electronics

Review

Interframe Forgery Video Detection: Datasets, Methods, Challenges, and Search Directions

Mona M. Ali [1], Neveen I. Ghali [1], Hanaa M. Hamza [2], Khalid M. Hosny [2], Eleni Vrochidou [3,*] and George A. Papakostas [3]

[1] Department of Digital Media Technology, Faculty of Computers and Information, Future University in Egypt (FUE), New Cairo 11835, Egypt; mona.almakhton@fue.edu.eg (M.M.A.); neveen.ghali@fue.edu.eg (N.I.G.)

[2] Department of Information Technology, Faculty of Computers and Information, Zagazig University, Zagazig 44519, Egypt; hmkamal@fci.zu.edu.eg (H.M.H.); k_hosny@zu.edu.eg (K.M.H.)

[3] MLV Research Group, Department of Informatics, Democritus University of Thrace, 65404 Kavala, Greece; gpapak@cs.duth.gr

* Correspondence: evrochid@cs.duth.gr; Tel.: +30-2510-462320

Abstract

The authenticity of digital video content has become a critical issue in multimedia security due to the significant rise in video editing and manipulation in recent years. The detection of interframe forgeries is essential for identifying manipulations, including frame duplication, deletion, and insertion. These are popular techniques for altering video footage without leaving visible visual evidence. This study provides a detailed review of various methods for detecting video forgery, with a primary focus on interframe forgery techniques. The article evaluates approaches by assessing key performance measures. According to a statistical overview, machine learning has traditionally been used more frequently, but deep learning techniques are gaining popularity due to their outstanding performance in handling complex tasks and robust post-processing capabilities. The study highlights the significance of interframe forgery detection for forensic analysis, surveillance, and content moderation, as demonstrated through both evaluation and case studies. It aims to summarize existing studies and identify limitations to guide future research towards more robust, scalable, and generalizable methods, such as the development of benchmark datasets that reflect real-world video manipulation diversity. This emphasizes the necessity of creating large public datasets of manipulated high-resolution videos to support reliable integrity evaluations in dealing with widespread media manipulation.

Keywords: deep learning; machine learning; interframe forgery detection; frame insertion forgery; frame deletion forgery; frame duplication forgery; double compression video; multimedia security

check for updates

Academic Editor: Yue Wu

Received: 17 May 2025
Revised: 16 June 2025
Accepted: 30 June 2025
Published: 2 July 2025

Citation: Ali, M.M.; Ghali, N.I.; Hamza, H.M.; Hosny, K.M.; Vrochidou, E.; Papakostas, G.A. Interframe Forgery Video Detection: Datasets, Methods, Challenges, and Search Directions. *Electronics* **2025**, *14*, 2680. https://doi.org/10.3390/electronics14132680

1. Introduction

Image and video data are the most important and commonly used forms of communication in today's society. The video consists of a series of still images displayed rapidly, creating the illusion of motion, and stored in various formats, such as AVI and MOV. Frame rate is the quantity of frames or images shown in a video each second, affecting the smoothness of motion. The video serves as reliable evidence and verified proof in many fields, such as criminal justice, forensics, journalism, and others. Video forensics endeavors to identify and document forensic indicators of diverse modifications implemented in a specified video's temporal and spatial domains [1]. The emergence of video applications and data has also given rise to the issue of video and image forgery. Although significant efforts have been dedicated to addressing image forgery, video forensics remains a formidable challenge, especially with interframe forgery videos, where videos can be manipulated in numerous ways, with frame insertion, deletion, and duplication being among the principal obstacles [2].

Digital technology has made it easier for individuals to create forged videos, raising significant concerns regarding information security and the authenticity of visual media [3]. Video evidence is one of the most important forms of evidence that can be utilized to identify incidents. With modern technologies, editing videos has become simple due to the ease of use and widespread accessibility of video editing software, such as Adobe Premiere and DaVinci Resolve. Video editing is manipulating and rearranging footage to create a new video, often involving cutting, transitions, and effects. These advancements in video technology have improved the lives of modern people; however, they are creating social issues due to intentionally malicious content modifications [4]. Malicious intent can lead to changes. Therefore, the fidelity of the audiovisual materials that will serve as evidence or to furnish individuals with accurate information is necessary.

Video forgery refers to the modification of digital video footage to create false or altered representations of reality. This survey provides a comprehensive and systematic examination of interframe video forensics. Numerous researchers have contributed survey articles in this domain. Nevertheless, their focus has primarily been on conducting surveys within all aspects of image or video forensics, rather than emphasizing interframe detection methodologies. Our work is motivated by the urgent need for effective techniques in diverse situations and how they can be enhanced to address evolving security challenges. We summarize various methods for detecting interframe video forgeries. We have looked for original research papers written in English in Springer, IEEE, and ScienceDirect databases. In this analysis, only papers published in journals and recorded in Scopus are considered. Figure 1 shows the publisher distribution for interframe video forgery. Papers were included based on the inclusion criteria. Terms searched within the paper title included Interframe Forgery Detection, Frame Duplication Forgery, Frame Insertion Forgery, Machine Learning, and Deep Learning.

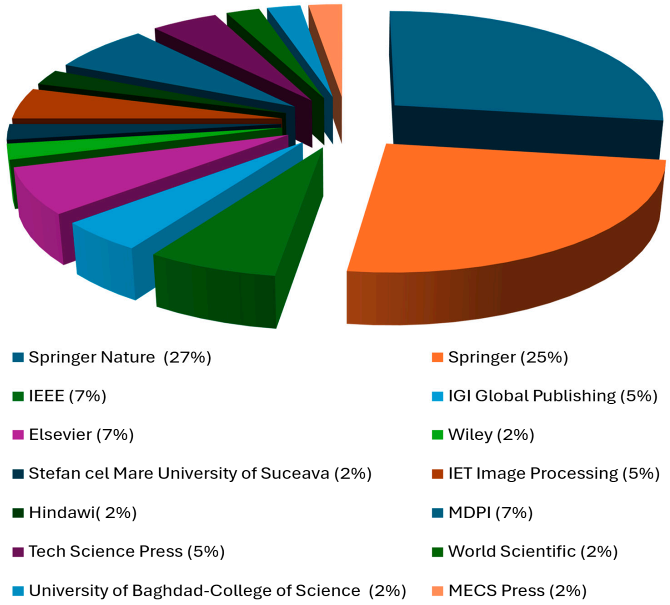

Figure 1. The distribution of publishers for interframe forgery video detection research work.

This survey aims to provide a comprehensive analysis of interframe forgery detection, examining literature and techniques, their key features, and understanding their scope and limitations for overcoming them. It also describes many efficient dataset types that researchers have used to test interframe forgery detection techniques. Current interframe forgery detection studies often rely on limited or outdated datasets, lack standardized performance metrics, and focus on narrow types of manipulation, resulting in inconsistent evaluations and poor generalization to real-world scenarios. In addition to providing a comprehensive analysis of dataset trends and limitations, our contributions include the identification of advanced detection architectures that exhibit improved performance and the provision of insights into model generalization and resilience to post-processing. This survey addresses critical gaps by systematically evaluating detection methods using established performance metrics, including accuracy, precision, recall, and F1-score. These findings have crucial implications for forensic investigations, surveillance systems, and digital content moderation platforms by offering more reliable tools for video integrity verification.

The remainder of the paper is organized as follows: Section 2 presents various types of video forgery. Section 3 presents video forensics datasets. Section 4 reports classifying interframe forgery video techniques according to traditional machine learning and deep learning methodologies. Section 5 analyses and summarizes the performance of the different approaches discussed in Section 4. Section 6 discusses the published studies between 2019 and 2024, their limitations and challenges, and the directions of their search. Finally, Section 7 concludes the study.

2. Video Forgery Types

The types of forgery detection in digital videos are divided into intra-frame forgery, also known as spatial forgery; interframe forgery, known as temporal forgery; and spatiotemporal forgery [5], as shown in Figure 2. Figure 2 outlines the key elements of each type of video forgery, including the specific operations made to the video and the hierarchical levels at which these alterations can be analyzed and detected [6].

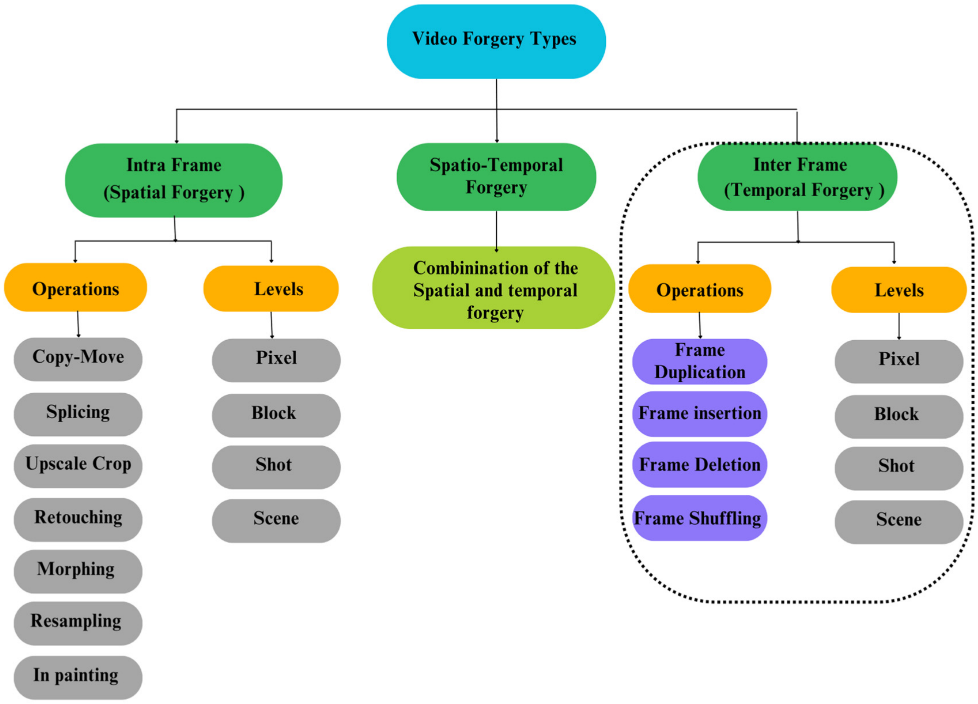

Figure 2. Video forgery classification.

Interframe forgery is a common technique in video tampering. It involves changing the video frames to manipulate the content [3]. The operations carried out in interframe forgery include frame duplication, frame insertion, frame deletion, and frame shuffling [7,8]. Frame duplication occurs when one or more frames are copied and reinserted into the video, often to extend the duration of an event or to slow down specific actions. This form of forgery may not always be noticeable to human viewers but introduces repetitive patterns that can be detected algorithmically. Researchers have developed methods based on spatial and temporal consistency, measuring pixel-level inconsistencies between adjacent frames to identify duplicated content [3,9].

Frame insertion involves adding new frames to the sequence, potentially from another video, which can introduce content not initially presented [10]. This manipulation can alter the event timeline and create a misleading narrative. Techniques such as optical flow inconsistency have been employed to detect frame insertion and analyze irregular movement patterns that occur when new content is added [8,11,12].

Frame deletion is another common technique for interframe forgery. It involves removing specific frames from the sequence, often hiding certain events or information. This manipulation disrupts the continuity of motion, making it appear that specific actions did not occur. The detection of frame deletion usually relies on motion analysis or velocity field techniques, where sudden jumps in the movement flow of objects or individuals indicate the presence of deleted frames [13,14].

Frame shuffling is an additional form of interframe forgery that involves rearranging frames in a different order than their original sequence [12]. This technique distorts the logical progression of events, causing actions to appear out of their proper order [15,16]. Detection methods for frame shuffling often rely on bitstream analysis, where inconsistencies in the compressed video data highlight irregularities in the frame order, as shown in Figure 3.

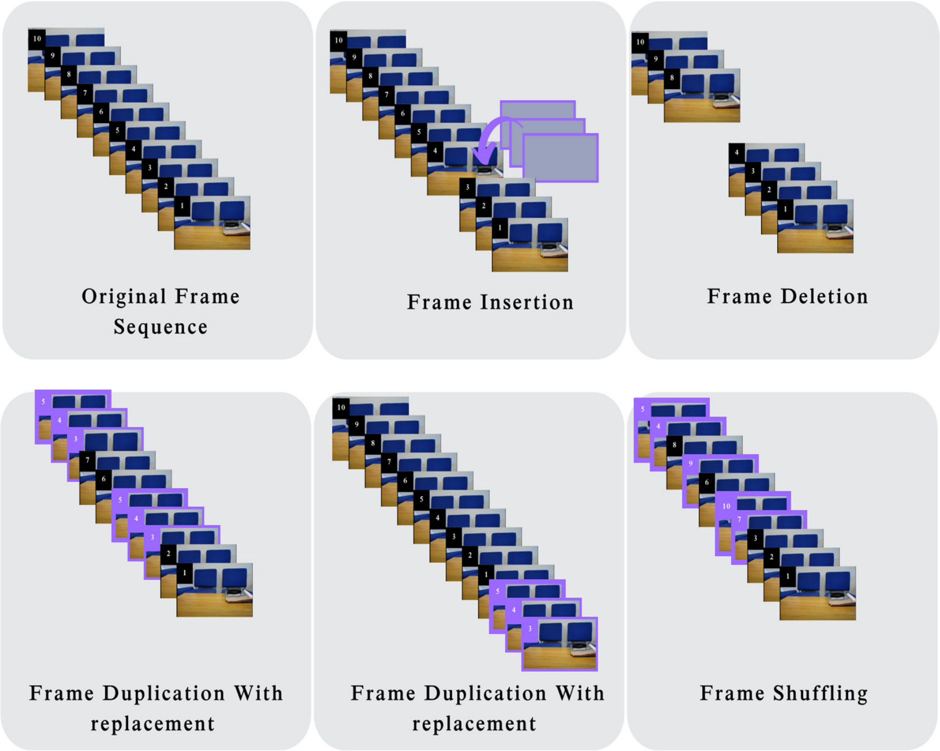
Original Frame Sequence
Frame Insertion
Frame Deletion
Frame Duplication With replacement
Frame Duplication With replacement
Frame Shuffling

Figure 3. Standard processes for various interframe video forgeries.

Interframe forgery can be analyzed at different levels of granularity, each offering a unique perspective on detecting manipulation. At the pixel level, individual pixel values within frames are compared for inconsistencies. This level of analysis helps detect subtle changes introduced by operations such as frame insertion or duplication [3]. Residual frame analysis and pixel-wise comparison techniques are often employed at this level, where pixel differences across frames can reveal duplicated or inserted frames.

A frame is divided into smaller segments at the block level, and the consistency between blocks across different frames is analyzed [17]. Block-level analysis reduces computational complexity by focusing on localized regions of the frame, which is particularly effective for detecting manipulations such as frame duplication or shuffling. Methods like DCT are commonly used to identify inconsistencies in block motion and correlation [7].

Shot-level analysis compares entire sequences of continuous frames, looking for inconsistencies across longer segments. This level is beneficial for identifying temporal manipulations, such as frame insertion or deletion, where tampering spans multiple frames or shots. Shot-level detection often involves keyframe comparison, where significant frames from different shots are extracted and analyzed for discrepancies. At the highest level of analysis is the scene level, where entire scenes are compared for inconsistencies. Scene-level analysis looks at the overall flow of events. It can detect larger-scale tampering, such as frame deletion or shuffling, that disrupts the narrative or timeline of a video. Scene-level detection often involves global feature analysis, such as identifying abrupt changes or irregular transitions between shots [5].

