IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 February 2023, accepted 12 April 2023, date of publication 17 April 2023, date of current version 17 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3267743

RESEARCH ARTICLE

A New Dataset for Forged Smartphone Videos Detection: Description and Analysis

YOUNES AKBARI[1], AL ANOOD NAJEEB[1], SOMAYA AL MAADEED[1], (Senior Member, IEEE), OMAR ELHARROUSS[1], FOUAD KHELIFI[2], (Member, IEEE), AND ASHREF LAWGALY[2]

[1]Department of Computer Science and Engineering, College of Engineering, Qatar University, Doha, Qatar
[2]Department of Computer and Information Sciences, Northumbria University, NE1 8ST Newcastle Upon Tyne, U.K.

Corresponding author: Somaya Al Maadeed (S_alali@qu.edu.qa)

This work was supported in part by the National Priorities Research Program (NPRP) from Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP12S-0312-190332, and in part by the Qatar National Library (QNL).

ABSTRACT The advancement of Internet technology has significantly impacted daily life, which is influenced by digital videos taken with smartphones as the most popular type of multimedia. These digital videos are extensively sent through various social media websites such as WhatsApp, Instagram, Facebook, Twitter, and YouTube. The development of intelligent and simple editing tools has favoured the transformation of multimedia content on the Internet. As a result, these digital videos' credibility, reliability, and integrity have become critical concerns. This paper presents a video forgery (Copy-move forgery) dataset in which 250 original videos are manipulated mainly by two forgery techniques: Insertion and Deletion. Inserting transparent objects into the original video without raising suspicion is one type of manipulation performed. Another type of forgery presented on the dataset is the removal of objects from the original video without notifying the viewers. The videos were collected from five different mobile devices, namely, IPhone 8 Plus, Nokia 5.4, Samsung A50, Xiomi Redmi Note 9 Pro and Huawei Y9-1. The forged videos were created using a popular video editing software called Adobe After Effects as well as usage of other software such as Adobe Photoshop and AfterCodecs. Since the source of the videos is known, PRNU-based methods can be suitable for applying to the dataset. Experiments were performed using classical and deep learning methods. The results are recorded and discussed in detail, showing that improvements are essential for the dataset. Furthermore, the forged videos of this dataset are comparatively large when compared to other datasets that performed copy-move forgery.

INDEX TERMS Dataset, video, mobile devices, copy-move forgery, deep learning.

I. INTRODUCTION

Due to the rapid advancement in the multimedia technology, multiple manipulating software's are being introduced which are freely available to the users. The majority of the users are dependent on multimedia content on a daily basis for various tasks. Transferring of contents has become common in the present scenario due to less restrictions. The use of cell phones has rapid development in the preceding century as a result of their cost-effectiveness, functionality, and convenience of use [1], allowing for the creation of digital audiovisual without any limitations of time, objects, locations, or network connections [2]. They have the potential to provide crucial information for criminal prosecution and forensic investigations [1]. These types of investigations could be useful in scientific disciplines such as medical, law, and surveillance systems, where the authenticity of photos and recordings is essential. Various video manipulating applications and software such as Adobe Premiere Pro, Photoshop, After Effects, Paintshop Pro, GIMP, and CorelDraw are being employed to perform different modifications in the multimedia content in a more convenient manner. Video forgery

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

can be easily done for manipulation and falsification using different editing software and modern smartphones. A multimedia forgery is the deliberate alteration of a visual media for misrepresentation, and it may be difficult for humans to determine the genuineness of those video recordings with bare eyes. As a result, it is crucial to evaluate and determine whether visual content is genuine or altered before using it as proof. To investigate the integrity and reliability of multimedia data, digital forgery detection mechanisms are necessary. The diagnosis of digital video forgery is a process that determines whether the digital video contents have been intentionally manipulated [3].

Forgery detection techniques can be classified into two groups, which are the active and passive mechanisms [3]. In active mechanism, the forged data could be extracted and the data is hidden in the form of a digital signature and watermarks to examine the morality and accuracy of the video whereas, the passive mechanism performs efficiently when the secret data is not hidden in videos. In this scenario, the detection of manipulated data is difficult in active approach. Therefore, in the past few years, the passive technique is gaining more attention in the research field.

There are different kinds of video forgery techniques which are categorized into two sections. They are Intra-frame forgery and Inter-frame forgery. These categories of forgery techniques can be accomplished using the finest tools in video editing such as Adobe After Effects, Adobe Premier Pro, and Adobe Photoshop. In Intra-frame forgery, the actual components of the individual frames are modified. This type of forgery is also known as spatial video manipulation. Various kinds of Intra-frame manipulation are in use and some of them are Copy-Move forgery, Splicing and Upscale Crop. Copy-Move forgery is the most popular forgery executed on graphical videos and images. Any objects can be incorporated and removed from the videos by the intruders. Simultaneously, it may be used to replicate objects in a video by cloning a section of the video frame and inserting it to a different spot in the same or different frame of the movie. Copy-move forgery also known as inpainting forgery is a technique for deleting specific items from digital photographs or videos and replacing them with information that resembles the background. In splicing based video forgery, a part of video frame is selected and inserted to another video frame. Merging of the two video frames are mostly happening in this method. In an upscale crop, the exterior section of the video frame is trimmed out to eliminate particular areas or features. Changing the arrangement of frames in videos in a variety of ways is carried out in inter-frame forgery. This type of forgery is also known as a temporal forgery. There are different types of inter-frame forgery which are in use. Some of them are Frame deletion, Frame duplication, Frame insertion, and Frame shuffling or replication. Frame deletion is the removal of certain frames in a video which is considered as an illegal activity. Frame duplication is the purposeful repetition of video frames. A similar type of frame deletion forgery known as frame mirroring is used in [4]. A version of some of the frames from the video sequence is duplicated and added at random positions in the same video is implemented in frame mirroring. Frame insertion is the addition of the frames taken from the same or different videos at various locations. Frame shuffling or replication is the mixing or shuffling of the frames of the same video.

