Hindawi
Applied Computational Intelligence and Soft Computing
Volume 2023, Article ID 6661192, 21 pages
https://doi.org/10.1155/2023/6661192

Review Article

Forged Video Detection Using Deep Learning: A SLR

Maryam Munawar,[1] **Iram Noreen**,[1] **Raed S. Alharthi**,[2] **and Nadeem Sarwar**[1]

[1]*Department of Computer Science, Bahria University Lahore Campus, Lahore 54782, Pakistan*
[2]*Department of Computer Science and Engineering, University of Hafr Al-Batin, Hafar Al-Batin 39524, Saudi Arabia*

Correspondence should be addressed to Nadeem Sarwar; nsarwar.bulc@bahria.edu.pk

Received 15 August 2023; Revised 4 October 2023; Accepted 12 October 2023; Published 25 October 2023

Academic Editor: Mominul Ahsan

In today's digital landscape, video and image data have emerged as pivotal and widely adopted means of communication. They serve not only as a ubiquitous mode of conveying information but also as indispensable evidential and substantiating elements across diverse domains, encompassing law enforcement, forensic investigations, media, and numerous others. This study employs a systematic literature review (SLR) methodology to meticulously investigate the existing body of knowledge. An exhaustive review and analysis of precisely 90 primary research studies were conducted, unveiling a range of research methodologies instrumental in detecting forged videos. The study's findings shed light on several research methodologies integral to the detection of forged videos, including deep neural networks, convolutional neural networks, Deepfake analysis, watermarking networks, and clustering, amongst others. This array of techniques highlights the field and emphasizes the need to combat the evolving challenges posed by forged video content. The study shows that videos are susceptible to an array of manipulations, with key issues including frame insertion, deletion, and duplication due to their dynamic nature. The main limitations of the domain are copy-move forgery, object-based forgery, and frame-based forgery. This study serves as a comprehensive repository of the latest advancements and techniques, structured, and summarized to benefit researchers and practitioners in the field. It elucidates the complex challenges inherent to video forensics.

1. Introduction

In the last few years, image production has increased exponentially. Around 1.4 trillion digital photos are expected to be created in 2020 alone, according to an estimate. Today, digital photographs are essential to our everyday lives since they not only serve as a means of saving photos of family and friends but also appear on the covers of all the main news publications, such as magazines, newspapers, and journals. Thanks to recent technological advancements, one may now effortlessly modify a digital picture or video using computer software or a mobile application. Identification theft is one instance in which someone's identity can be taken by a fraudster who has access to their personal and financial data. Law enforcement officials must utilize several automatic tools or approaches to determine if a person is clean-handed or the perpetrator in order to prevent such dire circumstances [1].

When authentic digital data are in short supply, the main objective of synthetic data generation is to produce something that is extremely close to the actual thing. Deepfake technology, which uses computer vision and graphics to swap out one person's face for another person's, is a major source of worry [2]. The credibility of media sources is therefore seriously compromised. Therefore, one must check the video to determine whether the content is unique. If the video's authenticity and uniqueness are affected, viewers may perceive it differently [3]. In recent years, there has been increased interest in object-based video forgery detection. However, up until recently, the most common object-based forgery detectors still relied on characteristics that had to be handcrafted, and their results were subpar [4]. Videos are vulnerable to manipulation attempts that change the intended meaning and trick the viewer. Previous methods of detecting video falsification discovered the altered areas using minute hints. Attackers can, however, circumvent

detection by erasing these hints using video compression or blurring [5].

With the advancement of multimedia editing capabilities in recent years, image and video alteration has become increasingly popular [6]. Current face forgery detection techniques based on frequency domain discover that, in contrast to authentic photographs, the generative adversarial network (GAN) fabricated images exhibit glaringly visible grid-like visual abnormalities in the frequency spectrum [7]. The most significant and commonly utilized type of communication nowadays is video and picture data. In a variety of fields, including law enforcement, forensic research, media, and others, it is utilized as proof and verified evidence. The issue of video and picture counterfeiting has emerged along with the growth of video applications and data [8].