Intra-frame forgery, also known as spatial forgery, involves tampering with individual frames in a video without altering the temporal sequence. This type of forgery targets the visual content within a single frame, making it appear that particular objects, people, or elements were always present or absent. Several operations are commonly used in intra-frame forgery, including copy-move, splicing, upscaling, cropping, retouching, morphing, and inpainting [18].

Copy-move is one of the most prevalent forms of intra-frame forgery. In this technique, part of an image or frame is copied and pasted elsewhere within the same frame. This operation often hides objects or fills gaps left after deleting elements. Since the copied region is from the same frame, it maintains consistent texture and color, making the forgery challenging for the eye to detect. However, advanced detection techniques, like key-point-based methods or block-matching algorithms, can detect minor inconsistencies in the copied areas [19].

Splicing involves merging parts from different images or frames into a single image or frame. Unlike copy-move, which uses content from the same frame, splicing introduces external content, creating significant discrepancies in lighting, shading, and noise patterns [20]. Techniques such as edge inconsistency analysis or deep learning models can detect splicing by identifying these variations [21].

The existing methods of splicing forgery are categorized into the following two groups: splicing from a different source and copy-move from a similar source. In the splicing technique, objects are taken from one video and inserted into another to create a new video. Detecting splicing involves identifying inconsistencies between the inserted objects and the background of the target video. Deep learning techniques, such as RNN [22] and CNN [21], are employed for effective splicing detection. Video copy-move forgery creates realistic manipulation by seamlessly blending with original content and can be classified as interframe, intra-frame, or a combination of both. It also includes additive forgery, which introduces new objects, and occlusive forgery, which conceals content using copied background elements [17].

Upscaling and cropping are other methods used in intra-frame forgery. This technique alters the resolution and size of objects within a frame by artificially enlarging or reducing portions of the image. It is often used in conjunction with other forms of tampering, such as copy-move or splicing, to resize or draw attention to manipulated areas. Although upscaling can blur or distort the copied area, techniques such as resampling detection can reveal inconsistencies introduced by the scaling process [23].

Retouching involves manipulating pixel values to enhance or obscure features in a frame. This operation is commonly used to adjust colors, sharpen objects, or blur specific regions to highlight or hide certain elements. While subtle, retouching can be detected through techniques that analyze pixel noise or color distribution across a frame, which may show unnatural transitions or smudging in retouched areas [19].

Morphing is another intra-frame manipulation technique that alters the shape and appearance of objects in the frame. Deep learning-based image editing tools often use this technique to subtly modify facial expressions, body shapes, or other objects. Detecting morphing is particularly challenging, as the alterations can be minimal. Still, geometric consistency checks, or shape-matching algorithms, can help identify unnatural changes in an object's structure [24].

Resampling involves re-encoding or resampling an image, which can introduce new artifacts or distortions. It is often used when frames are resized or rotated. Resampling detection usually relies on frequency domain analysis, where inconsistencies in pixel patterns or noise frequency can reveal tampering.

Inpainting is a more advanced form of forgery that involves filling in missing areas within a frame using surrounding information. Inpainting is often used to remove unwanted objects or restore damaged areas of a frame. While this technique can produce highly realistic results, detection is possible through texture consistency analysis, where subtle irregularities in the fill patterns or differences in texture grain can signal forgery. In some cases, deep learning models trained to identify the inconsistencies between inpainted and original regions have been significantly used [25].

Intra-frame forgery detection can be conducted at different levels, like interframe forgery. Individual pixels are analyzed for color, texture, and noise inconsistencies at the pixel level. Pixel-level detection is beneficial for identifying operations like retouching, copy-move, or inpainting, as these techniques often leave behind minor artifacts. The frame is segmented into smaller blocks at the block level, and comparisons are performed between these areas to find texture, color, or shading irregularities. This approach is practical for detecting copy-move operations, where specific blocks within a frame may have been duplicated.

While more commonly associated with interframe forgery, shot-level analysis can also apply to intra-frame manipulations in scene transitions or larger-scale spatial modifications. For instance, the large-scale splicing or upscaling of objects can disrupt the visual coherence of an entire shot, making detection possible through global motion or pattern analysis. Scene-level detection is more applicable when analyzing the overall coherence of the visual narrative. In cases where intra-frame forgery is used to alter multiple frames within a scene, scene-level analysis can identify inconsistencies in the overall visual flow, especially when objects or people are manipulated across different shots or angles [26].

Spatiotemporal forgery is an advanced video manipulation technique that combines spatial and temporal alterations to deceive viewers or distort video content. Spatial forgery, also known as intra-frame forgery, involves modifying individual frames by adding, removing, or altering objects, people, or backgrounds through splicing or copy-move operations. Temporal forgery, also known as interframe forgery, modifies the arrangement or timing of frames by duplicating, deleting, or rearranging them. Spatiotemporal forgery integrates both types of tampering, making detection more challenging, as it involves changes within frames and their temporal relationships, ensuring the manipulated frames are synchronized with the rest of the video to distort the portrayed events or sequences [24].

Sensor artifacts are vital for detecting interframe manipulations, as sensor pattern noise (SPN) uniquely links video frames to their originating sensor and reveals any alterations. Forged areas can still be identified even when such elements as SPN are present during the forgery detection techniques that compare noise residues across frames with a known reference SPN [27].

Moreover, when analyzed, sensor artifacts such as CFA artifacts in image blocks reveal embedded regions with their detection robust to different distortions, alongside which they can identify forgeries of at least a minimum size of 2 × 2 pixels [28]. Additionally, the continuity of noise between frames is essential in identifying forgeries, even when SPN is induced in the forged frames. This approach enhances the reliability of detecting interframe forgeries, especially in scenarios where attackers attempt to manipulate SPN to evade detection. Furthermore, the use of SPN in forgery detection methods demonstrates robustness against various distortions and compression techniques, ensuring the effectiveness of the detection process [20].

Double compression plays a crucial role in detecting interframe forgery, especially in the context of video forensics. Various studies highlight the significance of detecting double compression in videos to identify tampering and ensure the integrity of digital media [29–32]. Detecting double compression is challenging but essential, as tampered videos are often recompressed after manipulation, leading to artifacts that can be analyzed to identify forgery [29].

Various techniques have been presented to detect double compressions, which include analyzing periodicity in DCT coefficients and prediction errors of frames [33]. Additionally, techniques like BAS and VPF have been utilized for double compression detection in MPEG-4 videos [34]. Understanding the impact of double compression on residual data with different group-of-picture (GOP) structures and lengths is crucial for accurate interframe forgery detection [35]. Overall, the research underscores the importance of double compression detection in interframe forgery analysis to maintain the credibility and authenticity of digital videos [36].

3. Video Forensics Datasets

Multiple benchmark datasets for detecting forgery are currently available and have been used to assess the algorithm's performance. While researchers may choose to utilize the datasets outlined in Table 1 for their analysis, there is a shortage of datasets that researchers commonly employ. Typically, datasets curated explicitly for this objective predominantly emphasize copy-move and frame duplication, with only a minor fraction documented on dynamic backgrounds instead of static ones. As a result, scholars frequently resort to compiling their datasets utilizing videos from YouTube or other readily available sources. The accessibility of video manipulation datasets is vital for thoroughly analyzing and validating various tampering detection methods. This section provides a detailed examination of the datasets in Table 1 and carefully reviews and assesses the currently available video forgery datasets.

The HEVC-based Tampered Video Dataset (HTVD) [37] consists of many authentic and tampered video settings. The dataset comprises videos from indoor, outdoor, and surveillance contexts, as illustrated in the sample videos in Figure 4. It shall consist of 966 tampered videos, 60 original videos, masks, and related ground truth data. Each original video was captured using a smartphone with HEVC support. The dataset encompasses various object-based intra-frame forgeries, such as cloning, splicing, and inpainting, along with diverse interframe forgeries, including frame insertion, deletion, and duplication, as indicated in Figures 5–7, respectively. There are 8694 tampered videos, reflecting the variations in tampered videos based on the codec's encoding parameters, precisely the GOP size and frame types.

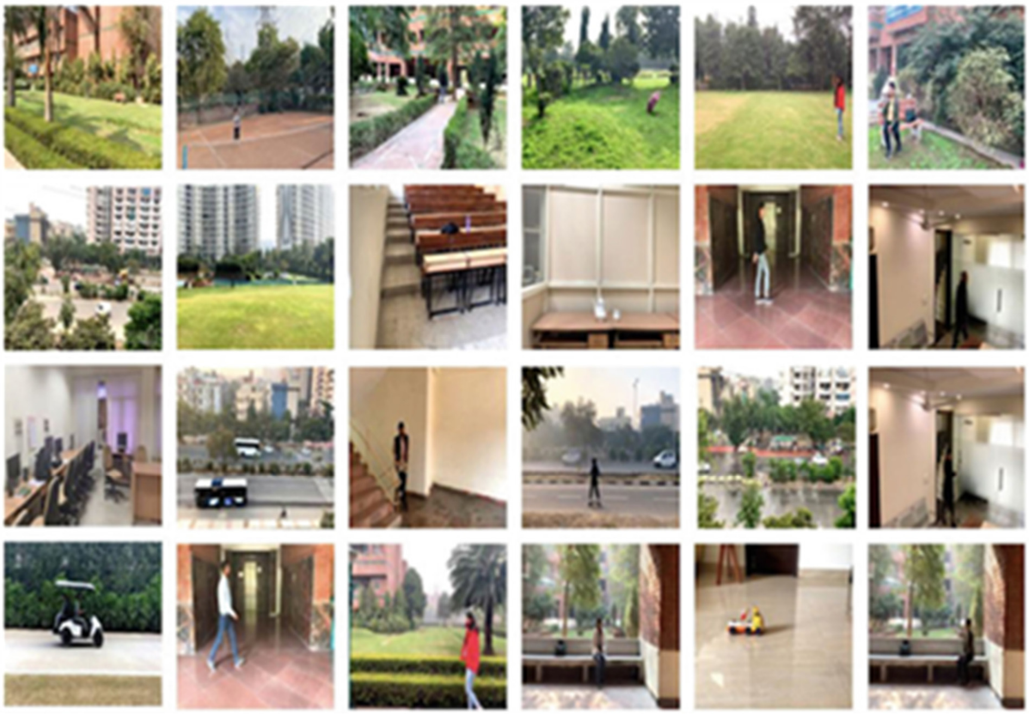

Figure 4. Visual examples extracted from the HTVD dataset [37], illustrating different categories of video tampering used for evaluating forgery detection algorithms.

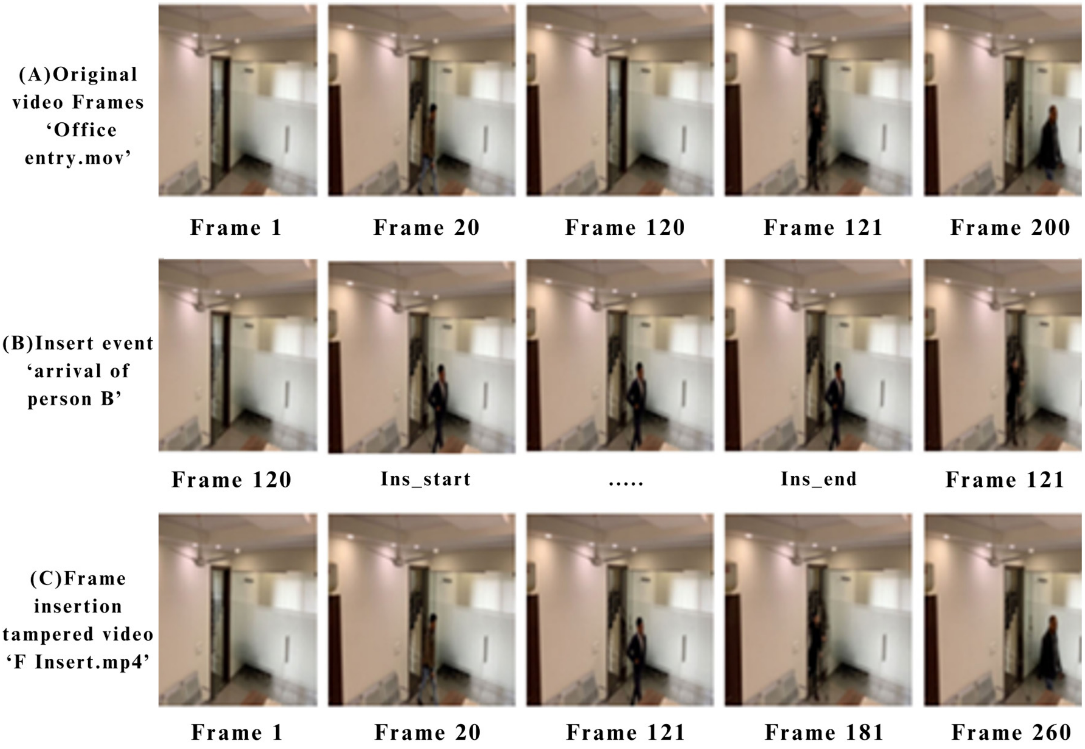
(A)Original video Frames 'Office entry.mov'
Frame 1
Frame 20
Frame 120
Frame 121
Frame 200
(B)Insert event 'arrival of person B'
Frame 120
Ins_start
.....
Ins_end
Frame 121
(C)Frame insertion tampered video 'F Insert.mp4'
Frame 1
Frame 20
Frame 121
Frame 181
Frame 260

Figure 5. Visualization of frame insertion tampering in a sample video from the HTVD dataset [37], illustrating the insertion of foreign frames into a video sequence to manipulate temporal events.

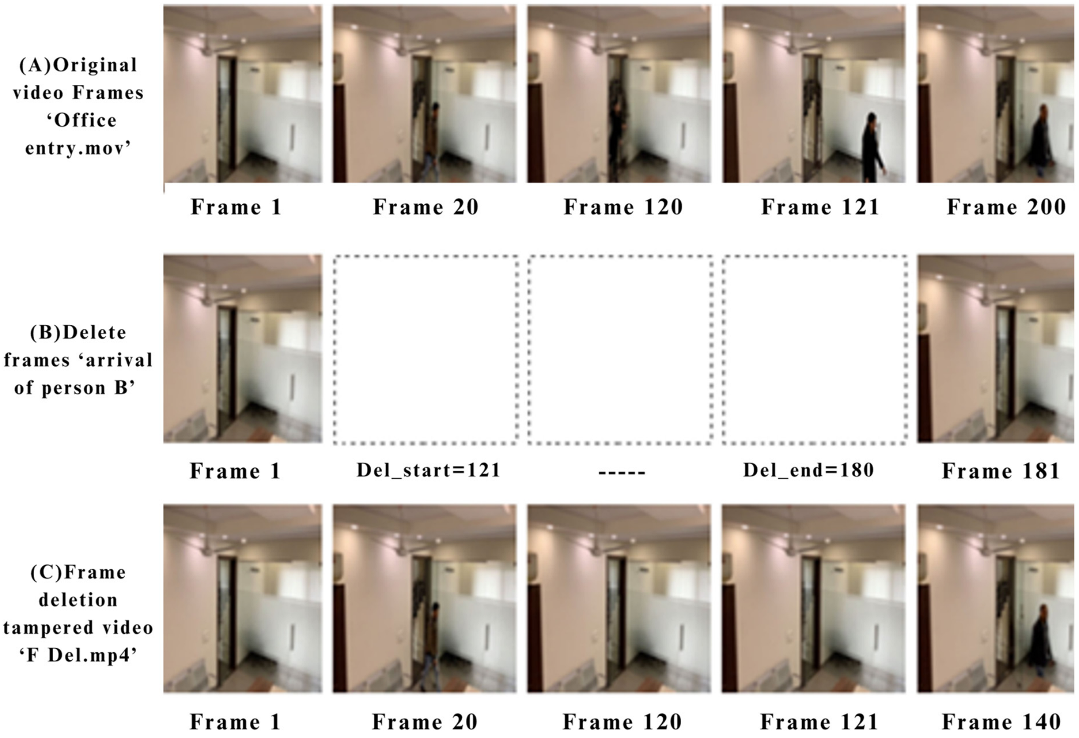

Figure 6. Visualization of frame deletion tampering in a video sample from the HTVD dataset [37], highlighting the forgery technique of eliminating specific frames to alter the perceived video content.