There are several video datasets for forgery detection that have been used for various purposes in recent years. Although most videos on the Internet are currently taken with smartphones, to our knowledge there has not been a video dataset where the videos were recorded with smartphones and the forgery was done on the videos recorded with the same smart phones. Therefore, this paper introduces the video dataset captured using mobile phones for the purpose of forgery where the major form of forgery is object insertion and deletion without giving a suspicion to the users that the videos are forged. A total of 250 videos were collected using 5 different smartphone brands, and 50 videos (25 insertions and 25 deletions) from each device is manipulated to form the forgery videos. The dataset's structure is discussed and given in detail in this paper. The use of footage from multiple smartphones is a distinctive feature in this dataset, as other datasets use different single-lens reflex cameras and other closed-circuit television camera, which are difficult to mobilize and expensive to purchase compared to smartphones [5]. Our knowledge of the source of the videos covers a wide range of methods, from Photo Response Non Uniformity (PRNU) methods to Deep Learning methods [6]. In the present scenario, smartphones are replacing digital single-lens reflex (DSLR) cameras with high-quality pictures and videos in addition to cloud backup facilities on smartphones for convenient storage. Moreover, new devices with the latest smartphone models were used to capture the videos. However, recording videos with smartphones is completely risk-free, allowing consumers to shoot videos without much restriction. The visual quality of the videos are sufficiently high as compared with the existing video datasets. Based on a qualitative evaluation that may change from one person to other persons, videos of some datasets were checked and in contrast to some existing video datasets that have fuzzy vision, the objects and scenarios in the recordings of our dataset are readily visible. Also, in terms of resolution as one quantitative evaluation, the majority of the video content in the dataset seems to have high resolution (1080 × 1920). This dataset is the only one with these many specifications and quality, thus making this forgery video dataset unique in every aspect. Moreover, the dataset is suitable for training a deep learning method, as did in the experiments.

The forthcoming section of the paper is assembled as follows. Section II denotes the review of different state-of-the-art datasets which used video data for forgery detection as well as the methods used in performing forgery detection. Our new video database is completely presented in Section III. Section IV describes database evaluation.

Section V provides the results of the experiments. Section VI discusses the results. Finally, the last section concludes this work.

II. RELATED WORKS

Forensic video and image analysis is the scientific inspection, comparison, and/or evaluation of video/images in legal approaches since this field is frequently employed in many high-profit cases, including international and conflict situations, that require highly trustworthy content in court [7]. Putting aside the significance of this sector and the requirement for a large number of datasets in this evaluation, the research field suffers from a shortage of either quality or quantity of datasets that may be used for this purpose. There are numerous image datasets [8], [9], [10], [11] till date that is being employed for image manipulations when compared with the video forgery datasets. Since this paper presents a dataset based on videos, the main focus will be on existing datasets related to video forgery detection.

Rossler et al. [20] suggested a face forensic video dataset used to detect the manipulation done on human faces. This dataset consists of manipulated data of more than 500000 images. These images were obtained from more than 1000 videos. The introduction of this dataset created a benchmark for the process of image forensics, segmentation, and classification and with generative refinement models, they provide a benchmark analysis for constructing manipulations utilizing existing ground truth.

Another facial video manipulation method known as MesoNet is proposed in 2018 by Afchar et al. [27]. This method is used to detect the forgery performed in videos using Deepfake and Face2Face methods by employing a deep learning technique. The dataset used for the experimentation of these techniques is obtained from publicly accessible videos over the internet [27] and a large scale dataset [20].

In 2012, a dataset named SULFA (Survey University Library for Forensic Analysis) [12] is introduced for benchmarking different forgery techniques. This dataset consists of 150 videos obtained from three camera devices, including both forgery and original videos. The length of the videos is 10 seconds with 30fps each and can be accessed online.

An expansion of [12] dataset is introduced in [13] which is composed of additional 10 videos i.e, 10 manipulated and 10 original videos apart from [12]. The videos are 30 fps with a resolution of 320 × 240 pixels with a duration of 10 seconds and can be found online. The disparities between the frames of the source material sequences and the fabricated sequences are contained in the dataset, which is beneficial in detecting video forgery.

In 2015, Ardizzone and Mazzola [14] developed a dataset using six non-identical videos from two other public datasets.[1],[2] This video dataset consists of 160 forged videos of different scenes from the regulation of traffic as well as parking scenarios with 30 fps. This dataset is built for detecting different video alternation methods and is employed by Digital Forensic Community.

Similarly, another dataset was introduced for object-based forgery in [15] consists of 100 video footage extracted from surveillance cameras having 25 frames per second each. The duration of each video in this dataset is about 11 seconds. This work is used to identify object-based manipulation performed on videos automatically.

A video forgery dataset was established for the purpose of forensic examination by Al-Sanjary et al. [16]. This dataset includes 33 videos collected from Youtube which are used to perform splicing, swapping as well as copy-move forgery. The length of each video of this dataset is 16 seconds and the frame rate is 30 frames per second.