Nowadays, especially considering the ubiquitous sharing of movies on social media and websites, a lot of attention is dedicated to spotting video forgeries. There are several video editing apps that work well for altering video footage or even producing videos [9]. The full exploration of deep learning models' broad representational capacity and their connection to different forensic aspects remains incomplete. Instead, existing approaches mostly concentrate on manually picked models and features for a limited task, such as copy-move or splicing [10]. The videos captured by surveillance cameras are frequently used in court as persuasive evidence. They are frequently employed to offer protection and security. Sometimes, after editing, various postprocessing steps are carried out to conceal the signs of counterfeiting. Since then, it has been critical to scientifically assess the veracity and integrity of surveillance videos.

1.1. Importance and Contribution. Several survey papers exist in the literature to cover different aspects of forged videos. Nayerifard et al. [11] conducted a literature review on traditional machine learning for image forensics, covering the time span from 2010 to 2021. Similarly, Stroebel et al. [12] conducted a systematic literature review (SLR) from 2021 to August 2022, highlighting the dominance of deep learning (DL) over machine learning (ML) in Deepfake detection, especially in nonmedical contexts. Another SLR conducted by Chauhan et al. [13] highlights the use of various algorithms, including deep neural networks (DNN), for detecting Deepfakes, particularly in the video game and cinema industries, and its focus was to identify loopholes, datasets, and contemporary techniques of Deepfake in the entertainment domain. Rana et al. [14] conducted an SLR on Deepfake detection covering the years from 2018 to 2020, categorizing methods into deep learning-based, classical machine learning-based, statistical, and blockchain-based techniques, concluding that deep learning-based methods are the most effective in detecting Deepfakes. Shahzad et al. [15] emphasized the need for novel methods to detect face manipulation in Deepfakes and suggested combatting this threat through policies, regulations, and technological advancements. Tolosana et al. [16] presented a comprehensive survey on techniques for facial image manipulations only and discussed methods for detecting manipulations related to Deepfakes. Yadav et al. [17] presented a survey on forgery and described Deepfake as an emerging AI-based technology to create convincing fake videos. They highlight its potential for misuse, such as character defamation of politicians and celebrities.

This study emphasizes the importance of video and image data in today's digital world in various domains such as law enforcement and forensic investigations. This systematic literature review (SLR) covers the primary research studies from 2016 to 2023, highlighting various research methodologies like deep neural networks, watermarking networks, hybrid model, etc. for detecting forged videos. Moreover, this study presents a discussion on key challenges like frame manipulation and serves as a valuable resource for researchers and practitioners in the field of video forensics, addressing the complexities and challenges involved. In this study, the primary objective is to collect information on fake video and image data and to collect information on techniques to identify the forged videos. The above summary concludes that no other SLR exists with the same publishing and scope period. The main contribution of this paper is as follows:

This study

(i) provides an overview of the evolution of video and image forgery detection

(ii) widely covers not only the traditional machine learning methods but also emerging methods such as deep learning, transfer learning, federated learning, water marking, generative adversarial networks (GANs) and attention mechanisms in improving video forensics

(iii) provides a state-of-the-art summary of notable work in the forensic domain for detection and verification of video authenticity

(iv) covers methodologies evaluated on different categories of private and public datasets such as surveillance, security, action, social media data, legal proceedings, wildlife, action detections, privacy or consent issues, and video forensics

(v) explores potential future trends and challenges of detection techniques for research community in video and image manipulation

The rest of the paper is organized as follows: The next section describes the planning, design, and execution of the SLR. Further sections present the state-of-the-art summary of the prominent work in the field, followed by a detailed discussion of various techniques. The last section describes domain challenges, followed by a conclusion in the end.

2. Systematic Literature Review

The three steps of SLR are planning, implementation, and reporting. The best research information is acquired and then utilized to evaluate research challenges, according to evidence-based software engineering (EBSE). The identified studies were examined based on their title, abstract, and

conclusions. The planning of SLR is presented in Figure 1. The next subsections are specifics on how the SLR is expected to be carried out.

This systematic literature review (SLR) paper stands out in the field of forged video detection by offering a comprehensive and up-to-date perspective. Covering primary research studies from 2016 to 2023, it ensures readers are presented with the latest advancements and techniques in video forensics. Unlike studies that focus on specific methodologies and domains like Deepfake, interframe, copy-move, and object-based detection, this review encompasses a wide array of research approaches, including deep neural networks, convolutional neural network, and hybrid models, providing a holistic view of detection techniques. Emphasizing the critical challenges in video forensics, particularly frame manipulation, Deepfake, copy-move manipulation, and object-based manipulation issues, it offers practical insights and recommendations for researchers and practitioners. Furthermore, this review takes a global perspective, considering the impact of forged videos across various domains, making it a valuable and distinctive resource for addressing this pervasive issue.