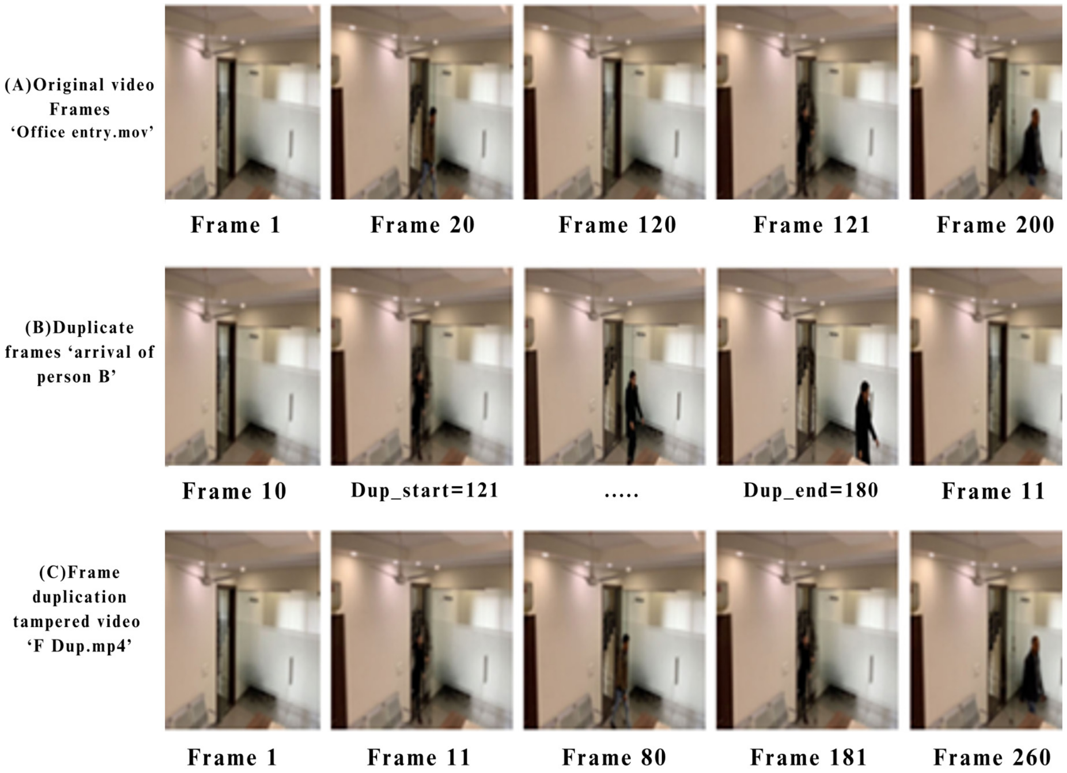
(A)Original video Frames 'Office entry.mov'
Frame 1
Frame 20
Frame 120
Frame 121
Frame 200
(B)Duplicate frames 'arrival of person B'
Frame 10
Dup_start=121
.....
Dup_end=180
Frame 11
(C)Frame duplication tampered video 'F Dup.mp4'
Frame 1
Frame 11
Frame 80
Frame 181
Frame 260

Figure 7. Visualization of frame duplication tampering in a video sample from the HTVD dataset [37], where illustrating duplicated frames within the same video stream.

The TDTVD dataset [38] contains several tampering videos and all temporal domain tampering. The resulting tampered video ranges in length from 6 to 18 s. Videos covering a variety of activities in different locations are included in the static and dynamic databases.

Researchers can access the full dataset publicly, which will be extremely helpful for testing their algorithms on such a large dataset. To help verify the tampering detection algorithms, each created tampered video additionally includes extensive ground truth information, such as the type of tampering, the modified frames, and the location of the tampering.

The VTL dataset uses the Video Trace Library (ASU) or YUV Video Sequences [39]. The manipulated videos have a resolution of 352 × 288 pixels and a frame rate of 30 frames per second (fps).

The SULFA dataset [40] widely utilized data collection containing 150 original videos. These videos were captured by Canon SX220, Nikon S3000, and Fujifilm S2800HD cameras, operating at a frame rate of 30 frames per second. Each video lasts approximately 10 s. Specifically, this dataset has been subjected to copy-move forgery in the spatial–temporal domain.

The REWIND dataset [41] consists of a combination of original and forged videos. Copy-move forgery is applied to the original video sequences in the spatial and temporal domains. Most of these original videos are from the SULFA library [40].

The GRIP video dataset [22,42] is subdivided into splicing and copy-move categories. Ten forged videos are accompanied by their original counterparts in the splicing dataset. On the other hand, the copy-move dataset comprises 60 original videos and 95 forged videos, all accompanied by comprehensive ground truth information.

To assess the method for detecting FD and FDS [43], original videos from the SULFA dataset [40], the LASIESTA dataset [44], and the IVY LAB dataset [45] were selected to create the dataset. The dataset comprises 51 videos (10 FD, 8 FDS, 5 with multiple forgeries, and 9 combining FD and FDS), all of which have been post-processed with Gaussian blur, brightness changes, and noise. Video durations range from 10–27 s.

The video tampering dataset (VTD) [46] includes 33 videos broken down into the following three categories: copy-move, splicing, and shuffling frames. This dataset has more tampered videos with longer durations than previous datasets. Each video lasts 16 s and has a resolution of 1280 × 720 pixels with a frame rate of 30 frames per second. This database is open to the public for use in studies on copy-move manipulation, splicing, and frame shuffling.

The TVD dataset [47] consists of 160 altered videos generated by duplicating objects from genuine traffic and surveillance scene environments at three distinct resolutions. The original videos are sourced from the SULFA [40] and CANTATA [48] datasets. The manipulated videos are generated through various geometric transformations, encompassing flipping, scaling, rotations, and shearing.

Table 1. Video forensics datasets.

Ref.	Dataset	Types of forgery	No. of Videos	Format	Resolution	Direct Download Link
[37]	HTVD	Frame insertion (Realistic)	108	MP4	1980 × 1080	http://rb.gy/2onorf (accessed on 14 September 2024)
		Frame insertion (Smart)	2700			
		Frame Deletion (Realistic)	108			
		Frame Deletion (Smart)	2700			
		Frame Duplication (Realistic)	108			
		Frame Duplication (Smart)	2700			
		Object cloning	90			
		Splicing	90			
		Inpaint	90			

Table 1. *Cont.*

Ref.	Dataset	Types of forgery	No. of Videos	Format	Resolution	Direct Download Link
[38]	TDTVD	Frame deletion Frame duplication Frame insertion	210	AVI	320 × 240	https://rebrand.ly/TVTVD (accessed on 14 September 2024)
[39]	VTL	Frame deletion	24	4:2:0 YUV	352 × 288	https://tinyurl.com/33v39d8u (Accessed: 14 September 2024)
[40]	SULFA	Copy-move	150	MOV, AVI	320 × 240	Not directly download
[41]	REWIND	Copy-move	20	MP4, AVI	320 × 240	https://github.com/ShobhitBansal/Video_Forgery_Detection_Using_Machine_Learning?tab=readme-ov-file (accessed on 1 January 2024)
[42]	GRIP	Copy-move	154	MP4	1280 × 720	http://www.grip.unina.it/ (accessed on 1 January 2024) (Not directly download)
		Splicing	10	AVI		
[43]	FD&FDs	Frame duplication Frame shuffling	53	AVI	320 × 240	Not directly download
[46]	VTD	Copy-move Splicing Frame shuffling	33	AVI	1280 × 720	https://tinyurl.com/4mfjy5f9 (accessed on 14 September 2024)
[49]	FVD	Frame Deletion Frame Insertion Frame Duplication Frame Duplication with Shuffling	32	AVI	320 × 240, 352 × 288, 704 × 576	https://rb.gy/roitjj (accessed on 14 September 2024)
[47]	TVD	Copy-move Geometric transformations	160	AVI	360 × 640, 576 × 768, 540 × 960	https://rb.gy/6br9m3 (accessed on 14 September 2024 (Not directly download)
[50]	VLFD	Frame duplication	210	AVI, MP4	3840 × 2160, 1920 × 1080	https://csepup.ac.in/vlfd-dataset/ (accessed on 14 September 2024) (Not directly download)
[51]	VIFFD	Frame insertion Frame deletion Frame duplication Frame shuffling	390	AVI	1920 × 1080, 720 × 404	https://rb.gy/pqtkpc (accessed on 14 September 2024)
[52]	Sondos	Frame insertion Frame deletion Frame duplication Frame shuffling	15	AVI	320 × 240 352 × 288 704 × 576	Not directly download
[53]	TRECVID	-	40	MPEG-1	720 × 576	https://trecvid.nist.gov/ (accessed on 30 September 2022)

Table 1. *Cont.*

Ref.	Dataset	Types of forgery	No. of Videos	Format	Resolution	Direct Download Link
[54]	SYSU-OBJFORG	Object-based forged	100	MP4, MPEG-4	1280 × 720	https://tinyurl.com/yzv6veum (accessed on 14 September 2024) (Not directly download)
[55]	UTF101	-	13K	AVI	320 × 240	https://tinyurl.com/4hk32upd (accessed on 14 September 2024)
[56]	MFC	-	4k	MP4, MOV	-	Not directly download
[57]	VIRAT	-	329	MP4	640 × 480	https://viratdata.org/ (accessed on 14 September 2024)

The VLFD dataset [50] contains original and forged videos with frame duplication forgery. It includes videos taken with smartphones of varying resolutions and camera technologies. The dataset repository comprises 50 stationary and 60 dynamic mode videos captured utilizing a smartphone camera with a 3840 × 2160 pixels resolution and a 30 frames per second frame rate. Additionally, the dataset comprises videos acquired from different smartphone cameras, including 50 stationary and 50 dynamic mode videos with a resolution of 1920 × 1080 pixels.

The forged video dataset (FVD) comprises 32 films that have undergone different manipulations, including duplication, insertion, deletion, and frame duplication with shuffling. The original videos used in this dataset were obtained from the SULFA [40], IVY LAB, and LASIESTA [44] datasets. The VIFFD dataset [51] has been gathered from five security cameras that capture real-world scene views under various lighting circumstances.

The Sondos dataset [52] comprises 15 forged videos, including almost all interframe forgeries. These forgeries occur at various resolutions. It is essential to mention that these videos are in the AVI format.

The TRECVID surveillance event detection (SED) evaluation [53] requested the participation of volunteers to create intricately manipulated videos. Five videos were recorded for each of these scenes—an elevator and an airport passageway—over five different days. Each video clip has a duration of 2 min.

SYSU-OBJFORG [54] represents the largest object-based video forgery dataset, comprising 100 authentic and 100 manipulated videos. Authentic clips are from static surveillance cameras, while forgeries involve adding, removing, or repositioning moving objects.

The UCF101 dataset [55] comprises a collection of realistic action videos from YouTube, encompassing 101 action categories. This dataset is an extension of the UCF50 dataset, which includes a smaller range of 50 action categories. Within the UCF101 dataset, there are a total of 13,320 videos spanning the 101 action categories. UCF101 stands out for its wide variety of sports-related actions, presenting significant challenges due to diverse factors in camera setup and illumination conditions. As a result, UCF101 has become the most challenging dataset to date for action recognition.

The Media Forensic Challenge (MFC) [56] introduced a benchmark dataset, serving as a platform for assessing the authenticity of video content. This comprehensive dataset encompasses a vast collection of 300,000 videos, including 11,000 high-provenance (HP) videos and 4000 manipulated videos. It provides invaluable historical data and annotations regarding the manipulation process. The MFC dataset further distinguishes itself by offering a diverse range of frame rates among its data samples.

The VIRAT Video Dataset [57] is a real-world video dataset meticulously crafted to reflect the real world. To accurately represent the challenges faced in video surveillance, this dataset incorporates a range of resolutions, varying background clutter, scene diversity, and a broad spectrum of human activity and event categories. As a result of these qualities, the VIRAT Video Dataset has emerged as a widely recognized benchmark within the computer vision community due to its ability to provide a comprehensive collection of diverse and complex visual events.

4. Classification of Interframe Forgery Video Techniques According to Methodology

The detection of interframe forgery in videos involves various techniques, each with its own methodological approach. These techniques typically follow a standardized pipeline to analyze video data, extract essential features, and identify any manipulation. Understanding this classification is crucial for evaluating the pros and cons of each approach, as well as their suitability for different scenarios. The diagram in Figure 8 presents a general framework for detecting interframe forgery. It begins with pre-processing the input video, which involves steps such as resizing, denoising, and adjusting the video quality to enhance subsequent analysis. Next is the feature extraction stage, where key characteristics such as frame differences, motion vectors, or texture details are collected. These features are then refined through feature pre-processing, which may involve normalization or dimensionality reduction to optimize the data for analysis.

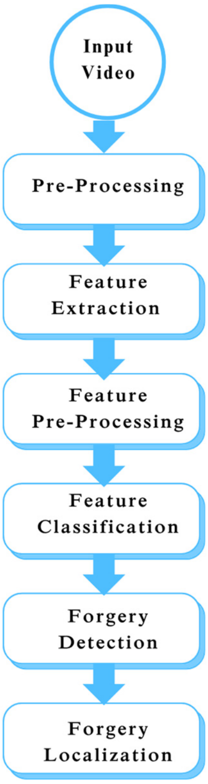

Figure 8. The general framework of video interframe forgery detection methods.

Feature classification involves using machine learning algorithms or rule-based techniques to categorize features as potential forgery or non-forgery. After that, forgery detection occurs, where the system confirms the presence of forgery based on the classification results. The process concludes with forgery localization, which identifies the exact frames or regions in the video that have been manipulated. This framework, depicted in Figure 8, offers a comprehensive and structured approach that can be applied across various methodologies used in interframe forgery detection.

4.1. Machine Learning Methods

Details on video forgery detection, implemented using machine learning methods and reviewed in this subsection, are summarized in Table 2. Aghamaleki and Behrad [58] proposed a methodology for identifying video forgery through spatial and temporal analysis of quantization effects on the I-frame. Conversely, the residual errors pertain to P-frames. Initially, they employed a Support Vector Machine (SVM) classifier, leveraging the first significant digit law of DCT coefficients within I-frames to detect single or double-compressed videos. Subsequently, they applied temporal analysis of the quantization effects on the residual errors of P-frames to uncover nefarious interframe forgery.

Singh and Aggarwal [59] propose two forensic methods to detect recompression and frame deletion in digital videos, utilizing key components of DCT coefficient analysis and optical flow analysis. They test their methods on digital images and videos from various sources, such as Columbia, Dresden, Sulfa, and change detection databases. The authors also propose a new method utilizing the high beam half of the optical flow to detect and localize, providing high detection accuracy rates for both recompression and frame deletion detection. For recompression detection, 100% accuracy is achieved for MPEG-2, MPEG-4, and H.264/AVC-encoded videos. For frame deletion detection, an accuracy of 99.3% is achieved in MPEG-x and H.264/AVC videos. They demonstrate the robustness and flexibility of their methods in terms of compression values, quality factors, bit rates, and anti-forensic operations.

