's performance was on par with, or even worse than, that of a number of filters that are regarded as state-of-the-art after training and testing it on videos that were comparable to one another. One of the suggested filters performed considerably better than the others when it was tested on datasets that were distinct from those used for training. That's what happened. As a direct result of this, we came to the conclusion that the suggested method had some promise. This is a compelling and insightful conclusion that might help to clarify the significance of such an automatic video verification method, particularly for content that can be accessed online and on social media.

References

1. Chen, S., Tan, S., Li, B., Huang, J.: Automatic detection of object-based forgery in advanced video. IEEE Trans. on Circ. Syst. Video Technol. 26(11), 2138–2151 (2016)
2. D’Amiano, L., Cozzolino, D., Poggi, G., Verdoliva, L.: Video forgery detection and localization based on 3D patchmatch. In: IEEE International Conference on Multimedia Expo Workshop (ICMEW) (2015)

3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

4. Labartino, D., Bianchi, T., Rosa, A.D., Fontani, M., Vazquez-Padin, D., Piva, A.: Localization of forgeries in MPEG-2 video through GOP size and DQ analysis. In: IEEE International Workshop on Multimedia and Signal Processing, pp. 494–499 (2013)

5. Katsaounidou, A.; Dimoulas, C.; Veglis, A. Cross-Media Authentication and Verification: Emerging Research and Opportunities; IGI Global: Hershey, PA, USA, 2018; pp. 155–188.

6. Arab, F.; Abdullah, S.M.; Hashim, S.Z.M.; Manaf, A.A.; Zamani, M. A robust video watermarking technique for the tamper detection of surveillance systems. Multimed. Tools Appl. 2016, 75, 10855–10885.

7. Chen, S.; Pande, A.; Zeng, K.; Mohapatra, P. Live video forensics: Source identification in lossy wireless networks. IEEE Trans. Inf. Forensics Secur. 2015, 10, 28–39. [CrossRef]

8. Amerini, I.; Caldelli, R.; Del Mastio, A.; Di Fuccia, A.; Molinari, C.; Rizzo, A.P. Dealing with video source identification in social networks. Signal Process. Image Commun. 2017, 57, 1–7.

9. Li, Z.H.; Jia, R.S.; Zhang, Z.Z.; Liang, X.Y.; Wang, J.W. Double HEVC compression detection with different bitrates based on co-occurrence matrix of PU types and DCT coefficients. In Proceedings of the ITM Web of Conferences, Guangzhou, China, 26–28 May 2017; p. 01020.

10. He, P.; Jiang, X.; Sun, T.; Wang, S. Double compression detection based on local motion vector field analysis in static-background videos. J. Vis. Commun. Image R 2016, 35, 55–
66. [CrossRef]

11. Zheng, J.; Sun, T.; Jiang, X.; He, P. Double H.264 compression detection scheme based on prediction residual of background regions. In Intelligent Computing Theories and Application; Springer: Cham, Switzerland, 2017; pp. 471–482.

12. Abdullahi, A., Bagiwa, M. A., Roko, A., & Buda, S. (2022). An Inter-Frame Forgery Detection Technique for Surveillance Videos Based on Analysis of Similarities. SLU Journal of Science and Technology, 4(1&2), 15–26. https://doi.org/10.56471/slujst.v4i.265

13. Fadl, S. M., Han, Q., & Li, Q. (2019). Inter-frame forgery detection based on differential energy of residue. IET Image Processing, 13(3), 522–528. https://doi.org/10.1049/iet- ipr.2018.5068

14. Hu, J., Liao, X., Liang, J., Zhou, W., & Qin, Z. (2022). FInfer: Frame Inference-Based Deepfake Detection for High-Visual-Quality Videos. Proceedings of the AAAI Conference on Artificial Intelligence, 36(1), 951–959. https://doi.org/10.1609/aaai.v36i1.19978

15. Huang, T., Zhang, X., Huang, W., Lin, L., & Su, W. (2018). A multi-channel approach through fusion of audio for detecting video inter-frame forgery. Computers and Security, 77, 412–426. https://doi.org/10.1016/j.cose.2018.04.013

16. Kingra, S., Aggarwal, N., & Singh, R. D. (2016). Video inter-frame forgery detection:

A survey. Indian Journal of Science and Technology, 9(44), 1–9. https://doi.org/10.17485/ijst/2016/v9i44/105142

17. Kingra, S., Aggarwal, N., & Singh, R. D. (2017). Inter-frame forgery detection in H.264 videos using motion and brightness gradients. Multimedia Tools and Applications, 76(24), 25767–25786. https://doi.org/10.1007/s11042-017-4762-2

18. Li, Q., Wang, R., & Xu, D. (2018). An inter-frame forgery detection algorithm for surveillance video. Information (Switzerland), 9(12). https://doi.org/10.3390/info9120301

19. N., S. K., & Chennamma, H. R. (2015). A Survey On Video Forgery Detection. IX(Ii), 17–27.

20. Oraibi, M. R., & Radhi, A. M. (2022). Enhancement Digital Forensic Approach for Inter- Frame Video Forgery Detection Using a Deep Learning Technique. Iraqi Journal of Science, 63(6), 2686–2701. https://doi.org/10.24996/ijs.2022.63.6.34

21. Patel, J., & Sheth, R. (2021). an Optimized Convolution Neural Network Based Inter-Frame Forgery Detection Model-a Multi-Feature Extraction Framework. 2570–2581. https://doi.org/10.21917/ijivp.2021.0364

22. Pu, H., Huang, T., Weng, B., Ye, F., & Zhao, C. (2021). Overcome the brightness and jitter noises in video inter-frame tampering detection. Sensors, 21(12), 1–21. https://doi.org/10.3390/s21123953

23. Shi, Y. Q., Kim, H. J., & Perez-Gonzalez, F. (2013). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7809 LNCS(October). https://doi.org/10.1007/978-3-642-40099-5

24. Xia, Z., Qiao, T., Xu, M., Wu, X., Han, L., & Chen, Y. (2022). Deepfake Video Detection Based on MesoNet with Preprocessing Module. Symmetry, 14(5), 1–14. https://doi.org/10.3390/sym14050939

25. Zhong, J. L., Gan, Y. F., Vong, C. M., Yang, J. X., Zhao, J. H., & Luo, J. H. (2022). Effective and efficient pixel-level detection for diverse video copy-move forgery types. Pattern Recognition, 122, 108286. https://doi.org/10.1016/j.patcog.2021.108286