2.1. Research Questions Formation. The research questions (RQ) addressed in this study is presented in Table 1.

2.2. Review Protocol Formation. This section outlines the SLR process and provides a quick overview of the SLR. The following subsections describe the search process, selection of papers, extraction of data, and analysis conducted.

2.2.1. Process of Search. Following that, we chose the keywords based on the language used in the forged video detecting sector. Then, a specialist looked up all the keywords' synonyms, alternatives, and hypernyms. To limit the search results, used the Boolean operators AND and OR, as well as the wildcard character "*" in the search term. The synonyms were combined using the "OR" operator. For instance, the wildcard (*) denotes either a single alphanumeric character or a collection of alphanumeric characters in accordance with the IEEE Xplore search criteria. The population and intervention terms are combined using the AND operator.

To identify which research should be regarded as primary, both primary and secondary searches, as well as snowball tracking, were employed. I looked for the years 2016 and 2023 but found nothing. The primary search was conducted by searching electronic journals, conference proceedings, and the grey literature. Search engines and online databases for research (IEEE, BMJ, Springer, ISPRS, SAGE, ICCIT, IAES, ECS, BMC, ELSEVIER, etc.) were also used. Although they were also used, search engines like Google and Google Scholar were excluded from the total number of studies since they only included excerpts from other known research databases.

The secondary search on the papers uses the titles, abstracts, and conclusions (identified by the first search). The

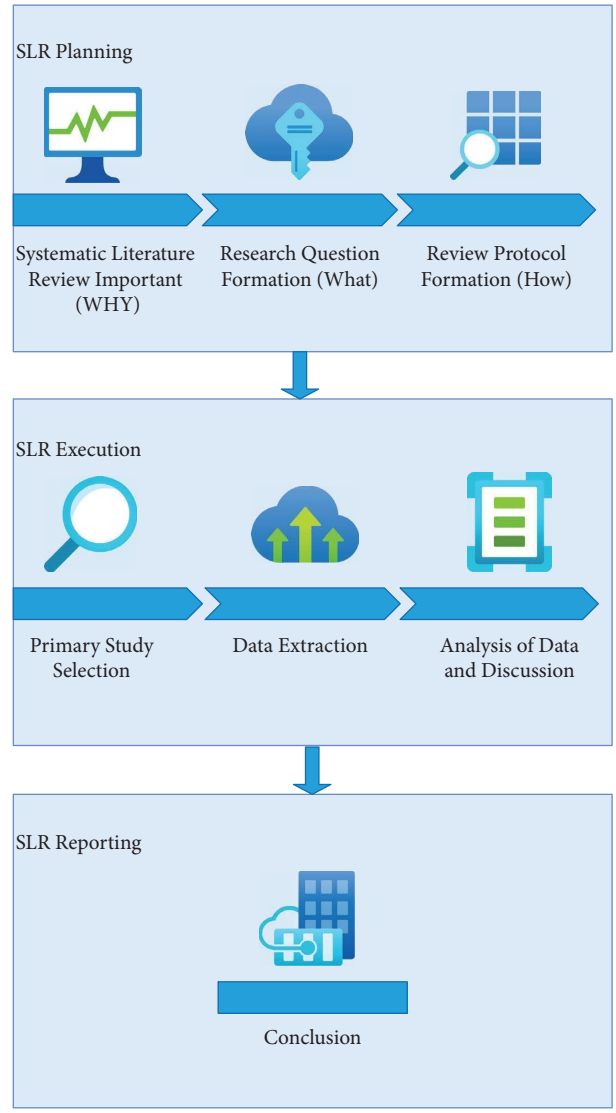

Figure 1: Systematic literature review planning.

secondary search results were then used to select the articles for analysis using the inclusion/exclusion criteria and quality criteria provided in the next section.

Snowball tracking was also carried out to make sure no pertinent studies were missed, which entails reading over the final primary research's reference list.

2.2.2. Exclusion Criteria of Study. Primary study is in English and with the complete text was necessary for consideration.