Another optical flow-based technique is proposed in [60]. It is a new framework for forgery detection in MPEG videos, utilizing spatial–time domain analysis of the quantization effect on the DCT coefficients of the I-frame and the residual errors of the P frame. The algorithm leverages the fact that video communication protocols introduce inconsistencies in the quantization levels of video frames. The algorithm can detect and localize interframe networks, such as frame deletion, insertion, and duplication. The algorithm is tested on a large dataset of publicly available network videos. The experimental results show that the algorithm requires 345 s of simulation time, and the detection accuracies for tampered videos of the scene elevator and airport passageway are 93.3% and 86.7%, respectively.

Wu et al. [61] introduce a novel framework that relies on velocity field coherence to identify video interframe forgery, especially frames involving successive deletions or duplications. The generalized extreme studentized deviation (ESD) test can be used to detect and analyze these outlier classes. It is easy to pinpoint specific locations where these changes occurred in the videos. The result demonstrates the effectiveness of the algorithm, achieving overall accuracies of 90.0%, 85.0%, and 80.0% for identifying original videos, frame deletion videos, and frame duplication videos, respectively.

Bao et al. [62] present a novel method for detecting anti-forensic video forgeries by analyzing noise transfer matrices, with a focus on detecting frame deletions in static scenes. The approach employs a three-layer pyramid structure for noise extraction and adaptive weight adjustments in the transfer matrix, thereby enhancing detection accuracy. Tested on a dataset of 80 videos (half of which were tampered with), the method demonstrates significant improvements in detection performance, particularly in scenarios with low false positive rates, outperforming both traditional and learning-based methods.

Saddique et al. [30] identify manipulated video segments and label the altered frames based on their texture and micro-patterns. The method employs DRLBP and the chrominance value of the difference between consecutive frames (CCD) to create a novel descriptor. This descriptor captures the inconsistency between frames at the interframe level, which arises from the forgery. Utilizing the SVM classifier, the technique determines whether consecutive frames are genuine or forged.

Prashant and Krishnrao [63] present an innovative forensic framework designed to identify frame-shuffling forgery within MPEG-encoded videos. The suggested methodology combines motion vector analysis with scale-invariant characteristics to identify and pinpoint instances of frame shuffling forgery, achieving a remarkable 100% detection accuracy across 15 evaluated videos.

Singla et al. [29] proposed two interframe forgery detection methods for HEVC-coded videos. The first stage detects anomalous points based on compression field features, and the second stage verifies these points by localizing forgery. The experimental results demonstrate that the proposed method is effective.

Bakas et al. [64] proposed a forensic methodology for identifying frame insertion, deletion, and duplication instances in video forgeries. The initial step involves identifying outlier frames using a Haralick-coded frame correlation. The subsequent step involves a more precise detection level, aiming to minimize false positives and enhance the efficiency of forgery detection.

Forgery duplication is a standard video manipulation technique, where traditional video tampering detection methods are often inefficient and ineffective for detecting spoofed videos in challenging environments. Girish and Nandini [12] present a novel model for detecting video forgery. Video sequences were initially obtained from the University of Surrey Forensic Analysis Library (SULFA) [40] and Sondos's interframe dataset [52]. Additionally, a spatial time-averaging technique has been developed for the acquired video sequences to extract the historical background information of moving subjects, facilitating effective video forgery detection. This is followed by feature extraction using the Google LeNet framework to generate feature vectors. Subsequently, the Unsupervised Feature Selection (UFS-MSRC) method, combined with multi-scene randomization collaboration, is applied to identify significant feature vectors that substantially decrease the training duration, while enhancing detection precision. Ultimately, networks utilizing video sequences are incorporated, with an extended long short-term memory (LSTM) network deployed for the detection task. The empirical analysis revealed that the UFS-MSRC integrated with the LSTM model achieved accuracy rates of 98.13% and 97.38% on the SULFA and Sondos datasets, respectively.

A method for detecting frame duplication attacks in passive forgery scenarios is proposed by Bozkurt and Ulutaş [7]. The approach involves visualizing feature vectors and generating a binary image from a feature matrix, which solves the problem by extracting a forged frame group template from the binary image. The proposed technique is effective for both compressed and uncompressed video formats. Its resilience to compression is assessed using videos encoded with MPEG-4 and H.264 standards, and the algorithm also shows robustness against blurring attacks. Additionally, it can accurately identify the location of the forgery and trace the source of the tampered frames. Experimental results highlight the algorithm's speed, accuracy, and resistance to compression.

Kang et al. [65] improve the existing frame deletion detection algorithm by combining the time and magnitude of the P-frame sequence error. It also introduces the anti-forensic approach by comparing the actual latent prediction errors. The paper demonstrates that the proposed methods are more effective than those for detecting frame deletion, anti-forensic modification, and video interframe forgery attacks. The proposed methods are also robust to motion estimation algorithms, and detecting video frame deletion achieved an accuracy of 96.9% and an Area Under the ROC Curve (AUC) of 99.3%. In conclusion, only a few methods have been found in the literature that address the detection of anti-forensically prepared forged videos.

Mohiuddin et al. [9] introduce an approach that identifies multiple instances of frame duplication in manipulated videos. Within this specific type of forgery, several frames are replicated and inserted into a video to conceal or emphasize specific incidents or objects. Detecting the duplicated frames in these digital videos poses a significant challenge, as they are difficult to discern visually and computationally. While previous studies have proposed various methods to address this issue, only a few techniques can effectively identify frame duplication in videos with static and non-static backgrounds. This study presents a straightforward yet efficient method for detecting frame duplication forgery in videos, addressing this gap. This method utilizes SSIM to evaluate the similarity between consecutive frames. Additionally, a search algorithm is employed to detect and localize tampered frames. For experimentation, original videos were collected from Urban Tracker, derf's collection, and REWIND [41]. Duplicating frames created databases and forged videos. The experimental results from these videos demonstrate that the proposed method achieves an impressive average accuracy of 98.90% in detecting frame duplication forgery.

Panchal and Shah [66] introduce a multi-level approach for detecting video frame insertion forgery. At the initial level, temporal characteristics between consecutive frames are analyzed using temporal perceptual information measurement, also known as Temporal Information. The z-score is used to identify anomalies and pinpoint locations of forgery. The second verification level involves calculating the structural similarity index measurement (SSIM) for the inserted frames detected in the first step. These SSIM values are then compared to those of the original video frames to confirm if the inserted frames originate from a different video source. The method is tested on the TDTVD dataset, with experimental results demonstrating an improved detection accuracy of 94.44% for frame insertion forgeries. A comparison is also made between the proposed technique and existing video forgery detection methods.

G. Singh and K. Singh [67] introduce a passive blind method to detect video frames and regions forgeries using correlation coefficient and variation algorithms. It involves extracting mean features, creating sequence matrices, and localizing errors to calculate region similarities. The scheme is tested on 30 videos from the SULFA dataset [40] and others from the internet.

Pandey and Kushwaha [68] present a histogram-based technique for detecting video forgeries by examining changes in pixel intensity across video frames, rather than relying on machine learning methods. It details processes such as extracting frames, calculating histograms, visualizing data, and performing analysis to identify manipulative alterations. The proposed approach demonstrates an accuracy rate of 87.5%. Although the method effectively tackles video forgery detection.

Ren et al. [69] present a methodology for detecting and localizing frame duplication forgeries based on similarity assessment using an enhanced Levenshtein distance metric. Initially, the manipulated video sequence is segmented into overlapping subsequences. Then, the similarities of each subsequence with all other subsequences are computed. The enhanced Levenshtein distance is employed to evaluate these similarities within this study. The analysis of subsequence similarities is conducted to identify potentially duplicated frame pairs. These identified duplicate frame pairs are then consolidated into a comprehensive duplicate sequence, facilitating the localization of frame duplication forgeries. Comprehensive experimental evaluations have been performed on various tampered videos sourced from the VTL dataset [39]. The findings indicate that the proposed technique's precision can reach 99.5%, surpassing the precision of contemporary leading methods. Additionally, the proposed technique can pinpoint the precise location of the duplicate and is capable of detecting frame duplication forgeries in static scenes.

Temporary copy-move or frame duplication is a standard method of video forgery where consecutive frames are copied within the same video. Mohiuddin et al. [70] present an ensemble approach to detect duplicate frames in video. First, the images are preprocessed, and three different features are extracted from each image. The frames are then ordered based on the values of their properties. A filtering method is employed to eliminate false detections and prevent the generation of duplicate sequences. A voting mechanism is used to predict duplicate images. The proposed method achieves high detection accuracy, true positive rate, and true negative rate on a dataset of 300 videos. The method is robust to various post-processing techniques and outperforms other state-of-the-art methods.

Table 2. Video forgery detection using machine learning methods.

Ref.	**Year**	**FrameDeletion**	**FrameInsertion**	**FrameShuffling**	**FrameDuplication**	**DoubleCompression**	**Short Description**
[63]	2024			✓			It identifies potential forgery by analyzing motion vectors and detecting abrupt variations in the mean values of frames, which indicate possible frame shuffling. Suspicious frames are validated using SIFT features and RANSAC homography to confirm forgery.
[62]	2024	✓					It involves three main steps—noise extraction, computation of the noise transfer matrix, and adjustment of the transfer matrix weights. Noise extraction uses a three-layer pyramid structure that captures local and structural noise features. The noise transfer matrix evaluates the distance between noise features in consecutive frames.
[3]	2024	✓	✓	✓	✓		Frame difference analysis by examining the pixel intensities of adjacent frames in a video. The original sequence of frames exhibits a smooth pattern of differences, while any manipulation disrupts this pattern, resulting in spikes. Sequencing alteration detection, which analyzes the differences between adjacent frames and identifies alterations and potential forgery. Location identification by pinpointing the presence of spikes.
[71]	2024	✓	✓		✓		A passive approach to detect and localize interframe forgeries using texture features like HoG, uniform, and rotation invariant LBP.
[66]	2023		✓				A two-level method for detecting video frame insertion forgeries. It combines temporal perceptual analysis using z-score anomaly detection with SSIM-based verification.

Table 2. *Cont.*

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[70]	2023				✓		It detects tampering artifacts by post-processing techniques, such as adjusting contrast, boosting brightness, blurring, and introducing noise. A voting mechanism is used to predict duplicate images. A filtering method removes false detections and generates duplicate sequences.
[29]	2023	✓				✓	It detects anomalous points based on compression field features. It verifies the irregular points associated with the localization of forgery.
[7]	2023				✓		The method extracts feature vectors from video frames using DCT and converts them into binary images. Applies edge detection to the binary images and identifies the candidate regions for forgery using histogram analysis and correlation coefficients. Searches for the corresponding source frames of the duplicated frames using SSIM.
[11]	2022	✓	✓	✓			Uses Li et al.'s robust hashing method, which is based on selective quaternion invariance, to compute hash values from the extended frames and compare them with a threshold to detect tampered frames.
[9]	2021				✓		The SSIM-based method measures the similarity between consecutive frames. The search algorithm detects and localizes the tampered frames.
[69]	2021				✓		The tampered video sequence is segmented into small overlapping subsequences, and the similarity between them is calculated using the enhanced Levenshtein distance. Subsequence merging and localization.
[49]	2020	✓	✓	✓	✓		The method utilizes HOG as a distinctive attribute extracted from each image to detect interframe forgery. Correlation coefficients are employed to identify instances of frame deletion and insertion, with abnormal points detected using Grubbs' test. MEI is applied to edge images of each shot to identify instances of frame duplication and shuffling, contributing to the overall capability of identifying all interframe forgeries.
[72]	2020	✓	✓		✓		Extracting the ENF signal from the suspicious video. The ENF signal is obtained using band-pass filtering and cubic spline interpolation techniques. Analyzing the correlation coefficient between adjacent periods in the interpolated ENF signal.

Table 2. *Cont.*

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[73]	2020				✓		A passive blind forgery detection technique for identifying frame duplication attacks in MPEG-4 videos. It employs a two-step algorithm that involves SIFT key points for feature extraction and a RANSAC algorithm for identifying duplicate frames.
[74]	2020	✓	✓				The Triangular Polarity Feature Classification (TPFC) framework for detecting video frame insertion and deletion forgeries. It utilizes a novel TPD and extracts discriminative features from the MLS framework.
[43]	2020			✓	✓		Using temporal average and GLCM features.
[30]	2019	✓	✓		✓	✓	Identify manipulated video segments and label the altered frames based on their texture and micro-patterns. The method employs DRLBP.
[34]	2019	✓	✓		✓		The technique leverages residue data extracted during video decoding by analyzing spatial and temporal energies.
[64]	2019	✓	✓		✓		The correlation between the Haralick-coded frames proved effective for static and dynamic videos.
[52]	2019	✓	✓	✓			The method calculates the temporal mean of non-overlapping subsequences of frames to minimize the number of comparisons and processing time. Utilizing the UQI is essential for evaluating the quality of neighboring TP images and identifying discrepancies that suggest tampering. By examining the UQI values of TP images, the method recognizes the interframe forgery.
[67]	2019				✓		Frames are converted to grayscale and transformed using DCT, and mean features are extracted. A sequence matrix is constructed, and correlation coefficients are calculated to identify duplicated frames. The mean and standard deviation of grayscale values are computed to calculate the coefficient of variation, which distinguishes between original and forged frames.
[58]	2017	✓	✓			✓	Analyzing the spatial and temporal effects on specific video frames. The algorithm classifies videos into three types—single-compressed, double-compressed without tampering, and double-compressed with tampering.
[31]	2016	✓	✓			✓	It detects quantization-error-rich areas in P frames and use them to calculate spatially constrained residual errors. It also employs a wavelet-based algorithm to enrich the traces of quantization error in the frequency domain.

Table 2. *Cont.*

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[65]	2016	✓					It detects frame deletion in the video with the magnitude of the fingerprint using the periodicity of the P-frame prediction error sequence.
[75]	2016	✓					Variation in prediction of residual magnitude and number of intra-macroscopic blocks.
[59]	2015	✓				✓	Visual inspection of DCT patterns on a logarithmic scale provides a simple but effective method for distinguishing between single and recompressed video images. Presented a new method for exploiting the semi-peak light in an optical flow to detect and establish frame deletion in digital videos.
[76]	2015	✓	✓				Tchebyshev's inequality is used twice in localization to determine the locations of insertions and deletions. It focuses on the inconsistency of QCCoLBP values at the tampered frame locations.
[60]	2014	✓	✓		✓		The algorithm consists of extracting DCT coefficients from I at the quantization level and extracting residual errors from the P-frame. It calculates the optical flow between consecutive frames after detecting fake locations based on the consistency of optical flow.
[61]	2014	✓	✓		✓		It checks the consistency of the velocity field in the video to detect interframe forgery. Then, use the generalized ESD test to pinpoint specific locations where these changes occurred in the videos.

Shehnaz and Kaur [71] aim to identify and pinpoint various interframe forgeries, like frame insertion, deletion, and duplication, using a passive-blind approach that does not rely on pre-embedded information. It employs texture features, such as HoG, U-LBP, and R-LBP extracted from grayscale frames. These features are compared using histogram similarity metrics, normalized, and quantized. An SVM classifier is trained on these features to detect tampering, achieving a 99% accuracy rate. The method also localizes tampered frames using Chebyshev's inequality, effectively highlighting forged frames in the video.

Jiang et al. [77] employ a PDTM to reduce MVC, while maintaining rate distortion performance. By utilizing 3D-Sobel operators, a Binocular Just Noticeable Difference (BJND) model, and a Distortion Quantization (D-Q) model, the PDTM estimates perceptual distortion thresholds, enabling the early termination of mode selection in interframe prediction. The results show a 52.9% reduction in computational complexity for dependent views, with a slight bitrate increase of 0.9% under the same subjective quality and 1.0% under the same PSNR, demonstrating significant coding time savings with minimal impact on video quality.