A large standard dataset[3] was introduced in 2017 which involves 2520 manipulated images and 23 videos. In 2017, another video dataset known the GRIP dataset, which consists of 10 forged videos is created by splicing the videos by using Adobe After Effects CC software. The description of this dataset is found in [17]. This dataset is available online.[4] The resolution of the videos is 720 × 1280 pixels. In another work, a similar dataset also named GRIP which is developed by D'Amino et al. [19] using copy-move forgery in 15 videos by utilizing the software Adobe After Effects Pro. This video forgery dataset is publicly accessible.[5]

Reference [20] is used as a benchmark dataset comprising of more than 1000 videos which is used as a baseline for basic image forensic operations including identifying and segmenting fabricated images. In 2018, Korshunov and Marcel [21] presented a public video dataset of low and high quality video sequences comprised of 620 deepfake videos developed using GAN based method which is obtained from VidTIMIT dataset.[6]

A large-scale dataset known as the MFC dataset [22] was introduced in 2019, which constitutes 11000 HP videos, 4000 manipulated videos and 300000 video clips. This is a massive media forensic analysis baseline dataset for monitoring the effectiveness of quantitative multimedia forensic techniques. The results of the analysis can be utilized to assess existing tasks. Some of the other existing video forgery datasets available are [5], [23], [24], [25], and [26]. More detailed summary of the existing video forgery dataset is described in Table 1.

Even though developing high-quality realistic crafted videos with normal editing tools take a long time, only a few small datasets comprising classic manipulations such as copy-move and splicings are available on the internet. Therefore, this paper also presents a video dataset collected using 5 different smart phones and the videos are manipulated and experimented using traditional and deep learning approaches

[1] http://sulfa.cs.surrey.ac.uk/

[2] http://www.hitech-projects.com /euprojects/cantata/datasets_cantata /dataset.html

[3] https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation

[4] http://www.grip.unina.it/

[5] http://www.grip.unina.it/download/prog/ForgedVideosDataset/Copymove/

[6] https://conradsanderson.id.au/vidtimit/

TABLE 1. Summary of the existing video dataset used for different types of video forgery techniques.

Dataset	Title	Type of multimedia	Type of forgery	Size	Year
[12]	SULFA	Videos	copy-paste forgery	150 videos	2012
[13]	REWIND	Videos	Copy-move	10 forged and 10 original videos	2013
[14]	-	Videos	Copy-move	160 forged videos	2015
[15]	SYSU-OBJFORG	Videos	Object based forgery	100 original video sequences and 100 manipulated video sequences	2016
[16]	VTD	Videos	Splicing forgery, Copy-move forgery, Swapping Frames	33 forged videos	2016
[17]	GRIP	Videos	Splicing	10 forged videos	2017
[18]	Test dataset	Videos	Frame Duplication	31 forged videos	2018
[19]	GRIP	Videos	Copy-move	15 forged videos	2018
[20]	FaceForensics	Videos	Face forgery detection	more than 1000 forged videos	2018
[21]	DeepFakes	Videos	Face swap	620 deepfake videos	2018
[22]	MFC	Videos	Performance evaluation of media forensic system	11,000 HP videos, 4,000 manipulated videos, 300,000 video clips	2019
[23]	Celeb-DF	Videos	Deep face forensics	5, 639 videos	2020
[24]	TDTVD	Videos	Multiple tampering - frame deletion, frame duplication or frame insertion	210 videos	2020
[25]	DeeperForensics-1.0	Videos	Face forgery detection	60, 000 videos	2020
[26]	ForgeryNet	Videos	Image and video classification, spatial and temporal localization	221,247 videos	2021
[5]	VLFD	Videos	Detecting video interframe forgeries	210 native videos	2021

TABLE 2. Description of the forgery-video dataset listing different devices used for creating forgery videos.

Mobile devices	Total no.of videos manipulated	Resolution (width × height)	Type of manipulation	# Forged	# Original
IPhone 8 Plus	50	1080 × 1920	Insertion	25	25
			Deletion	25	25
Nokia 5.4	50	1080 × 1920	Insertion	25	25
			Deletion	25	25
Samsung A50	50	1080 × 1920	Insertion	25	25
			Deletion	25	25
Xiomi Redmi Note 9 Pro	50	1080 × 1920	Insertion	25	25
			Deletion	25	25
Huawei-Y9	50	720 × 1280	Insertion	25	25
			Deletion	25	25

thus producing better forgery detection results compared with the existing dataset. This dataset contains a large number of forgery videos when compared with some of the existing datasets such as [12], [13], [14], [16], and [19] which also performed copy-move forgery. This database is unique in that it includes a list of the source smartphones used to make the original recordings, which will aid in the deployment of PRNU-based video authentication procedures. There is no other dataset that contains such a vast number of authentic and faked videos, as well as the number of smartphones used to record them.