2.2.3. Inclusion Criteria of Study. The research's relevance to RQs was one of the selection criteria. Primary studies from business and/or research viewpoints were considered if they offered an empirical evaluation. The validity of SLR significantly depends on the calibre of the chosen study. Therefore, we only included peer-reviewed studies that met our standards for quality.

Table 1: Research questions.

ID	Research question	Motivation
1	In the literature, which technologies, models, methods, and practices are discussed?	To get a deeper grasp of the motivation behind identifying forged videos as well as the most recent models, procedures, and techniques for forged video identification and localization
2	What mitigation techniques are advised to be used to recognize forged videos?	Identification of selection processes and barriers for forged videos is essential for deployment
3	What variables determine the forged videos are identified successfully using different technique?	To establish procedures, guidelines, standards, and collective experiences for effective forged video detection deployment, as well as likely cause and risk factors
4	When implementing forged video detection, what techniques/models/methods are developed for metric selection?	Examine how certain strategies may be utilized to alleviate the problems associated with identifying faked video faces

2.2.4. Criteria of Quality Assessment. SLR must be used with high-quality research in order to produce reliable results and conclusions. This calls for sound SLR planning, appropriate keywords, and well-stated exclusion and inclusion criteria. A snowball tracking activity that involved perusing the reference lists of each primary study listed (see Table 2) was then performed. Criteria to further analyse the validity of the research are presented in Table 2.

2.2.5. Extraction of Data. Each RQ could be extracted in a structured, uniform, and consistent manner thanks to the data extraction forms made in Microsoft Excel. The results were entered in the forms for further analysis and investigation. You may find definitions of certain RQ-related data (see Table 3) lower down this page.

2.2.6. Empirical Study. Identified the empirical research method that was applied in each primary study. Empiricism techniques are categorized (see Table 4).

2.2.7. Tool/Model. The measurement planning model aids software businesses in carrying out their measurement operations in a manner that helps them accomplish their objectives. We reviewed the key studies, together with their empirical validation, to identify what kinds of measurement planning models and related tools are available.

2.3. Method of Conducting SLR

2.3.1. Primary Study of Research. In Figure 2, the illustration outlines the systematic process of identifying and selecting primary studies. The initial step involved conducting a comprehensive primary search, which yielded a pool of 450 prominent papers to serve as our initial reference point. Subsequently, further exploration led to the discovery of several potential primary studies, as indicated. In addition, Table 5 complements this process by providing valuable information regarding the impact factor and the most recent updates of these studies.

2.3.2. Data Analysis and Extraction. Data extraction forms are used to extract data (see data extraction section). In our analysis, we used both qualitative and quantitative methods. Figure 3 displays the distribution of primary research based on their publication year and the number of research publications per year. The average number of publications was the same every three to four years. These findings are rather surprising given that fake videos have been available for a long time. In past years, it was expected to see more empirical research. This might be due to the paucity of researchers in this subject and the lack of access to businesses that utilize falsified video identification. As a result, the group may be unable to exchange experiences and learn from one another. Figure 4 depicts the empirical technique for primary research categorized as conference and journal.

Table 5 shows the total number of articles, conferences, and search engines for dataset searches. Table 6 represents the state-of-the-art of all previous 7 years publications, whereas Table 7 represents the dataset detail. The pie chart in Figure 5 shows the bar chart of publisher from research that has been done, and Figure 2 represents the total ratio and proportion of research articles used in this study. Figure 6 represents the primary study selection process and criteria.

Table 7 gives a thorough overview of the many datasets used in computer vision and video analysis, spanning multiple years, and research references. These datasets cover a wide range of video formats, including MP4, MPEG-4, and high-resolution video coding standards, and range greatly in size from thousands to millions of samples. In addition, these dataset's dimensions, which indicate the qualities of the video data, range greatly from low resolutions to multidimensional feature vectors. From action recognition and Deepfake detection to video frame analysis and high-provenance image and video datasets, the datasets span a wide spectrum of applications. These datasets can be used by computer vision researchers and practitioners to create and test new algorithms, ultimately improving video analysis and artificial intelligence in a variety of real-world applications. These datasets support developments in the study of artificial intelligence and video-based research by catering to a wide range of applications, including Deepfake detection, action identification, and picture and video analysis.

3. Categories of Models

This section provides a discussion on models used in the domain of video forgery detection. The structure of models and prominent work based on these models is discussed.