Using a robust hashing algorithm, Niwa et al. [11] propose a novel method for detecting temporally operated videos, such as frame insertion, deletion, reordering, and replacement. The method uses Li et al.'s robust hashing method, which captures both spatial and chromatic features of images using quaternions and generates binary hash values of 120 bits in length. The method also introduces the concept of extended frames, which are synthesized from multiple frames to increase the sensitivity to temporal operations. The technique is assessed on six datasets prepared by applying temporal operations to three videos and uploading them to Twitter and Instagram. The results demonstrate that the method achieves high accuracy and average precision and is robust against recompression and resizing operations. The proposed method is expected to be useful for media forensics and deepfake detection.

Alsakar et al. [16] present an innovative technique for passively detecting video forgeries, utilizing a third-order tensor representation with low computational complexity. The technique capitalizes on the strong correlation of video data in both the spatial and temporal domains, constructing a third-order tensor tube-fiber mode to represent the video subgroups. By selecting an arbitrary number of core tensors and applying an orthogonal transformation, essential features are extracted to detect and locate two types of interframe forgeries—insertion and deletion. The scheme employs correlation analysis, Harris corner detection, and SVD feature extraction to assess the discontinuities in frame sequences and identify forged frames. Through experimental results, it is demonstrated that the proposed technique achieves high accuracy and precision in detecting and locating both types of forgeries across a range of video datasets, encompassing static and dynamic scenes, fast-moving objects, and zooming effects.

Aghamaleki and Behrad [31] introduce a new passive approach for detecting and localizing tampering in MPEG-encoded videos. The proposed algorithm identifies frame insertion, deletion, and double compression, even in the presence of varying GOP structures and lengths. The method is built upon a mathematical analysis of quantization error traces in the residual errors of P-frames. Based on these insights, a new algorithm is developed to detect regions in P-frames with a significant quantization error, while minimizing the impact of motion on residual errors. A wavelet-based technique is then applied to amplify the quantization error traces in the frequency domain. The processed P-frames and spatially constrained residual errors enable the detection and localization of video forgery in the temporal domain. Experimental results and a comparative analysis demonstrate the method's superior performance, particularly for highly compressed videos, and show that it requires an average of 3 s of simulation time.

Zhang et al. [76] present a method for identifying the insertion and deletion of frames within digital videos. This technique consists of two primary stages—feature extraction and detecting irregular points. In the feature extraction phase, the LBP encodes each frame in the video. Then, the ratios of correlation coefficients between successive LBP-encoded frames are calculated. In the detection of abnormal points stage, the localization of insertions and deletions is accomplished by utilizing the Tchebyshev two times, then detecting irregular points using a decision-thresholding approach.

Yu et al. [75] present a method for identifying frame deletions in video streams. Two distinct characteristics have been devised to evaluate the extent of variation in prediction residuals and the number of intra-macroblocks. By combining these attributes, a consolidated index is formed to detect abnormal and sudden changes in video streams. The effectiveness of this approach is tested on a dataset comprising six subsets. The detection capability is evaluated at both the video and GOP levels. The proposed approach exhibits consistent performance across various configurations and surpasses existing methods in different configurations.

Fadl et al.'s [49] suggested approach to detect interframe forgery is based on using HOG and MEI. HOG, extracted from each image, serves as a key distinguishing feature. Correlation factors are then utilized, and abnormal points are detected using Grubb's test to identify frame deletion and insertion instances. Additionally, MEI is applied to the edge images of each shot to detect frame duplication and shuffling. Experimental results demonstrate that the proposed approach can effectively identify all interframe forgeries, while maintaining high accuracy and reducing execution time.

Also, Fadl et al. [78] present a system that identifies interframe forgeries, specifically frame deletion, insertion, and duplication. The suggested method extracts spatiotemporal information and performs deep feature extraction using 2D-CNN. Moreover, we employ a Gaussian Radial Basis Function multi-class support vector machine (RBF-MSVM) for the classification task. Our experiments' findings illustrate the proposed system's efficacy in identifying all categories of interframe tampering, even when the altered videos have experienced a range of post-processing modifications, including Gaussian noise, Gaussian blurring, alterations in brightness, and compression.

Shekar et al. [3] proposed a method for detecting forgery that is simpler and more efficient than existing methods. It requires less computational power and time. It capitalizes on the principle that any tampering with a video alters its original frame sequence, detectable through changes in pixel intensity between adjacent frames. This method analyzes the video by separating the frames and examining the pixel intensity variations to pinpoint counterfeit instances. A consistent pattern of differences between neighboring frames characterizes an unaltered video, whereas alterations introduce noticeable spikes. Identifying these spikes allows for the detection of counterfeit content and its specific location within the video.

Wang et al. [72] proposed an algorithm for detecting video manipulations between adjacent frames. Using band-pass filtering and cubic spline interpolation, the algorithm leverages the ENF signal extracted from the suspected video. Unlike other ENF-based detection methods, this novel approach does not require a reference ENF signal from the power grid. Given the scarcity of ENF signal databases, this reference-free algorithm proves more practical. The core detection mechanism analyzes the correlation factors between neighboring periods in the interpolated ENF signal. A sudden decrease in this coefficient indicates a forgery's presence and precise location. According to the results, the suggested algorithm identifies interframe video forgeries well, such as frame insertion, duplication, and deletion.

Kharat and Chougule [73] proposed a passive blind forgery detection technique for identifying frame duplication attacks in MPEG-4 videos. The method involves a two-step algorithm that utilizes SIFT key points for feature extraction and RANSAC for locating duplicate frames. The research tested the method on 20 videos, achieving an impressive 99.8% average accuracy in detecting tampered frames. The proposed method needs an average of 33 s of simulation time.

Huang et al. [74] introduced the TPFC framework to detect video frame insertion and deletion forgeries. Key contributions of this work involve the introduction of the novel TPD and the extraction of discriminative features using the MLS framework. The effectiveness of the proposed method is evaluated on the open-source Recognition of Human Actions Database, demonstrating its capability in detecting these types of forgeries. The results demonstrate high accuracy in detecting manipulated video frames.

Fadl et al. [43] present a novel method for detecting frame duplication and shuffling video forgery using temporal average and GLCM features. The authors propose a feature extraction method that utilizes the temporal average of each video shot. It employs GLCM to locate statistical textural information and is robust to frame ordering. This approach improves accuracy and reduces computational time. The method is tested on a dataset comprising original videos and videos with various types of forgeries, achieving high accuracy rates with a precision of 0.99 and recall of 0.98, even when the videos are subjected to post-processing operations, such as Gaussian blurring, noise addition, brightness modification, and compression, with an average precision of 0.94 and recall of 0.96. The proposed method needs an average of 5.79 s of simulation time.

Fadl et al. [34] introduce a novel approach to identify interframe forgery in surveillance videos, encompassing frame duplication, insertion, and deletion. The researchers propose a novel framework that leverages residue data obtained during the decoding phase, in conjunction with spatial and temporal energies, to pinpoint anomalies suggestive of forgery. A revised Thompson tau test for anomaly detection is introduced, allowing the use of noise ratios to distinguish between insertion and duplication attacks. Experimental validation is conducted on datasets sourced from SULFA [40], LASIESTA [44], and IVY LAB [45], showcasing notable accuracy and minimal computational overhead, rendering it conducive for real-world forensic applications. A comparative analysis reveals the advantages of the technique over current methodologies, particularly in scenarios involving double compression. The proposed method achieves outstanding performance in detecting interframe forgery, achieving 99% precision, recall, and F1-score for insertion-type forgery. It also attains 97% precision, 99% recall, and 98% F1-score for duplication forgery and 97% precision, 95% recall, and 96% F1-score for deletion forgery. The study concludes by outlining possibilities for enhancing the detection of shuffling forgery and adjusting the methodology to tackle noise-induced distortions in videos. The authors documented that the average execution duration for the suggested approach, utilizing the insertion process, is 53 s, compared to 57 s for the deletion process and 59 s for duplication.

Fadl et al. [52] recommend using the Universal Image Quality Index (UQI) to assess temporal averages of non-blending subsequences of frames, resulting in reduced computational time. The methodology was evaluated through experimentation on a dataset obtained from SULFA [40], LASIESTA [44], and IVY LAB [45] repositories, which included 15 manipulated videos. The results demonstrate precision, recall, and F1-scores of 99% for detecting insertion frame attacks, 96%, 97%, and 96%, respectively, for identifying shuffling attacks, and 98%, 99%, and 98% for detecting deletion frame attacks in video forgery. Additionally, the average processing time per frame remains under 0.03 s.

4.2. Deep Learning Methods

Details on video forgery detection, implemented using deep learning methods and reviewed in this subsection, are summarized in Table 3. Deep learning has seen a significant increase in favor because of its capacity to deliver exceptional results in managing large datasets [79]. This increase can be attributed to the era of the internet, which has enabled access to vast amounts of video data, resulting in a rise in the creation of counterfeit and forged video content. The rapid increase in forgery videos has become a notable concern in various domains, posing challenges related to authenticity and credibility [80]. This section will cover the various recent deep learning techniques used in interframe forgery detection, and Table 3 provides a comprehensive analysis of proposed methods and their key features:

In the phase of video forgery classification, the method evaluates the manipulated video for insertion, forgery, and deletion and then reassesses its originality. The proposed method is designed to be applied in different scenarios, and its performance has been evaluated on two datasets. The results of our proposed model show that our approach achieves a forgery detection accuracy of 91% on the VIFFD dataset [51] and 90% on the TDTV dataset [38]. Additionally, it achieves a forgery classification accuracy of 82% for insertion and deletion forgeries on the VIFFD dataset [51] and 86% on the TDTV dataset [38]. This study contributes to the analysis of original and edited videos across various fields.

In Nguyen et al. [81], the authors posit an innovative strategy employing CNN architectures retrained on ImageNet information. The suggested approach attains a detection precision of 99.17% by leveraging video spatial–temporal correlations. It significantly outperforms current methods and substantially enhances the efficiency and precision of detecting video counterfeits.

Kohli et al. [82] has provided a new approach to detecting and localizing object-based forgery in high-definition videos, which utilizes CNNs. The procedure is executed in two steps. First, a temporal convolution network is used to determine whether frames in a video sequence are solid or two times compressed by looking at the motion residuals using a classifying sliding window approach. Then, a second or spatial CNN is applied to locate the forgery in the solid frames, which are divided into non-overlapping blocks and classified into forged or clean blocks. The focus of the study is the SYSU-OBJFORG dataset [54] comprising 100 clean and 100 forged videos of 720p resolution clips. The method was proven effective, achieving 97.49% frame accuracy, 98.94% double compression detection accuracy, and 96.04% forgery detection accuracy, respectively. It also showed promising results, surpassing the currently accepted approaches in terms of error rate and F1-score, both on the frame level and block level. The method was also found to maintain its effectiveness even after tampering with specific frame properties, such as adjusting video content frame rates. Therefore, the work discussed in this paper makes evident that CNNs are capable of efficiently modeling the spatiotemporal aspects in detecting video forgeries and object-based forgery localization and detection.

Gowda and Pawar [14] focuses on the detection and localization of forgeries between frames, specifically concerning frame insertion and deletion, which are key aspects in digital video forensics. The authors propose using the 3DCNN model for the interframe video forgery detection and localization of forged segments, employing a multi-scale algorithm to measure the structural similarity index. Additionally, they introduce an absolute difference algorithm that facilitates the differentiation of video frames, thereby minimizing temporal redundancy and enabling the identification of counterfeit artifacts in video frames. This enhancement significantly increases the effectiveness of the 3DCNN model in detecting cases of frame insertion and deletion forgery.

Zampoglou et al. [32] presents a video piracy detection technique dependent on forensics and deep learning. The authors use two new forensic filters, Q4 and Cobalt, designed to be manually analyzed by human experts and combine deep steering neural networks (CNNs) for scene classification to generate filters RGB images that reveal potential anomalies in video informational manifestations, such as chiseled regions or compression artifacts. CNNs are trained to distinguish between true and truncated frames using filter output.

Patel and Sheth [83] introduce a new model for detecting interframe forgeries in surveillance videos. The model has three phases—pre-processing, feature extraction, and forgery detection. In the pre-processing phase, video frames are enhanced by removing noise and motionless frames. In the feature extraction phase, six features are extracted from the frames. These features are then utilized in the forgery detection phase, where a CNN is employed to classify frames as either manipulated or pristine. The use of a hybrid optimization approach adjusts the CNN weights. The model also performs tamper localization if a frame is detected as tampered with.

Chandru and Priscilla [84] introduce an advanced method that combines CNNs for frame-level analysis with RNNs to capture temporal dynamics in video sequences. This multi-modal framework enables the system to detect subtle inconsistencies and manipulation artifacts within video content effectively. Although the specific dataset used for training and evaluation is not explicitly mentioned, the model is built with adaptability in mind, utilizing domain adaptation and adversarial training techniques to enhance its resilience against emerging deepfake technologies. The results demonstrate that the system achieves high accuracy in differentiating authentic videos from manipulated ones, while being optimized for real-time performance. The key contribution of this study is the development of a scalable and efficient video integrity detection system that not only enhances the accuracy of identifying altered content but also fosters trust and transparency in digital media, ultimately supporting a more secure digital environment.

Akhtar et al. [8] propose a new method combining 2D-CNN for feature extraction. LSTM/GRU upscales autoencoder to analyze long-term dependencies on video frames. This approach addresses significant challenges in existing processes, including limited applicability, poor generalization, and the localization of forgery. The method achieves an accuracy of above 90% in detecting spoofing and localization across various video formats, frame rates, and compression properties, demonstrating its strength and effectiveness.

Kumar et al. [10] proposed a method that can detect these forms of falsification and is unique in its ability to identify the location of the forgery. It establishes the relationship between neighboring frames by utilizing the correlation coefficient. Furthermore, it determines the interframe correlation distance among frames and assesses the minimum distance score in conjunction with statistical properties. It analyzes the upper and lower threshold values, including the sigma coefficient, to identify the localization of forgery. This approach also differentiates between types of forgery insertion and deletion by applying threshold-controlled parameters. The effectiveness of this method has been verified through its application to the VIFFD dataset [51]. This proposed method achieves a frame-level accuracy of 97% and a video-level accuracy of 83% in identifying forgery.

Singla et al. [85] present a video forensic technique that uses CNNs to generate and match integrity embeddings of surveillance videos to detect frame duplication forgery. The method evaluates various deep learning models for embedding generation and prepares a dataset comprising over 400 original and forged HEVC-coded surveillance videos for experimentation. The results show that the technique outperforms traditional methods that rely on hand-crafted or compression domain features in terms of accuracy and reliability.

Table 3. Deep learning methods for video forgery detection.

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[13]	2024	✓	✓				The system involves four stages—preprocessing, feature extraction using VGG-16, feature selection with KPCA, and correlation analysis to detect forgeries.
[8]	2024	✓	✓				It involves 2D-CNN for feature extraction, an autoencoder for dimensionality reduction, and LSTM/GRU for analyzing long-range dependencies in video frames.
[12]	2023	✓	✓	✓			A spatiotemporal averaging method was used to extract background and moving objects for video forgery detection. Features were extracted using the GoogleNet model to obtain feature vectors. The UFS-MSRC method was employed to select the most crucial feature vectors, thereby accelerating training and enhancing detection accuracy. An LSTM network was applied to detect forgery in different video sequences.