III. DATASET DESCRIPTION

The video dataset[7] consists of a collection of forgery videos and their respective original videos which are collected from different mobile devices. Five different mobile brands were used for performing the video forgery. The mobile brands are Iphone 8 plus, Huawei-Y9, Samsung A50, Nokia 5.4, Xiomi Redmi Note 9 pro. The original videos are from the Qatar University Forensic Video Database (QUFVD) [28], which consists of 6000 videos from 20 current cell phones of five different brands. Each of the devices consists of 300 videos. The whole QUFVD is divided into training, testing and validation data. Hence, this structure has employed classical PRNU and machine learning methods thus producing promising results for the source camera identification problems. Therefore, videos are selected from this dataset that are suitable for video forgery detection. The video contents in the dataset were captured using the main camera of the smartphones. Regarding the motion information within the videos, it is worth mentioning that the recording devices had slight movements when most of the videos were being captured. In some other videos, the objects in the scene can be seen moderately moving when the camera is at rest. The rest of the videos in the dataset are still videos where the movement of the camera as well as the objects in the videos are immobile. Therefore these videos are selected to perform forgery. The original videos were captured using the camera of different scenarios, primarily static and still objects of shopping malls, parking areas, gardens, streets, roads, sky,

[7] https://www.dropbox.com/sh/14djt7wggyljcjh/AAB08WbooHO0wKjGqbZkbaCsa?dl=0

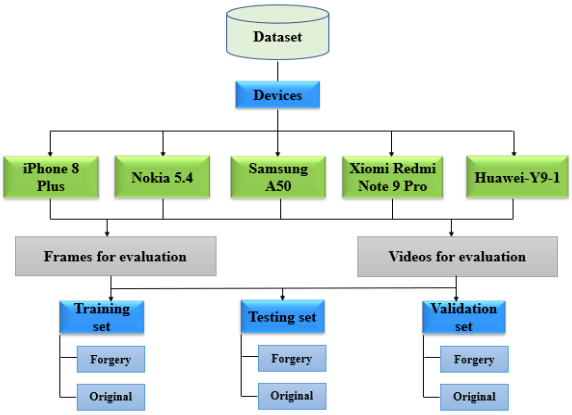

FIGURE 1. Structure of the folders and subfolders comprising the dataset.

sea, shops and homes. Out of the 300 videos, 50 videos were taken from five different mobile brands for manipulation. The video duration is 11-15 seconds and the formats are MPEG-4 and H.264. Out of the 50 videos from each device, 25 videos were used for manipulation by inserting objects and 25 videos were used for manipulation by removing objects. During insertion, the entire frame of the video is changed by placing the object in a desired position when compared with the original video. While removing the object from the video, the frames of the video where the void of the removed object when seen is replaced with the matching background. The matching background is fixed using the content aware fill of the Adobe Photoshop software. Therefore, every frame of the videos will be adjusted by creating reference frames. More details about video manipulation are discussed in the upcoming sections. The details of the devices and the number of manipulations for each device are shown in Table 2.

The process of insertion and removal of objects can be accomplished by using video editing software such as Adobe After Effects as well as Adobe Photoshop. The main reason for selecting Adobe After Effects software lies in its effectiveness to produce exclusive motion graphics. It also enhances the visual effects of the footage by reducing the time taken for rendering. This software is better at video editing as it consists of more inbuilt features and more options can be added with the help of different plugins. On the other hand, Adobe Photoshop is a powerful tool when used for object removal. This software works well with all types of file formats and the time consumed for editing an image is considerably less compared to other editing softwares. Moreover, this software is used to fill and match the background during object removal.

This dataset consists of original and manipulated data at the frame and video levels. The dataset's structure is first divided into two sections: FramesForEvaluation and VideosForEvaluation. The FramesForEvaluation consists of Training, Testing and Validation data of the frames whereas, VideosForEvaluation consists of videos which is also divided into Training, Testing and Validation data. The Training, Testing and Validation data of both frames and video consists of Forgery and Original data. The Forgery data in frames and videos are named based on ''Brand_Model_Ins\Del_(Video No.)''. The original data in the videos are named based on ''Brand_Model_Ori_Ins\Del_(Video No.)'' The format of the data of the videos is MOV for Iphone 8 and rest of devices is MP4. The Overview of the structure of the dataset is shown in Figure 1.

FIGURE 2. Sample videos of the forgery video dataset using different mobile brands.

A. VIDEO MANIPULATION

After collecting the videos from each device, the videos were selected for executing two types of manipulation. Inserting objects is the first kind of manipulation and removing objects is the second kind of manipulation performed on the videos. The software used for these types of manipulation is Adobe After Effects 2020 as well as Adobe Photoshop 2020. The resolution of the videos of Nokia 5.4, Samsung A50, Iphone 8 and Xiomi Redmi Note 9 Pro in terms of width and height is 1080 × 1920 whereas, the resolution of the videos taken using Huawei-Y9 is 720 × 1280. The forged and original videos have the same file format, resolution and frame rate before and after manipulation. Samples videos of our dataset is shown in Figure 2.

1) INSERTING OBJECTS

This manipulation requires an additional transparent object to be inserted into each of the frames of the videos. The first step is selecting the video from the dataset to insert the object. Then, select the track camera option from the Animation option tab to track the motion of the moving video. The third step is to select a tracking point as a reference area to stick the object to insert in this place. In order to edit the track solid that is created, choose the pre-compose option to create a nested composition. Then, the composition should be selected and the object should be transparent to match the scenes of the video. This process requires adjusting other parameters as well as depending on the recorded video, changing the background, and adding more animation until