3.1. Convolution Neural Network (CNN). An advanced deep learning architecture known as a convolutional neural network (CNN) is specifically created to process and analyse visual input, such as pictures and videos. Due to its capability to automatically learn hierarchical features from the input data, it has revolutionized computer vision jobs. Figure 7 shows the CNN structure.

3.1.1. Structure of CNN. The main components of a CNN are described as follows.

The foundational elements of the CNN are convolutional layers. Each layer is made up of a collection of filters (also known as kernels) that conduct convolutional operations by sliding over the input data, like an image. The filters are in charge of spotting various elements in the input, such as edges, corners, and textures.

An activation function, frequently a ReLU (rectified linear unit), is applied elementwise after each convolution process to provide nonlinearity to the network. This aids the CNN model in detecting more intricate links and patterns in the data.

Table 2: Selection for quality criteria of primary study.

Type	Definition
Internal	Study's purpose, assumptions, and background are presented
External	The findings should be relevant in industry as well as academia
Construct	The relationship between research questions, evaluation/measurement components, and outcomes is well characterized
Conclusion	The conclusions and analysis are based on theoretical/empirical research and correspond to the RQs

Table 3: Data extraction.

Purpose	Meta-data
Generic information	Title of the paper, author(s), location of publication, and date of publication
Specific information	There is coverage of empirical research, measurement planning tool/model, measurement purpose, implementation level, measurement entity focus, metrics selection methodologies, failure/success factors, and challenge mitigation measures

Table 4: Methods of empirical research.

Type of study	Explanation
Case study	Study that is planned and carried out with particular goals in order to evaluate a research topic
Industrial report	Study using empirical evaluations based on industry experiences but no explicit study objectives or questions
Experiment	An experiment that is planned and carried out according to the standard operating procedures stated in the guidelines
Action research	The research technique, as well as the study that is designed and carried out to address a research topic
Survey	The study was planned and carried out in order to collect evidence utilising quantitative and qualitative research approaches
Not defined	If the research technique is not provided implicitly and is unable to define it, the study is labelled as "not-defined."

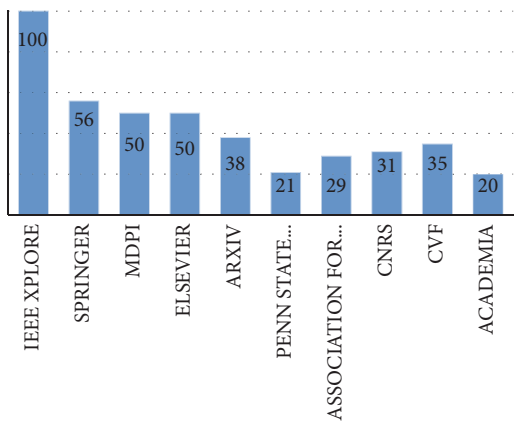

Figure 2: Percentage of research paper.

Pooling layers are used to control overfitting and minimize the spatial dimensions of the data. The most popular pooling method, known as max-pooling, subsamples the maximum value from a tiny area of the feature map. This helps to keep key characteristics while reducing computational complexity.

The data are sent via fully connected layers after several convolutional and pooling layers. These layers link each neuron in one layer to each neuron in the one above. They are responsible of making predictions using the high-level characteristics that the preceding layers have learnt.

Each convolutional block in the CNN architecture is made up of convolutional layers, followed by activation and pooling layers. For classification tasks, a softmax layer is frequently placed after the final fully linked layers and offers probability ratings for various classes.

3.1.2. Discussion of CNN Based Approaches. Ganguly et al. [2] proposed a deep learning model improved with the visual attention approach to distinguish false videos and images from authentic ones. To create the feature maps, they first extracted the facial region from the video frames and then applied the extracted region through the Xception model that has been previously trained. Next, they concentrated mostly on the Deepfake video modification of the remaining artifacts with the aid of the visual attention mechanism. FaceForensics++ and Celeb-DF, two publicly accessible datasets, were used to test their model (V2). Kumar et al. [3] described that the deep characteristics are an important aspect in identifying the fake and abnormal fluctuations in the film. They used a parallel "CNN" model to extract deep features to uncover the disassociation between the consecutive frames and detect video counterfeiting. Their model also determined that how far the correlation coefficient is

from the deep features. Tan et al. [4] combined two-dimensional/three-dimensional recurrent neural network and the convolutional neural network for the first time in a unique hybrid deep learning network. They used it for object-based video forgery detection with sophisticated encryption formats. Zhou et al. [5] offered a network for video watermarking to detect manipulation. They trained a decoder to forecasts the tampering mask and a 3D-UNet-based watermark embedding network.