Table 3. *Cont.*

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[14]	2023	✓	✓				It uses a 3DCNN model to detect and locate video forgeries between frames. It uses a multi-scale algorithm to measure the similarity of video frames and find the forged segments. It uses an absolute difference algorithm to reduce the redundancy of video frames and spot the fake artifacts.
[85]	2022				✓		CNN is employed to detect duplicates, locate video frames, and create and match integrity embeddings.
[36]	2022		✓				The method uses a VFID-Net to extract deep features from adjacent frames and computes the correlation coefficients and distances between them. It also employs a threshold control parameter and a minimum distance score to classify the frames as either original or inserted.
[10]	2022	✓	✓				It calculates the correlation factors of the deep features of neighboring frames to detect forgeries.
[15]	2022	✓	✓	✓	✓	✓	3D-CNN is the foundation upon which the suggested model is constructed. It uses pre-processing techniques to provide different frames that detect differences between consecutive adjacent frames.
[83]	2021	✓	✓		✓		It uses a hybrid optimization algorithm, called MO-BWO, that combines the mayfly optimization (MO) and the black widow optimization (BWO) algorithms to fine-tune the weights of the CNN and improve its performance. It uses prediction residual gradient (PRG) and optical flow gradient (OFG) to automatically locate the tampered frames and identify the type of forgery within the GOP structure.
[2]	2021				✓		It extracts features and converts videos into frames. The I3D network then detects frame-to-frame duplication by examining an original video alongside a forged one. RNN is responsible for detecting sequence-to-sequence forgery in videos.
[16]	2021	✓	✓	✓			It exploits the high correlation of video data in both spatial and temporal domains and constructs a third-order tensor tube-fiber mode to represent video subgroups. It uses correlation analysis, Harris corner detection, and SVD feature extraction to measure the discontinuity of frame sequences and identify the forged frames.
[78]	2021	✓	✓		✓	✓	To extract spatiotemporal information from video frames, a 2D-CNN is utilized to identify the video frames, and a Gaussian radial basis function (RBF-MSVM) is utilized.
[82]	2020					✓	It detects and localizes object-based forgery in high-definition videos using CNNs. The approach consists of two steps, as follows: it uses a temporal CNN to categorize the frames of a video as either solid or double compressed, primarily based on the motion residuals extracted from a sliding window technique.

Table 3. *Cont.*

Ref.	Year	FrameDeletion	FrameInsertion	FrameShuffling	FrameDuplication	DoubleCompression	Short Description
[81]	2020	✓	✓	✓	✓		It uses retrained CNN models (e.g., MobileNetv2, ResNet18) on the ImageNet dataset. It combines residuals of adjacent frames and optical flow for robust detection.
[32]	2019	✓	✓				It uses two novel forensics filters, Q4 and Cobalt, designed for manual verification by human experts, and combines them with deep CNNs for visual classification.

Shelke and Kasana [13] proposed a method for detecting forgery that incorporates various types of forgeries in digital videos using the VGG-16 deep neural model and KPCA. It spans pre-processing, feature extraction, feature selection, and analyzing correlations with other variables, such as noise or brightness level variations, before ultimately determining the types of forgeries that were applied to our video when it was being composed or rendered, among many others. The system attained an exceptional accuracy level of 97.24% and a precision of 96.86%.

Oraibi and Radhi [15] proposed a system based on the 3D-CNN model, which can extract spatial and temporal features derived from videos. The system also uses different frames to capture successive frames 'changes, providing valuable information for forgery detection. The system is evaluated on two datasets—the UCF101 dataset [55], which contains 101 human action categories, and a selected dataset comprising videos recorded by a surveillance camera. The system achieves an average accuracy of 99.14% for detecting different types of interframe video forgery.

Paper [36] proposes a novel method for detecting and locating interframe video forgeries based on deep features and correlation analysis. The method uses VFID-Net to extract deep features from adjacent frames and compute the correlation coefficients and distances between them. The method also employs a threshold control parameter and a minimum distance score to classify the frames as either original or inserted. The method is evaluated on two standard datasets—VIFFD and SULFA—which contain videos with insertion-type forgeries.

Munawar and Noreen [2] expand on the challenge of identifying frame duplication across diverse frame rates by applying a deep learning approach, specifically a pioneering deep learning architecture incorporating inflated 3D and Siamese-based RNN. The initial step in the proposed framework involves extracting features and converting videos into frames. The I3D network detects frame-by-frame duplication by examining an original video alongside a forged one. Subsequently, several frames are combined to create a scene that is passed to the Siamese-based RNN. This RNN is responsible for detecting frame-by-frame forgery in videos. To evaluate the proposed model, the relatively large MFC dataset, which encompasses different frame rates and a considerable volume of videos, as well as the VIRAT datasets, are utilized. The accuracy of the proposed method, as evaluated using the VIRAT dataset, is 86.6%. With the MFC dataset, it reaches 93%.

Hu et al. [86] propose an Attention-Erasing Stripe Pyramid Network (ASPNet) as an advanced approach to face forgery detection, designed to improve generalization across different forgery datasets. The latter is achieved by leveraging high-frequency noise and integrating RGB and fine-grained frequency clues to enhance detection accuracy. Towards not losing complementary features, ASPNet employs the Stripe Pyramid Block (SPB), which simultaneously extracts multi-scale and multi-granularity features. To further refine the model's focus and suppress unwanted noise, the authors incorporate a Two-Stage Attention Block (TSAB), filtering out pixel- and channel-wise interference. Finally, an attention erasing (AE) structure is implemented to dynamically direct the model's focus, prompting it to analyze different areas of the human face. Extensive experiments conducted on the FaceForensics++ dataset report ASPNet's performance metrics AUC of 77.4% and accuracy of 70.85%, surpassing existing benchmarks and proving its effectiveness.

5. Performance Analysis of Interframe Forgery Detection Techniques

In this section, we present a comprehensive summary of various state-of-the-art methods for detecting interframe video forgeries, detailing their performance metrics across different datasets. It is crucial for researchers in multimedia forensics, as it offers a consolidated view of the effectiveness of different approaches. Each method is evaluated based on its accuracy, precision, recall, and F1-score, which are critical performance indicators in real-world applications. The previous section highlights techniques, from traditional spatial and temporal analysis methods to advanced deep learning models, such as CNNs and LSTM networks. By comparing these metrics, the table enables a better understanding of the trade-offs between different approaches, their robustness against various types of forgeries, and their measurability across multiple datasets. This comparison highlights areas that require further investigation and development, while also helping to identify the most feasible approaches.

5.1. Evaluation Metrics

Table 4 summarizes the key performance metrics used across the reviewed studies, including Forged Frame Accuracy (FFACC), Detection Accuracy (DA), and F1-score. These metrics collectively provide insights into classification reliability, precision under imbalanced conditions, and detection consistency. Metrics like the Probability of Error (Pe) and True Positive Rate (TPR) further clarify the system's error behavior under varying conditions.

Table 4. Evaluation metrics and their descriptions.

Metric	Description	Optimal Value	Mathematical Formula
Double-compressed frame accuracy (DFACC)	The accuracy of correctly classifying double-compressed frames is determined by dividing the number of correctly classified double-compressed frames by the total number of double-compressed frames, accurately reflecting the effectiveness of the classification process in identifying double-compressed frames.	1	$\text{DFACC} = \frac{\text{correctly classified double compressed frames}}{\text{total no.of double compressed frames}}$

Table 4. *Cont.*

Metric	Description	Optimal Value	Mathematical Formula
Forged frame accuracy (FFACC)	The accuracy of correctly classifying forged frames is calculated by dividing the number of correctly classified forged frames by the total number of forged frames. This metric evaluates the system's ability to accurately identify forged frames.	1	$\text{FFACC} = \frac{\text{correctly classified forged frames}}{\text{total no.of forged frames}}$
Frame accuracy (FACC)	The accuracy is calculated by dividing the number of correctly classified forged frames by the total number of forged frames. It is used to assess the performance of frame-level detection. This metric provides insight into how accurately the system detects forgeries at the individual frame level.	1	$\text{FACC} = \frac{\text{correctly classified frames}}{\text{total no.of frames}}$
Precision	Precision measures the proportion of true positive predictions among all positive predictions, indicating how many of the detected positives are correct. It is calculated by dividing the number of true positive predictions (TP) by the total number of positive predictions, including both true and false positives (FP).	1	$\text{P} = \frac{\text{TP}}{\text{TP+FP}}$
Recall	Recall measures the proportion of true positive predictions relative to all actual positive instances, reflecting the system's ability to identify all relevant instances. It is calculated by dividing the number of true positive predictions (TP) by the total number of actual positives, including both true and false negatives (FN).	1	$\text{R} = \frac{\text{TP}}{\text{TP+FN}}$
Probability of error (Pe)	The error rate measures how often a classification system makes mistakes. It is calculated by subtracting the correct predictions (true positives and true negatives) from the total number of instances. This shows the chance of the system making a wrong prediction, including false positives and negatives.	0	$\text{P}_\text{e} = 1 - \frac{\text{TP+TN}}{\text{P+N}}$

Table 4. *Cont.*

Metric	Description	Optimal Value	Mathematical Formula
F1-score	The F1-score is a metric that balances precision and recall, providing a single measure of a model's accuracy, particularly useful when the data has an uneven class distribution. It is calculated by taking the harmonic mean of precision and recall to assess the model's performance, combining precision and recall into a single value that considers both false positives and false negatives.	1	$\mathrm{F1} = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P}+\mathrm{R}}$
Detection accuracy (DA)	This refers to the proportion of correct predictions (whether authentic or forged) out of the total number of predictions made by the model. It measures the model's overall effectiveness in correctly classifying authentic and forged instances.	1	$\mathrm{DA} = \frac{\mathrm{TP}+\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}}$
The proportion of positive cases (TPR)	TPR stands for the percentage of positive cases—forged frames that are successfully categorized.	1	$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
Video accuracy classification (VAC)	The percentage of accurately categorized video segments (CCFVS) to all video segments (N) in a video.	1	$\mathrm{VAC} = \frac{\mathrm{CCFVS}}{\mathrm{N}}$

5.2. Performance of Machine Learning-Based Video Forgery Detection Methods

Table 5 includes the performance of various machine learning-based forgery detection methods. The table contains details on the employed method, as well as the dataset used.

Table 5. Performance of various machine learning-based video forgery detection methods.

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[59]	DCT coefficient and optical flow analysis	SULFA, change detection video database	0.993	-	-	-
[60]	Optical flow analysis and anomaly detection	TRECVID	[0.867, 0.933]	-	-	-
[61]	Consistency of the velocity field	TRECVID	[0.80, 0.90]	-	-	-
[30]	Texture analysis for spatial video forgery detection and localization	SULFA	0.96	-	-	-

Table 5. *Cont.*

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[29]	A methodology for detecting and localizing forgery using similarity metrics and feature classification	VTL	0.9492	0.9589	0.9655	0.9622
[7]	Detection and localization of frame duplication using a binary image template	SULFA, VLFD, VTL	-	621	0.9863	0.993
[9]	Duplicate frame detection using sequence matching	REWIND	0.989	-	-	-
[66]	Passive video forgery detection based on temporal information and Structural Similarity Index	TDTVD	94.44%	-	-	-
[71]	Detection and localization of multiple interframe forgeries in digital videos	SULFA	[HOG 0.992, R-LBR 0.996]	[HOG 0.992, R-LBR 0.994]	[HOG 0.991, R-LBR 0.992]	[HOG 0.991, R-LBR 0.995]
[69]	An algorithm for detecting and localizing frame duplication forgeries based on the increased Levenshtein distance	VTL	-	0.995	0.1	0.9975
[70]	An ensemble approach to detect frame duplication	FD&FDs	0.9932	-	-	-
[11]	A robust hashing-based temporally operated video detection method	SELF-MADE	0.9825	1	-	-
[31]	Quantization errors in video coding for utilizing the inherent effects of double compression	VTL	-	0.89	0.86	0.87
[65]	Video interframe forgery counter-forensics	VTL	0.969	-	-	-
[76]	Detection of inconsistencies in correlations between locally binary pattern-coded frames for effective video frame insertion and deletion	VTL	-	0.95	0.92	0.93

Table 5. *Cont.*

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[77]	Detecting frame deletion by detecting abrupt modifications in video streams	VTL	-	0.72	0.66	0.69
[49]	Using a motion energy image and an oriented gradient histogram	VTL	-	0.98	0.99	0.98
[3]	Detecting and localizing interframe video forgery using differences	VIFFD, TDTVD	[VIFFD 0.866, TDTVD 0.9192]	-	-	-
[73]	A method of passively detecting blind forgeries to detect frame duplication	SELF-MADE	0.998	0.999	0.997	-
[74]	A new approach to video forgery detection using triangle polarity feature classification	Recognition of Human Actions	-	0.9576	0.9826	-
[43]	An approach for detecting frame duplication and shuffling forgeries in surveillance recordings that uses a temporal average and a gray-level co-occurrence matrix	FD&FDs	-	[0.94, 0.99]	[0.96, 0.98]	-
[34]	Detection of interframe forgeries using residue differential energy	SULFA, LASIESTA, IVY LAB	-	[Dup0.97, Ins0.99, Del0.97]	[Dup 0.99, Ins 0.99, Del 0.95]	[Dup 0.98, Ins 0.99, Del 0.96]
[52]	Using the temporal average of the universal picture quality index for surveillance video authentication	SULFA, LASIESTA, IVYLAB	-	[Ins0.99, Shuf0.96, Del0.98]	[Ins 0.99, Shuf0.97, Del0.99]	[Ins 0.99, Shuf0.96, Del0.98]
[63]	A method for detecting frame shuffling in MPEG-coded video	SELF-MADE	-	1	1	-
[64]	Variations in the correlation distribution of frames coded with Haralick	VTL	-	0.85	0.89	0.87
[58]	Quantization effect studies in the spatial and temporal domains applied to MPEG videos	VTL	83.39	-	-	-

Table 5. *Cont.*

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[67]	Detection of video frame and region duplication forgeries using the correlation and coefficient of variation	SULFA	0.995	1	0.99	0.994

5.3. *Performance of Deep Learning-Based Video Forgery Detection Methods*

Table 6 includes details on the performance of deep learning-based forgery detection methods.

Table 6. Performance of various deep learning-based video forgery detection methods.

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[13]	Digital video multiple forgery detection using a deep neural network based on VGG-16 and KPCA	SULFA, VTL	0.97	0.96	-	0.958
[8]	Spatiotemporal analysis-based deep learning identification and localization of different kinds of interframe forgery videos	SELF-MADE	[Ins 0.9898, Del 0.9418]	-	-	-
[82]	CNN-based localization of the forged region	SYSU-OBJFORG	[FACC 0.974, DFACC 0.989]	-	-	-
[12]	Using the LSTM network and UFS-MSRC algorithm for forgery detection	SULFA, Sondos	[SULFA 0.9813, Sondos 0.9738]	-	-	-
[14]	Identification and localization of forgeries in videos using deep learning	VIFFD, UCF-101	0.98	0.96	0.95	0.97
[32]	Multimedia forensics and deep learning for tampered video detection	FVC	0.85	-	-	-
[36]	Using parallel convolutional neural network—VFID-Net	VIFFD, SULFA	[VIFFD 0.865, SULFA 0.92]	-	-	0.87
[83]	CNN-based detection model with multi-feature extraction framework	-	0.85	0.825	-	0.826

Table 6. *Cont.*

Ref.	Method	Dataset	Accuracy	Precision	Recall	F1-Score
[85]	Frame duplication detection with PCA and agglomerative clustering using CNN-based features	SELF-MADE	0.94	-	-	-
[10]	Multiple forgery detection in videos using convolution neural network	VIFFD, TDTVD	[VIFFD 0.82, TDTVD 0.86]	-	-	-
[15]	Deep learning technique for improving the digital forensic approach for interframe video forgery detection	UCF-101	0.9914	-	-	-
[2]	Using Siamese-based RNN to duplicate frame video forgery detection	MFC, VIRAT	[MFC 0.866, VIRAT 0.933]	[MFC 0.875, VIRAT 0.933]	[MFC 0.866, VIRAT 0.933]	[MFC 0.865, VIRAT 0.933]
[16]	Using third-order tensor representation with low computational complexity as the basis	VTL	-	0.99	0.99	0.99
[78]	Detecting surveillance video forgeries using CNN spatiotemporal characteristics and fusion	VIRAT, SULFA, LASIESTA, IVY LAB	0.983	-	-	-
[81]	Using a convolutional neural network model to identify video interframe forgeries	VIFFD	0.9917	-	-	-

6. Discussion

In this section, a statistical analysis was conducted on the research papers published in renowned academic journals, such as Springer, IEEE, Science Direct, and other Scopus-indexed sources, between 2019 and 2024. Based on the comparative analysis presented in Section 4, The classification of these publications was based on the methodologies employed, specifically focusing on whether they utilized deep learning or machine learning techniques, as illustrated in Figure 9.