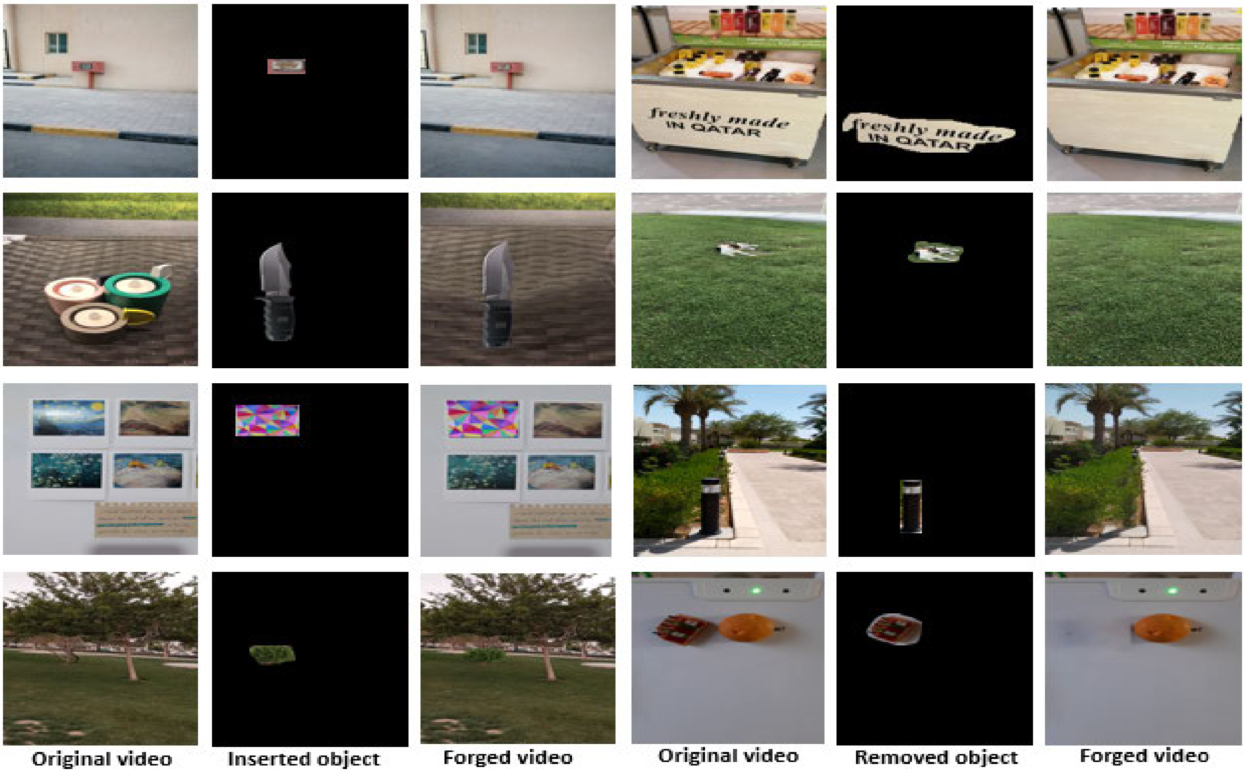

FIGURE 3. Sample of the original videos, inserted object, removed object and the final forged video of the dataset.

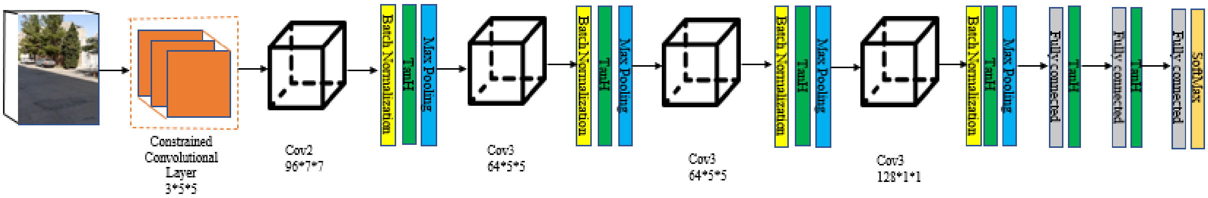

FIGURE 4. Architecture of the ConstrainedNet (Based on [29]).

reaching the required format. After placing the objects on each frame, the video should be rendered using Adobe Media Encoder to get the desired form. Each of the videos is in MPEG-4 format.

2) REMOVING OBJECTS

This type of manipulation is carried out on moving videos. The objects in the videos are selected by using different shape tools. The object is masked using the properties of the mask. After masking, the reference layer is created using Photoshop software. Using this software, background scenario of the object will be matched in accordance with the video. Using the content aware fill on Photoshop, the background of the object is filled. The video will then be rendered, resulting in the forged video. The removal of objects from the videos is flawless, and it is appropriate for the background scenarios of the videos. When compared to other videos from the existing datasets, the rendered videos will be of fine standards.

Samples of the original videos, inserted and removed objects are shown in Figure 3.

IV. DATASET EVALUATION

The methods presented in [19] as a successful classical method and MISLnet [29] as a deep learning method are used to evaluate the quality of the dataset. The classical method begins by computing appropriate features that are identical to multiple spatial, temporal, and intensity transformations. The features are derived compactly on a spatiotemporal grid instead of being at conspicuous keypoints, enabling for both additive and occlusive forgeries to be detected. Following that, a Nearest-Neighbor Field (NNF) is constructed, which links each characteristic to its perfect match. An ad hoc video-oriented variant [30], [31] of PatchMatch [32], [33] is utilized for this, which takes advantage of the NNF's intrinsic coherence to minimize search complexity. Ultimately, the NNF is postprocessed to identify locations with consistent spatiotemporal deformation as potential copy-move

TABLE 3. The statistics for training, testing, and validation at both the frame and video levels.