Fadl et al. [32] presented an interframe forgeries (frame deletion, insertion, and duplication) detection system using a "2D convolution neural network" (2D-CNN) with spatiotemporal fusion and information for deep automated feature extraction. For classification, a "Gaussian RBF multiclass support vector machine" (RBF-MSVM) is employed. Zheng et al. [33] offer a brand-new end-to-end structure with two key phases. A completely temporal convolution network makes up the initial level "fully temporal convolutional network" (FTCN). Surprisingly, discover that this unique architecture can help the model extract temporal cues and increase its capacity to generalize. Temporal transformer network, which is used in the second step, seeks to investigate long-term temporal coherence. The suggested system is all-encompassing and adaptable, allowing for direct training from the start without the need of pretraining models or outside datasets.

Hau Nguyen et al. [41] are retraining the existing CNN models that were trained on the ImageNet dataset in order to identify video interframe forgeries. The suggested techniques are based on retrained CNN models that take use of spatial-temporal correlations in a video to effectively identify interframe forgeries. It is suggested to use a confidence score rather than the raw output score from networks to account for network mistakes. It has been demonstrated through the results of tests that the suggested strategy is both much more efficient and accurate than more current methods. Long et al. [75] offer a novel method for forensic analysis that relies on the local spatiotemporal correlations inside a video segment to identify frame deletions. Suggest modifying the "Convolutional 3D Neural Network" (C3D) for the detection of frame drops. Rao and Ni [86] automated the learning of hierarchical representations from the input RGB colour photographs by a "convolutional neural network" (CNN), a novel method for detecting image forgeries based on deep learning. The suggested CNN is made particularly for applications like copy-move detection and picture splicing.

3.2. Deep Artificial Neural Network (ANN)

3.2.1. Structure of Deep ANN. An artificial neural network (ANN) with numerous layers in between the input and output layers is known as a deep neural network (DNN) (see Figure 8). Since it has several hidden layers, it can learn and model complicated patterns and representations in the data, which is why it is dubbed "deep" learning. DNNs are a crucial part of deep learning, a branch of machine learning that has attracted substantial interest and achieved success in several fields, including speech recognition, natural language processing, and computer vision. Deep neural networks have shown to be incredibly effective at a variety of difficult tasks, including picture and audio recognition, interpreting spoken language, playing games, and many more activities involving a lot of data and high-dimensional input.

3.2.2. Discussion of Deep ANN Based Methods. Kaur et al. [37] centred on a very effective strategy for the usage of "deep convolutional neural network" (DCNN) to reveal interframe manipulation in the videos. The suggested approach will identify forgeries without the need for extra preembedded frame data. The classification of the fabricated frames by our algorithm is based on the correlation between frames and the detected irregularities using DCNN, which is another important aspect of preexisting learning techniques. The decoders used for batch input normalization speed up training. Zhong et al. [38] presented a method to efficiently extract from the video multidimensional dense moment features. Second, a brand-new way of feature representation concatenates each feature submap index, which represents each dimension of the feature, into a 9-digit dense moment feature index. Third, an interframe best match approach is suggested to locate the best matches among each pixel's 9-digit dense moment feature index. The best match map is created by all the greatest matches. D'Avino et al. [70] propose deep learning detection with an architecture based on autoencoders and recurrent neural networks. The autoencoder learns an intrinsic representation of the source during a training phase on a few clean frames. The forged material is then identified as anomalous since it does not conform to the taught model and is encoded with a substantial reconstruction error. To leverage temporal relationships, recursive networks with the long short-term memory model are utilized. Preliminary findings on forged videos demonstrate the approach's potential.