A five-year study of scholarly publications found that fifteen papers utilized deep learning methods for interframe forgery detection, while thirty studies used machine learning techniques. These trends show a gradual shift from traditional feature-based methods to data-driven deep learning frameworks. Conventional methods, such as SSIM, optical flow estimation, and histogram analysis, remain helpful, because they are efficient and easy to understand. However, they often do not perform well in dynamic scenes or post-processing conditions, such as SSIM, which might miss refined temporal inconsistencies when global illumination changes. Additionally, optical flow-based techniques often have difficulty with fast camera motion or complex background dynamics. Deep learning models, such as VGG-16 and LSTMs, perform well in challenging situations. VGG-16 is great at learning spatial features to spot forgeries in single frames. LSTMs and combined CNN-LSTM models are good at understanding how things change over time. This helps detect insertions, deletions, and shuffles, particularly in videos with significant movement.

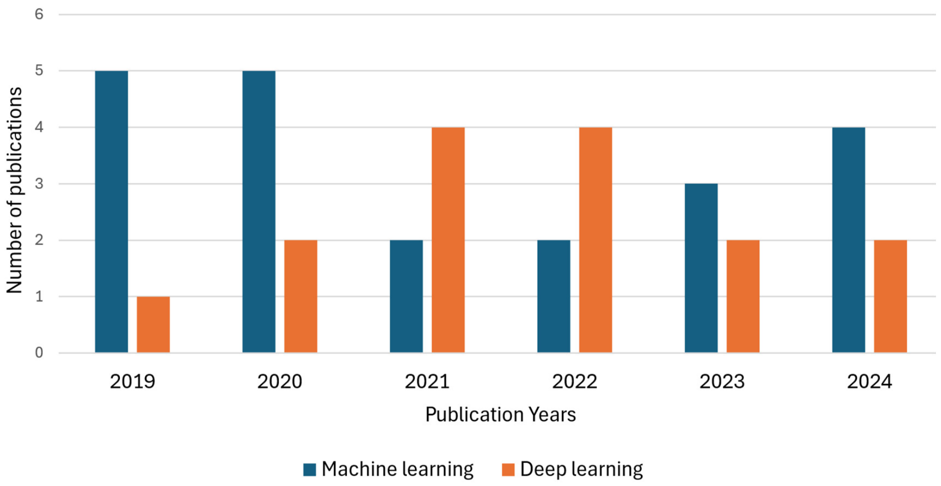

Figure 9. The number of published studies between 2019 and 2024.

In the field of interframe forgery detection, researchers have used various datasets in their studies. Some datasets, such as VTL and SULFA, have been widely used, appearing in 20.69% and 18.97% of the studies, respectively, indicating their strong relevance and reliability. However, some datasets are widely adopted, such as SULFA, but their usage has decreased over the last few years due to the difficulty of accessing them online. Another commonly used dataset is VIFFD, which has been adopted in 10.34% of the studies. On the other hand, datasets such as IVY LAB LASIESTA, FD&FDs, UCF-101, and TRECVid have been utilized to a lesser extent, each in 3.45% of the papers. Some datasets, like MFC and SYSU-OBJFORG, have seen marginal use at 1.72% due to unspecified forgery detection.

Figure 10 illustrates these trends, demonstrating the varying levels of dataset adoption, and suggesting that while some datasets are widely accepted, alternative datasets are still valuable for specific research needs. When resolution is a key factor in the study.

HTVD, VLFD, and VIFFD are the top choices among datasets. Additionally, VIFFD, HTVD, and TDTVD contain the most significant number of videos related to forgery detection.

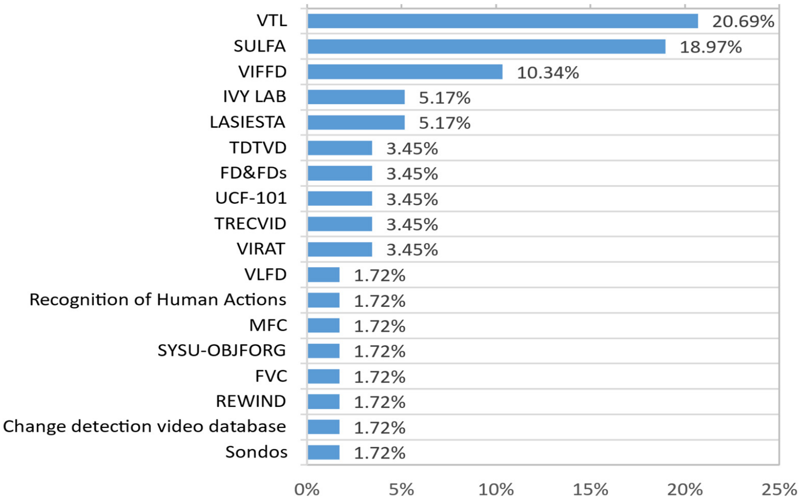

Figure 10. Percentage of datasets used in research papers between 2019 and 2024.

6.1. Limitations and Challenges

The analysis of interframe forgery detection techniques in Sections 4 and 5 reveals critical limitations and challenges that hinder their effectiveness. Below, we elaborate on these issues, while integrating insights from recent studies.

6.1.1. Challenges in Detecting Video Frame Shuffling and Duplication

Challenges in detecting video frame shuffling and duplication include the following aspects:

- Differentiating between natural video compression artifacts and inconsistencies caused by frame shuffling [35,60].
- Shuffled frames often retain similar statistical properties to authentic frames, making detection difficult [43].
- Advanced editing tools produce forgeries that closely resemble authentic frames, complicating detection efforts and reducing the efficacy of traditional methods, such as SSIM or optical flow estimation [62].
- Manipulations in real-time videos present significant challenges, especially when no reference videos are available [1].
- Processing large volumes of video data, especially in real time, requires significant computational resources [77].
- Few methods can effectively detect duplicate frames regardless of their order or the total count of forged frames.
- Detection accuracy decreases in videos with dynamic camera movements, limiting the application of specific detection methods [1].

6.1.2. Challenges in Detecting Frame Insertion and Deletion

Challenges in detecting frame insertion and deletion include the following aspects:

Many methods fail to detect anomalies in static scenes, especially when specific frames are removed and overlooked [75].

- Deleted frames can go unnoticed if the remaining sequence maintains visual consistency [62].
- Frame-by-frame analysis required for detecting insertion and deletion in real time is computationally expensive [76].

6.1.3. Challenges in Detecting Double Compression and Post-Processing Forgeries

Challenges in detecting double compression and post-processing forgeries include the following aspects:

- Differentiating between video compression artifacts and genuine forgery traces, especially with multiple compression layers [35].
- Techniques such as noise addition can mask forgery traces, reducing the accuracy of detection algorithms [60].
- Current methods are often not robust enough to detect forgeries after extensive post-processing, such as compression or noise addition [58].

6.2. Search Directions

Interframe forgery detection in digital videos involves various search directions to ensure the integrity and authenticity of multimedia data:

- Real-Time Forgery Detection Limitations:
  Current methods struggle with the real-time analysis of large-scale videos due to computational obstacles and low detection accuracy. Future work should focus on lightweight architectures and perceptual distortion thresholds to strike a balance between speed and precision [1,77].

- GOP Size and Forgery Classification:
  Fixed GOP sizes and the absence of forgery classification constrain current studies. Future research could enhance detection by investigating how adaptive GOPs impact forgery identification, and it could also explore the classification of forgery types based on fundamental manipulation techniques [35,60].
- Specific Forgery Types:
  Current methods are primarily designed for specific types of forgery, such as frame insertion and deletion, but may not be effective against other types of manipulation. Hybrid frameworks combining multiple forgery detection strategies, such as Siamese-RNN for duplication [2] and CNN-LSTM for shuffling [12], could enhance adaptability.
- Applicability to Dynamic Scenes:
  Most techniques are limited to static scenes, and expanding to dynamic or varied forgery operations would enhance robustness and generalizability. Leveraging MEI [49] and texture analysis [30] can improve robustness against camera motion and background variability.
- Computational Complexity:
  High computational costs limit real-time deployment. Optimizing algorithms for parallel processing and integrating low-complexity models, such as UFS-MSRC, with LSTM networks [12] could reduce resource demands.
- Post-Processing Robustness:
  Detection accuracy can be affected by sudden changes in video features, such as compression or noise addition. Designing anti-forensics-resistant algorithms [62] and multi-layer compression artifact isolation techniques [31,60] is critical for resilience.
- Integration with Machine Learning/Deep Learning:
  Hybrid models can enhance performance by combining deep learning techniques, such as CNN-LSTM and VGG-16, with more conventional approaches, like LBP and SSIM. For example, VGG-16 excels in spatial feature extraction for forged frames [13], while LSTM networks capture temporal inconsistencies [12].
- Comprehensive Datasets:
  There is a need for more extensive datasets that cover multiple types of forgery operations, facilitating the development of more generalized models.

7. Conclusions

Our survey on interframe forgery video detection has highlighted the importance of addressing this vital issue in digital video forensics. Through a comprehensive analysis of various datasets, methods, challenges, search directions, and applications, it is evident that detecting interframes is a rapidly evolving and intricate process, requiring advanced methods and ongoing research. Key findings reveal that deep learning methods provide better flexibility in dynamic scenes. They consistently outperform traditional techniques in various situations. Our study serves as a guide by presenting current methods, identifying weaknesses in existing datasets, and highlighting gaps in evaluation that limit generalization and real-time use. These insights have significant potential for critical real-world applications, such as forensic investigations, digital surveillance, video authentication, and maintaining the integrity of multimedia content. Future research should focus on hybrid detection models that combine traditional and deep learning methods, thereby enhancing real-time flexibility. Despite recent developments, interframe forgery detection still presents unresolved issues, making it a promising field for both current and future research. Researchers can develop scalable and practical solutions that preserve the authenticity of video content in a rapidly evolving digital landscape.

Author Contributions: Conceptualization, K.M.H.; methodology, N.I.G. and K.M.H.; investigation, M.M.A. and H.M.H.; resources, M.M.A. and H.M.H.; data curation, M.M.A.; writing—original draft preparation, M.M.A., N.I.G., H.M.H. and K.M.H.; writing—review and editing, N.I.G., H.M.H., K.M.H., E.V. and G.A.P.; visualization, N.I.G. and K.M.H.; supervision, K.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Networks
2D-CNN	Two-Dimensional Convolutional Neural Network
3DCNN	Three-Dimensional Convolutional Neural Network
RNN	Recurrent Neural Networks
DCT	Discrete Transform
BAS	Block Artifact Strength
VPF	Variation In Prediction Footprint
SSIM	The Structural Similarity Index Measure
LSTM	Long Short-Term Memory Network
LBP	Local Binary Pattern
DRLBP	The Discriminative Robust Local Binary Pattern
U-LBP	Uniform Local Binary Pattern
R-LBP	Rotation Invariant Local Binary Pattern
GLCM	Gray Level Co-Occurrence Matrix
MEI	Motion Energy Image
HOG	Histogram of Oriented Gradients
ENF	Electrical Network Frequency
SIFT	Scale Invariant Feature Transform
RANSAC	Random Sample Consensus
MLS	Multi-Level Subtraction
TPD	Triangular Polarity Detector
VGG	Visual Geometry Group
KPCA	Kernel Principal Component Analysis
MVC	Computational Complexity in Multi-View Video Coding
PDTM	Perceptual Distortion Threshold Model

References

1. Nam, S.-H.; Park, J.; Kim, D.; Yu, I.-J.; Kim, T.-Y.; Lee, H.-K. Two-stream network for detecting double compression of H.264 videos. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 111–115.
2. Munawar, M.; Noreen, I. Duplicate Frame Video Forgery Detection Using Siamese-based RNN. *Intell. Autom. Soft Comput.* **2021**, *29*, 927–937. [CrossRef]
3. Shekar, B.; Abraham, W.; Pilar, B. A Simple Difference Based Inter Frame Video Forgery Detection and Localization. In Proceedings of the International Conference on Soft Computing and its Engineering Applications, Anand, India, 7–9 December 2023; pp. 3–15.
4. Verdoliva, L. Media forensics and deepfakes: An overview. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 910–932. [CrossRef]
5. El-Shafai, W.; Fouda, M.A.; El-Rabaie, E.-S.M.; El-Salam, N.A. A comprehensive taxonomy on multimedia video forgery detection techniques: Challenges and novel trends. *Multimed. Tools Appl.* **2024**, *83*, 4241–4307. [CrossRef]
6. Sandhya; Kashyap, A. A comprehensive analysis of digital video forensics techniques and challenges. *Iran J. Comput. Sci.* **2024**, *7*, 359–380. [CrossRef]
7. Bozkurt, I.; Ulutaş, G. Detection and localization of frame duplication using binary image template. *Multimed. Tools Appl.* **2023**, *82*, 31001–31034. [CrossRef]