Data	Frame				Video			
	Original		Forged		Original		Forged	
	Insertion	Deletion	Insertion	Deletion	Insertion	Deletion	Insertion	Deletion
Training	29986	28181	29986	28181	79	77	79	77
Testing	10507	9109	10507	9109	28	25	28	25
Validation	6718	8399	8399	6718	18	23	18	23

TABLE 4. The results of the video level in terms of accuracy (%) for both implementations (2D and 3D).

Type of forgery	[19] (2D)			[19] (3D)		
	Forged	Original	Overall	Forged	Original	Overall
Insertion	25.00	83.50	54.25	40.00	83.50	61.75
Deletion	37.00	68.20	52.60	43.00	68.20	55.60
All	31.00	75.85	53.42	39.47	75.85	57.66

candidates. This technique is based on 2D and 3D properties that are invariant. Both of the implementations are tested on the proposed dataset.

In the deep learning method, which is also used in [34] a Constrained Convolutional layer is added at the beginning of a CNN to perform forensic tasks, as shown in Figure 4. The network used a constrained convolutional layer that was added as the first layer which uses three kernels with size 5 as shown in Figure 4. This layer is constructed in such a way that there are relationships between adjacent pixels that are independent of the content of the scene. As a result of the layer, low-level features are extracted that suppress the image content. To design the layer, the convolutional layer filters are enforced by the following constraints:

$$\begin{cases} \omega_{k_j}^{(1)}(0,0) = -1 \\ \sum_{m,n\neq 0} \omega_{k_j}^{(1)}(m,n) = 1 \end{cases} \tag{1}$$

where $j = \{1, 2, 3\}$. Moreover, $\omega_{k_j}^{(1)}$ denotes the jth kernel of the kth filter in the first layer of the network. The experiments showed that the layer can improve results compared with deep learning architectures without the layer.

Our implementation of the architecture is based more on [34] considered for source camera identification. The stochastic gradient descent (SGD) is used to train the model. The batch size is set to 100 and the momentum and decay parameters of the stochastic gradient descent are set to 0.95 and 0.0005, respectively. The CNN is trained for 10 epochs in each experiment.

This result provides a baseline for the accuracy of forgery detection in the dataset and can be used for comparison with other methods. The division of the dataset for the experiments is that 80% of these videos are considered training set and the remaining 20% are considered test set. Also, 20% of the training data is considered as validation data. The structure can be used for both classical and machine learning methods. In the training step, the locations of the forged videos are not considered and only the labels for the two classes are based on the type of videos (forgery and original), resulting in real conditions during the training step. The statistics for training, testing, and validation at both the video and frame levels are shown in Table 3. For example, in classical methods, for a fair comparison, testing set can be considered for evaluation. All devices have video in both the training and test sets. It also ensures that for a video, both the original video and forgery videos are included in one of the training and test sets.

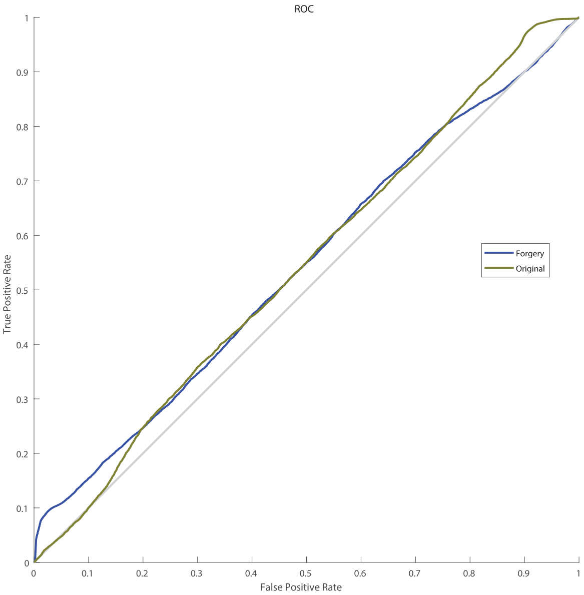

FIGURE 5. True and false positive rates (ROC) obtained based on deep learning method at frame level.

V. EXPERIMENTAL RESULTS

As mentioned in dataset evaluation section, the methods [19] and [29] are used for the experimental analysis. A 2-class problem should be considered for forgery video detection. Also, it can be considered as a one-class problem in which only original videos should be trained. To detect a video by its frames, all frames in the test set are considered.

In [19], a threshold is used to decide whether a video belongs to the classes of fake or original. The performance of the method is measured by computing the accuracy based on video-level in both original and forgery (insertion and deletion) videos. The video-level results for the method are shown in Table 5.

The scores obtained by the CNN based on the highest probability show which patch belongs to which class. At the

TABLE 5. The results of patch-level, frame-level, and video-level [29] in terms of accuracy (%).

Type of forgery	Patch			Frame			Video		
	Forged	Original	Overall	Forged	Original	Overall	Forged	Original	Overall
Insertion	41.50	80.25	60.75	30.60	74.88	52.74	45.50	62.00	53.75
Deletion	39.50	82.50	61.00	31.00	80.52	55.76	48.00	71.50	59.75
All	40.50	81.37	60.87	30.80	77.70	54.25	46.75	66.75	56.75

TABLE 6. Confusion matrix of deep learning method [29] at frame level.