Yadav and Salmani [17] present a survey article showing that the Deepfake approach combined with a generative adversarial network can produce results that appear realistic to human eyes. A false picture is created by the "Generative Adversarial Network" by fusing together two separate photographs, but the images of Persons A and B must be comparable in terms of facial features and skin tone, and they must have been shot under the same lighting conditions. Deepfake can be used in two beneficial ways: it can be implemented in the education sector to change the faces of historical figures and use them as study materials; or it can be used in the arts to change the faces of actors in movies, which will save a lot of money by doing away with the need for CGI and VFX in addition to Deepfake. Hosler et al. [44] enumerate the features, make-up, and collecting process of video-ACID, which contains films with obvious markings for testing camera model recognition software. Finally, utilising cutting-edge deep learning algorithms, we present baseline camera model identification findings on these evaluation films. Shou et al. [61] explicitly address the difficulties in training ODAS models by suggesting three unique techniques. Three techniques are used to deal with the lack of training data. Carry out substantial research utilising ActivityNet and THUMOS'14.

Table 5: Potential of primary journals and conferences.

Publish journals and conferences	Impact factor	Paper count
Journals database		
Pattern analysis and application	3.9 (2022)	15
Multimedia tools and application	2.5 (2022)	35
Signal processing: image communication	3.4 (2022)	30
Cornell hospitality quarterly	3.7 (2023)	26
Symmetry	2.7 (2022)	24
The visual computer	2.8 (2022)	35
Intelligent automation and soft computing	3.4 (2023)	32
Journal of imaging	3.2 (2022)	15
Signal processing	4.7 (2023)	11
Wireless personal communications	2.0 (2023)	10
Information science	8.2 (2021)	10
IEEE transactions on circuits and systems for video technology	8.4 (2022)	12
American journal of computer sciences and applications	0.3 (2022)	10
Iran journal of computer science	1.9 (2022)	10
International journal of information technology	2.5 (2021)	8
International workshop on information forensics and security	2.4 (2021)	9
Computer vision foundation	8.3 (2023)	11
International journal advanced research computer and communication engineering	8.1 (2023)	10
Multimedia system	2.6 (2023)	10
Transactions on circuits and systems for video technology	8.4 (2022)	12
Society for imaging science and technology	0.5 (2023)	8
Information hiding and multimedia security	0.2 (2023)	8
Turkish journal of electronic engineering and computer sciences transaction on multimedia	0.8 (2023)	7
International journal of electronic security and digital forensics	0.4 (2023)	9
Science direct	10.7 (2023)	7
Southwest symposium on image analysis and interpretation	1.0 (2020)	5
Information science	8.2 (2021)	10
Conference database		
CVF international conference on computer vision (ICCV)	8.3 (2022)	18
International conference on computer communication	2.2 (2023)	6
International conference of intelligent computing and control system	0.14 (2020)	4
International symposium on information technology convergence	0.7 (2022)	3
Computer vision and pattern recognition	23.46 (2020)	15
International multiconference on systems, signals and devices	0.4 (2023)	5

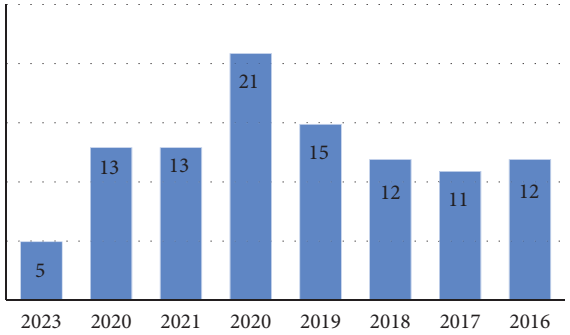

Figure 3: Yearly distribution of literature.

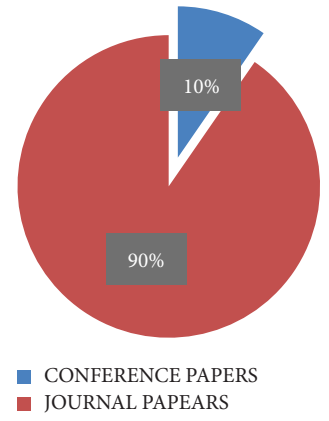

Figure 4: Methods used in primary study.

3.3. Hybrid Neural Network Model

3.3.1. Structure of Hybrid Models. When parts of several neural network architectures or machine learning models are combined, the result is a hybrid neural network model (see Figure 9). The objective is to combine the advantages of each model to produce a more robust and adaptable system that can successfully complete a range of tasks. To process both image and sequence input concurrently, a hybrid neural network, for instance, may incorporate

Table 6: Comparison of State of art approaches.