8. Akhtar, N.; Hussain, M.; Habib, Z. DEEP-STA: Deep Learning-Based Detection and Localization of Various Types of Inter-Frame Video Tampering Using Spatiotemporal Analysis. *Mathematics* **2024**, *12*, 1778. [CrossRef]
9. Mohiuddin, S.; Malakar, S.; Sarkar, R. Duplicate frame detection in forged videos using sequence matching. In Proceedings of the International Conference on Computational Intelligence in Communications and Business Analytics, Kalyani, India, 27–28 January 2021; pp. 29–41.
10. Kumar, V.; Kansal, V.; Gaur, M. Multiple Forgery Detection in Video Using Convolution Neural Network. *Comput. Mater. Contin.* **2022**, *73*, 1347–1364. [CrossRef]
11. Niwa, S.; Tanaka, M.; Kiya, H. A Detection Method of Temporally Operated Videos Using Robust Hashing. In Proceedings of the 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 18–21 October 2022; pp. 144–147.
12. Girish, N.; Nandini, C. Inter-frame video forgery detection using UFS-MSRC algorithm and LSTM network. *Int. J. Model. Simul. Sci. Comput.* **2023**, *14*, 2341013. [CrossRef]
13. Shelke, N.A.; Kasana, S.S. Multiple forgery detection in digital video with VGG-16-based deep neural network and KPCA. *Multimed. Tools Appl.* **2024**, *83*, 5415–5435. [CrossRef]
14. Gowda, R.; Pawar, D. Deep learning-based forgery identification and localization in videos. *Signal Image Video Process.* **2023**, *17*, 2185–2192. [CrossRef]
15. Oraibi, M.R.; Radhi, A.M. Enhancement digital forensic approach for inter-frame video forgery detection using a deep learning technique. *Iraqi J. Sci.* **2022**, *63*, 2686–2701. [CrossRef]
16. Alsakar, Y.M.; Mekky, N.E.; Hikal, N.A. Detecting and locating passive video forgery based on low computational complexity third-order tensor representation. *J. Imaging* **2021**, *7*, 47. [CrossRef]
17. Zhong, J.-L.; Gan, Y.-F.; Vong, C.-M.; Yang, J.-X.; Zhao, J.-H.; Luo, J.-H. Effective and efficient pixel-level detection for diverse video copy-move forgery types. *Pattern Recognit.* **2022**, *122*, 108286. [CrossRef]
18. Tyagi, S.; Yadav, D. A detailed analysis of image and video forgery detection techniques. *Vis. Comput.* **2023**, *39*, 813–833. [CrossRef]
19. Singh, G.; Singh, K. Copy-Move Video Forgery Detection Techniques: A Systematic Survey with Comparisons, Challenges and Future Directions. *Wirel. Pers. Commun.* **2024**, *134*, 1863–1913. [CrossRef]
20. Fayyaz, M.A.; Anjum, A.; Ziauddin, S.; Khan, A.; Sarfaraz, A. An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues. *Multimed. Tools Appl.* **2020**, *79*, 5767–5788. [CrossRef]
21. Verde, S.; Bondi, L.; Bestagini, P.; Milani, S.; Calvagno, G.; Tubaro, S. Video codec forensics based on convolutional neural networks. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 530–534.
22. D'Avino, D.; Cozzolino, D.; Poggi, G.; Verdoliva, L. Autoencoder with recurrent neural networks for video forgery detection. *arXiv* **2017**, arXiv:1708.08754. [CrossRef]
23. Thajeel, S.A.; Sulong, G.B. State of the art of copy-move forgery detection techniques: A review. *Int. J. Comput. Sci. Issues (IJCSI)* **2013**, *10*, 174.
24. Mizher, M.A.; Ang, M.C.; Mazhar, A.A.; Mizher, M.A. A review of video falsifying techniques and video forgery detection techniques. *Int. J. Electron. Secur. Digit. Forensics* **2017**, *9*, 191–208. [CrossRef]
25. Srivalli, D.; Tech, M.; Sri, D.; Begum, M.; Prakash, C.; Kumar, S.; Pallavi, P. Video Inpainting with Local and Global Refinement. *Int. J. Sci. Res. Eng. Manag.* **2024**, *8*, 3. [CrossRef]
26. Nabi, S.T.; Kumar, M.; Singh, P.; Aggarwal, N.; Kumar, K. A comprehensive survey of image and video forgery techniques: Variants, challenges, and future directions. *Multimed. Syst.* **2022**, *28*, 939–992. [CrossRef]
27. Mondaini, N.; Caldelli, R.; Piva, A.; Barni, M.; Cappellini, V. Detection of malevolent changes in digital video for forensic applications. In Proceedings of the Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, USA, 29 January–1 February 2007; pp. 300–311.
28. Li, Q.; Wang, R.; Xu, D. A video splicing forgery detection and localization algorithm based on sensor pattern noise. *Electronics* **2023**, *12*, 1362. [CrossRef]
29. Singla, N.; Nagpal, S.; Singh, J. A two-stage forgery detection and localization framework based on feature classification and similarity metric. *Multimed. Syst.* **2023**, *29*, 1173–1185. [CrossRef]
30. Saddique, M.; Asghar, K.; Bajwa, U.I.; Hussain, M.; Habib, Z. Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames. *Adv. Electr. Comput. Eng.* **2019**, *19*, 97–108. [CrossRef]
31. Aghamaleki, J.A.; Behrad, A. Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding. *Signal Process. Image Commun.* **2016**, *47*, 289–302. [CrossRef]
32. Zampoglou, M.; Markatopoulou, F.; Mercier, G.; Touska, D.; Apostolidis, E.; Papadopoulos, S.; Cozien, R.; Patras, I.; Mezaris, V.; Kompatsiaris, I. Detecting tampered videos with multimedia forensics and deep learning. In *Proceedings of the MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, 8–11 January 2019*; Proceedings, Part I 25; Springer: Berlin/Heidelberg, Germany, 2019; pp. 374–386.
33. Sitara, K.; Mehtre, B. Detection of inter-frame forgeries in digital videos. *Forensic Sci. Int.* **2018**, *289*, 186–206.

34. Fadl, S.M.; Han, Q.; Li, Q. Inter-frame forgery detection based on differential energy of residue. *IET Image Process.* **2019**, *13*, 522–528. [CrossRef]
35. Kingra, S.; Aggarwal, N.; Singh, R.D. Video inter-frame forgery detection approach for surveillance and mobile recorded videos. *Int. J. Electr. Comput. Eng.* **2017**, *7*, 831. [CrossRef]
36. Kumar, V.; Gaur, M.; Kansal, V. Deep feature based forgery detection in video using parallel convolutional neural network: VFID-Net. *Multimed. Tools Appl.* **2022**, *81*, 42223–42240. [CrossRef]
37. Singla, N.; Singh, J.; Nagpal, S.; Tokas, B. HEVC based tampered video database development for forensic investigation. *Multimed. Tools Appl.* **2023**, *82*, 25493–25526. [CrossRef]
38. Video Tampering Dataset Development in Temporal Domain for Video Forgery Authentication. Available online: https://drive.google.com/drive/folders/1y_TVO6-ow2yoKGSLLvj-GAYf_-HtCLw4 (accessed on 14 September 2024).
39. Vtl Video Trace Library. Available online: http://trace.eas.asu.edu/yuv/index.html (accessed on 24 July 2023).
40. Qadir, G.; Yahaya, S.; Ho, A.T. Surrey university library for forensic analysis (SULFA) of video content. In Proceedings of the IET Conference on Image Processing (IPR 2012), London, UK, 1–4 July 2012; p. 121.
41. Quiros, J.V.; Raman, C.; Tan, S.; Gedik, E.; Cabrera-Quiros, L.; Hung, H. REWIND Dataset: Privacy-preserving Speaking Status Segmentation from Multimodal Body Movement Signals in the Wild. *arXiv* **2024**, arXiv:2403.01229.
42. Grip Dataset. Available online: http://www.grip.unina.it/web-download.html (accessed on 3 August 2020).
43. Fadl, S.; Megahed, A.; Han, Q.; Qiong, L. Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix. *Multimed. Tools Appl.* **2020**, *79*, 17619–17643. [CrossRef]
44. Cuevas, C.; Yáñez, E.M.; García, N. Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA. *Comput. Vis. Image Underst.* **2016**, *152*, 103–117. [CrossRef]
45. Sohn, H.; De Neve, W.; Ro, Y.M. Privacy protection in video surveillance systems: Analysis of subband-adaptive scrambling in JPEG XR. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 170–177. [CrossRef]
46. Al-Sanjary, O.I.; Ahmed, A.A.; Sulong, G. Development of a video tampering dataset for forensic investigation. *Forensic Sci. Int.* **2016**, *266*, 565–572. [CrossRef]
47. Ardizzone, E.; Mazzola, G. A tool to support the creation of datasets of tampered videos. In *Proceedings of the Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, 7–11 September 2015*; Proceedings, Part II 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 665–675.
48. CANTATA D4.3 Datasets for CANTATA Project. Available online: https://www.hitech-projects.com/euprojects/cantata/datasets_cantata/dataset.html (accessed on 16 May 2025).
49. Fadl, S.; Han, Q.; Qiong, L. Exposing video inter-frame forgery via histogram of oriented gradients and motion energy image. *Multidimens. Syst. Signal Process.* **2020**, *31*, 1365–1384. [CrossRef]
50. Sharma, H.; Kanwal, N. Video interframe forgery detection: Classification, technique & new dataset. *J. Comput. Secur.* **2021**, *29*, 531–550.
51. Nguyen, X.H.; Hu, Y. VIFFD—A dataset for detecting video inter-frame forgeries. *Mendeley Data* **2020**, *6*, 2020.
52. Fadl, S.; Han, Q.; Li, Q. Surveillance video authentication using universal image quality index of temporal average. In *Proceedings of the Digital Forensics and Watermarking: 17th International Workshop, IWDW 2018, Jeju Island, Korea, October 22–24, 2018*; Proceedings 17; Springer: Berlin/Heidelberg, Germany, 2019; pp. 337–350.
53. NIST Trec Video Retrieval Evaluation. Available online: http://trecvid.nist.gov/ (accessed on 30 September 2022).
54. Sysu-Objforg Dataset. Available online: http://media-sec.szu.edu.cn/sysu-objforg/index.html (accessed on 18 May 2024).
55. Soomro, K. UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild. Available online: https://www.crcv.ucf.edu/data/UCF101.php (accessed on 29 June 2025).
56. Guan, H.; Kozak, M.; Robertson, E.; Lee, Y.; Yates, A.N.; Delgado, A.; Zhou, D.; Kheyrkhah, T.; Smith, J.; Fiscus, J. MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation. Available online: https://www.nist.gov/publications/mfc-datasets-large-scale-benchmark-datasets-media-forensic-challenge-evaluation (accessed on 16 May 2025).
57. Oh, S.; Hoogs, A.; Perera, A.; Cuntoor, N.; Chen, C.-C.; Lee, J.T.; Mukherjee, S.; Aggarwal, J.; Lee, H.; Davis, L. A large-scale benchmark dataset for event recognition in surveillance video. In Proceedings of the Computer Vision and Pattern Recognition (CVPR) 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 3153–3160.
58. Abbasi Aghamaleki, J.; Behrad, A. Malicious inter-frame video tampering detection in MPEG videos using time and spatial domain analysis of quantization effects. *Multimed. Tools Appl.* **2017**, *76*, 20691–20717. [CrossRef]
59. Singh, R.D.; Aggarwal, N. Detection of re-compression, transcoding and frame-deletion for digital video authentication. In Proceedings of the 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), Chandigarh, India, 21–22 December 2015; pp. 1–6.

60. Wang, W.; Jiang, X.; Wang, S.; Wan, M.; Sun, T. Identifying video forgery process using optical flow. In *Proceedings of the Digital-Forensics and Watermarking: 12th International Workshop, IWDW 2013, Auckland, New Zealand, 1–4 October 2013*; Revised Selected Papers 12; Springer: Berlin/Heidelberg, Germany, 2014; pp. 244–257.
61. Wu, Y.; Jiang, X.; Sun, T.; Wang, W. Exposing video inter-frame forgery based on velocity field consistency. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 2674–2678.
62. Bao, Q.; Wang, Y.; Hua, H.; Dong, K.; Lee, F. An anti-forensics video forgery detection method based on noise transfer matrix analysis. *Sensors* **2024**, *24*, 5341. [CrossRef]
63. Prashant, K.J.; Krishnrao, K.P. Frame Shuffling Forgery Detection Method for MPEG-Coded Video. *J. Inst. Eng. Ser. B* **2024**, *105*, 635–645. [CrossRef]
64. Bakas, J.; Naskar, R.; Dixit, R. Detection and localization of inter-frame video forgeries based on inconsistency in correlation distribution between Haralick coded frames. *Multimed. Tools Appl.* **2019**, *78*, 4905–4935. [CrossRef]
65. Kang, X.; Liu, J.; Liu, H.; Wang, Z.J. Forensics and counter anti-forensics of video inter-frame forgery. *Multimed. Tools Appl.* **2016**, *75*, 13833–13853. [CrossRef]
66. Panchal, H.D.; Shah, H.B. Multi-Level Passive Video Forgery Detection based on Temporal Information and Structural Similarity Index. In Proceedings of the 2023 Seventh International Conference on Image Information Processing (ICIIP), Solan, India, 22–24 November 2023; pp. 137–144.
67. Singh, G.; Singh, K. Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation. *Multimed. Tools Appl.* **2019**, *78*, 11527–11562. [CrossRef]
68. Pandey, R.; Kushwaha, A.K.S. A Novel Histogram-Based Approach for Video Forgery Detection. In Proceedings of the 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 28–30 August 2024; pp. 827–830.
69. Ren, H.; Atwa, W.; Zhang, H.; Muhammad, S.; Emam, M. Frame duplication forgery detection and localization algorithm based on the improved Levenshtein distance. *Sci. Program.* **2021**, *2021*, 5595850. [CrossRef]
70. Mohiuddin, S.; Malakar, S.; Sarkar, R. An ensemble approach to detect copy-move forgery in videos. *Multimed. Tools Appl.* **2023**, *82*, 24269–24288. [CrossRef]
71. Shehnaz; Kaur, M. Detection and localization of multiple inter-frame forgeries in digital videos. *Multimed. Tools Appl.* **2024**, *83*, 71973–72005. [CrossRef]
72. Wang, Y.; Hu, Y.; Liew, A.W.-C.; Li, C.-T. ENF based video forgery detection algorithm. *Int. J. Digit. Crime Forensics (IJDCF)* **2020**, *12*, 131–156. [CrossRef]
73. Kharat, J.; Chougule, S. A passive blind forgery detection technique to identify frame duplication attack. *Multimed. Tools Appl.* **2020**, *79*, 8107–8123. [CrossRef]
74. Huang, C.C.; Lee, C.E.; Thing, V.L. A novel video forgery detection model based on triangular polarity feature classification. *Int. J. Digit. Crime Forensics (IJDCF)* **2020**, *12*, 14–34. [CrossRef]
75. Yu, L.; Wang, H.; Han, Q.; Niu, X.; Yiu, S.-M.; Fang, J.; Wang, Z. Exposing frame deletion by detecting abrupt changes in video streams. *Neurocomputing* **2016**, *205*, 84–91. [CrossRef]
76. Zhang, Z.; Hou, J.; Ma, Q.; Li, Z. Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames. *Secur. Commun. Netw.* **2015**, *8*, 311–320. [CrossRef]
77. Jiang, G.; Du, B.; Fang, S.; Yu, M.; Shao, F.; Peng, Z.; Chen, F. Fast inter-frame prediction in multi-view video coding based on perceptual distortion threshold model. *Signal Process. Image Commun.* **2019**, *70*, 199–209. [CrossRef]
78. Fadl, S.; Han, Q.; Li, Q. CNN spatiotemporal features and fusion for surveillance video forgery detection. *Signal Process. Image Commun.* **2021**, *90*, 116066. [CrossRef]
79. Mohsen, H.; Ghali, N.I.; Khedr, A. Offline signature verification using deep learning method. *Int. J. Theor. Appl. Res.* **2023**, *2*, 225–233.
80. Koshy, L.; Shyry, S.P. Detection of tampered real time videos using deep neural networks. *Neural Comput. Appl.* **2024**, *37*, 7691–7703. [CrossRef]
81. Nguyen, X.H.; Hu, Y.; Amin, M.A.; Khan, G.H.; Truong, D.-T. Detecting video inter-frame forgeries based on convolutional neural network model. *Int. J. Image Graph. Signal Process.* **2020**, *10*, 1. [CrossRef]
82. Kohli, A.; Gupta, A.; Singhal, D. CNN based localisation of forged region in object-based forgery for HD videos. *IET Image Process.* **2020**, *14*, 947–958. [CrossRef]
83. Patel, J.; Sheth, R. An optimized convolution neural network based inter-frame forgery detection model-a multi-feature extraction framework. *ICTACT J. Image Video Process* **2021**, *12*, 2570–2581. [CrossRef]
84. Chandru, R.; Priscilla, R. Video Integrity Detection with Deep Learning. In Proceedings of the 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 14–15 March 2024; pp. 1288–1292.

85. Singla, N.; Nagpal, S.; Singh, J. Frame Duplication Detection Using CNN-Based Features with PCA and Agglomerative Clustering. In *Communication and Intelligent Systems: Proceedings of ICCIS 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 383–391.

86. Hu, Z.; Duan, Q.; Zhang, P.; Tao, H. An Attention-Erasing Stripe Pyramid Network for Face Forgery Detection. *Signal Image Video Process.* **2023**, *17*, 4123–4131. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.