		Target calss		
		Forgery	Original	
Output class	Forgery	6035	4374	58.0 %
	Original	13581	15242	52.9 %
		30.8 %	77.7 %	54.2 %

TABLE 7. Error rates in terms of FRR (%), FAR (%) AND AER (%) for two implementation of [19] (2D and 3D) and deep learning method [29].

Methods	FAR	FRR	AER
[19] (2D)	72.50	25.00	48.75
[19] (3D)	61.00	24.55	42.77
[29] (Deep Learning)	74.00	23.20	48.60

frame level, the frame is considered fake if any of the patches belongs to the fake class. Also at the video level, the video is considered fake if any of the frames fall into the forgery class. The performance of the deep learning method is measured by computing the accuracy based on patch-level, frame-level and video-level in both original and forgery (insertion and deletion) videos. The results for the method are shown in Table 5. Table 6 shows the confusion matrix obtained for the Deep Learning method at the frame level. The confusion matrix can show misclassifications between all classes. As shown in the table, misclassifications between two classes occur the same. Figure 5 provides a more comprehensive picture of forgery detection performance to check the quality of the CNN in frame level by presenting the Receiver Operating Characteristic (ROC) curve. Two values are calculated for each threshold: True Positive Ratio (TPR) and False Positive Ratio (FPR). The TPR of a given class, e.g. forgery, is the number of outputs whose actual and predicted class is forgery divided by the number of outputs whose predicted class is forgery. The FPR is calculated by dividing the number of outputs whose actual class is not a forgery, but whose predicted class was a forgery by the number of outputs whose predicted class is not a forgery.

Also three important measures, i.e, FAR (False Accept Rate), FRR (False Reject Rate) and AER (Average Error Rate) are reported in Table 7 for both methods. The FRR and FAR is obtained by the ratio of the number of the rejected original test videos to the total number of submitted original test videos and the ratio of the number of accepted forgeries, to the total number of forgeries, respectively.

A 64-bit operating system (Ubuntu 18) with a CPU E5-2650 v4 @ 2.20 GHz, 128.0 GB RAM, and four NVIDIA GTX TITAN X. was used in order to run the experiments.

VI. DISCUSSION

Although, recently, Deep Learning methods have been introduced to solve this problem, traditional methods have obtained promising results in the field. As mentioned earlier, the proposed dataset is also evaluated using both deep learning method and traditional method developed to solve this problem. Overall, the results at the video level show that the traditional method is more successful than deep learning method for the forgery detection problem, but they does not work well on the dataset.

As shown in Table 4, [19] based on the 3D approach achieves better results compared to 2D. When the original videos are tested with the method compared to forgery videos, the method achieves better results. In the forgery type, deleted objects are recognised better than inserted objects. Overall, however, the detection of object insertion videos achieves about 60 % better results than that of object deletion videos for both original and fake videos.

As shown in Table 5, the Deep Learning method is also more successful in detecting original videos compared to the forgery videos. At the video level, it achieves 66.75 % accuracy in detecting the original videos. For the forgery videos, the method is more accurate in detecting object deletion videos compared to object insertion videos.

The result of the two tables is that the method presented in [19] and deep learning methods are better at detecting original videos. This result can also be confirmed by Table 7. In the table, the less values for FAR, FRR and AER shows that which of the methods obtained better results. For example, the best result is when the measure is Zero (0). The table shows the deep learning method has obtained better result for FRR. It means that the method is successful for forgery videos compared to [19], which is better for FAR that shows the method detects original videos better than deep learning method.

Aside from the experimental results, this dataset contains a greater number of high-resolution forgery and original videos. The object insertion manipulation is accomplished impressively by inserting a random transparent object that corresponds to the video sequences. Each video consists of the fixing of various objects that fit the scenes of the video and are fixed on the entire frame of the video. When compared to the previously released dataset, this is a unique feature of this dataset.

VII. CONCLUSION

This paper presents a new video dataset based on smartphones for forgery detection. The dataset includes five popular smartphone brands, 250 original and 250 forged videos (125 insertion videos and 125 deletion videos). The entire dataset is provided with an evaluation analysis for use by the research community. The dataset is suitable for use by deep learning methods and traditional methods. Object insertion and deletion are the fabrications performed on the forgery video dataset. The experimental results also show that the proposed dataset offers the research community a sufficient amount of challenges which could well be faced in real scenarios to deal with forgery detection on digital videos. Another distinguishing feature of this dataset is that the videos were taken using smartphones which are manipulated by employing copy-move forgery. The manipulation is carried out on every frames of the videos making the dataset more distinctive as compared with the traditional dataset.

Different deep learning methods could be fused at the score level in the future to enhance the current performance. Also, different augmentation methods could be used to tackle the problem of limited training sample size in deep learning. It would also be worth investigating the traces of manipulation in relation to the editing tools used (i.e., Adobe After Effect and Photoshop) and this remains an open research question. Finally, since the database is accessible in video format in its current form, other frame formats such as TIFF can be considered as input data for easy access and processing.

ACKNOWLEDGMENT

The statement made herein are solely the responsibility of the authors.

REFERENCES

[1] H. Tian, Y. Xiao, G. Cao, Y. Zhang, Z. Xu, and Y. Zhao, "Daxing smartphone identification dataset," *IEEE Access*, vol. 7, pp. 101046–101053, 2019.

[2] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics," *APSIPA Trans. Signal Inf. Process.*, vol. 1, no. 1, 2012, Art. no. e2.

[3] N. A. Shelke and S. S. Kasana, "A comprehensive survey on passive techniques for digital video forgery detection," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 6247–6310, Feb. 2021.

[4] G. Ulutas, B. Ustubioglu, M. Ulutas, and V. Nabiyev, "Frame duplication/mirroring detection method with binary features," *IET Image Process.*, vol. 11, no. 5, pp. 333–342, Jan. 2017.

[5] H. Sharma and N. Kanwal, "Video interframe forgery detection: Classification, technique & new dataset," *J. Comput. Secur.*, vol. 29, no. 5, pp. 531–550, 2021.

[6] Y. Akbari, S. Al-Maadeed, O. Elharrouss, F. Khelifi, A. Lawgaly, and A. Bouridane, "Digital forensic analysis for source video identification: A survey," *Forensic Sci. Int., Digit. Invest.*, vol. 41, Jun. 2022, Art. no. 301390.

[7] T. Gloe, A. Fischer, and M. Kirchner, "Forensic analysis of video file formats," *Digit. Invest.*, vol. 11, pp. S68–S76, May 2014.

[8] T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. de Rezende Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 7, pp. 1182–1194, Jul. 2013.

[9] J. Dong, W. Wang, and T. Tan, "CASIA image tampering detection evaluation database," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 422–426.

[10] T.-T. Ng, J. Hsu, and S.-F. Chang. *Columbia Image Splicing Detection Evaluation Dataset*. Accessed: Sep. 19, 2019. [Online]. Available: http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm

[11] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD—New database for copy-move forgery detection," in *Proc. ELMAR*, Sep. 2013, pp. 49–54.

[12] G. Qadir, S. Yahaya, and A. T. S. Ho, "Surrey University Library for Forensic Analysis (SULFA) of video content," in *Proc. IET Conf. Image Process. (IPR)*, London, U.K., 2012, pp. 1–6, doi: 10.1049/cp.2012.0422.

[13] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Local tampering detection in video sequences," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2013, pp. 488–493.

[14] E. Ardizzone and G. Mazzola, "A tool to support the creation of datasets of tampered videos," in *Image Analysis and Processing—ICIAP 2015* (Lecture Notes in Computer Science), vol. 9280, V. Murino and E. Puppo, Eds. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-23234-8_61.

[15] S. Chen, S. Tan, B. Li, and J. Huang, "Automatic detection of object-based forgery in advanced video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 11, pp. 2138–2151, Nov. 2016.

[16] O. Ismael Al-Sanjary, A. A. Ahmed, and G. Sulong, "Development of a video tampering dataset for forensic investigation," *Forensic Sci. Int.*, vol. 266, pp. 565–572, Sep. 2016.

[17] D. D'Avino, D. Cozzolino, G. Poggi, and L. Verdoliva, "Autoencoder with recurrent neural networks for video forgery detection," *Electron. Imag.*, vol. 29, no. 7, pp. 92–99, Jan. 2017.

[18] G. Ulutas, B. Ustubioglu, M. Ulutas, and V. V. Nabiyev, "Frame duplication detection based on BoW model," *Multimedia Syst.*, vol. 24, no. 5, pp. 549–567, Oct. 2018.

[19] L. D'Amiano, D. Cozzolino, G. Poggi, and L. Verdoliva, "A PatchMatch-based dense-field algorithm for video copy–move detection and localization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 669–682, Mar. 2019.

[20] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics: A large-scale video dataset for forgery detection in human faces," 2018, *arXiv:1803.09179*.

[21] P. Korshunov and S. Marcel, "DeepFakes: A new threat to face recognition? Assessment and detection," 2018, *arXiv:1812.08685*.

[22] H. Guan, M. Kozak, E. Robertson, Y. Lee, A. N. Yates, A. Delgado, D. Zhou, T. Kheyrkhah, J. Smith, and J. Fiscus, "MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2019, pp. 63–72.

[23] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for DeepFake forensics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3207–3216.

[24] H. D. Panchal and H. B. Shah, "Video tampering dataset development in temporal domain for video forgery authentication," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 24553–24577, Sep. 2020.

[25] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2889–2898.

[26] Y. He, B. Gan, S. Chen, Y. Zhou, G. Yin, L. Song, L. Sheng, J. Shao, and Z. Liu, "ForgeryNet: A versatile benchmark for comprehensive forgery analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4360–4369.

[27] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2018, pp. 1–7.

[28] Y. Akbari, S. Al-Maadeed, N. Al-Maadeed, A. A. Najeeb, A. Al-Ali, F. Khelifi, and A. Lawgaly, "A new forensic video database for source smartphone identification: Description and analysis," *IEEE Access*, vol. 10, pp. 20080–20091, 2022.

[29] B. Bayar and M. C. Stamm, "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2691–2706, Nov. 2018.

[30] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient dense-field copy–move forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2284–2297, Nov. 2015.

[31] L. D'Amiano, D. Cozzolino, G. Poggi, and L. Verdoliva, "Video forgery detection and localization based on 3D patchmatch," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun. 2015, pp. 1–6.

[32] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 24:1–24:11, Jul. 2009.

[33] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized PatchMatch correspondence algorithm," in *Computer Vision—ECCV 2010* (Lecture Notes in Computer Science), vol. 6313, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010, doi: 10.1007/978-3-642-15558-1_3.

[34] D. Timmerman, S. Bennabhaktula, E. Alegre, and G. Azzopardi, "Video camera identification from sensor pattern noise with a constrained ConvNet," 2020, *arXiv:2012.06277*.