Sr. no	Reference	Approach	Contribution	Dataset	Model accuracy	Limitation
1	Wang et al. [18]	Convolutional neural network	Spatiotemporal model (3D ConvNet). Novel training strategy (AltFreezing)	FaceForensics++ [19], Celeb-DF (V2) [20]	99%	Data augmentations/ Model evaluation
2	Liu et al. [21]	Neural network	Generalized residual federated learning method	FaceForensics++ [19], some YouTube data	99.7%	Privacy protection/Data privacy
3	Tyagi and Yadav [1]	Survey	Deep learning, visual imagery forgery detection	Self-collected data	NA	Generalized methods
4	Ganguly et al. [2]	Deep learning model	Soft attention mechanism, visual attention	FaceForensics++ [19], Celeb-DF (V2) [20]	70.1%	Low accuracy
5	Kumar et al. [3]	Convolutional neural network	Extract deep features, distance of correlation coefficient	VIFFD [22], surrey university library for forensic analysis (SULFA) [23]	86.5% and 92% for video level/99.9% frame level	Limited mention of false positives/information about methodology
6	Tan et al. [4]	2D-convolutional neural network	Bidirectional long-short-term-memory	SYSU-OBJFORG [24]	99%	No comparison with emerging deep learning architectures
7	Zhou et al. [5]	Watermarking network	Robust watermarking network for video forgery detection (RWVFD) tampering localization, 3d-unet-based watermarking embedding network	Davis [25], YouTube-VOS	NA	Limited scope of video types
8	Kim et al. [6]	Convolutional neural network	Symmetrically overlapped motion residual	SULFA 14, REWIND18, SYSU-OBJFORG 15	98%	Diverse tampering should be considered
9	Wang et al. [7]	Discrete cosine transform-based forgery clue augmentation network (FCAN-DCT)	Compact features extraction (CFE), frequency temporal attention (FTA)	Wild-Deepfake [26], Celeb-DF [20]	86%, 99%	Lack of current real-world exploration
10	Munawar and Noreen [8]	Siamese-based RNN, I3D (inflated 3 dimension)	Siamese based RNN integrated with I3D to find the duplicate frame rate	Media forensic challenge (MFC) [27], video and image retrieval and analysis tool (VIRAT) [28]	93.3%, 86.6%	Transfer learning not explored
11	Alsakar et al. [9]	SVD (single value decomposition), inter-frame forgery	First phase is 3D-Tensor decomposition, second phase is forgery detection, third phase is forgery locating	Randomly selected eight videos	99%	Enhance detection and location for variety
12	Jin et al. [10]	ResNet50 model, LSTM-EnDec, DMAC, noiseprint	Object based video forgery detection, multi features fusion, dual stream	GRIP [29], VTD (video tampering dataset) [30], SULFA [23], REWIND [31]	NA	Limited evaluation of real-world scenarios
13	Fadl et al. [32]	2D-CNN, SSIM, gaussion RBF multiclass support vector machine (RBF-MSVM)	Passive forensics, CNN (convolutional neural network), SSIM, spatiotemporal features, inter-frame forgeries	SULFA [23]	99.9%	Detecting multiple forgeries in videos
14	Zheng et al. [33]	Spatiotemporal convolutional,	2D R50 network structure, 3D R50 network structure, 3D R50-FTCN (fully temporal convolutional network)	Deepfake [34], FaceSwap, Face2Face	99%	Limited real-world application evaluation
15	Huang et al. [35]	Cross-model authentication	Localization on live surveillance videos	Run time evaluation	95.1%	Hardware and environment scalability

Table 6: Continued.

Sr. no	Reference	Approach	Contribution	Dataset	Model accuracy	Limitation
16	Verde et al. [36]	Convolutional neural network (CNN)	Focal: Forgery localization framework based on video coding self-consistency	60 encoded videos	88.9%	Assess scalability, improve model fusion
17	Kaur and Jindal [37]	Deep convolutional neural network (DCNN)	ANN (artificial neural network), convolutional layer, ReLU activation layer, max pool layer, correlation classification	REWIND [31], GRIP [29]	98%	Consideration of hardware constraints
18	Zhong et al. [38]	Interframe best match algorithm	A unified moment framework, 9-digit dense, moment feature index, best match algorithm	REWIND [31], SULFA [23]	75%	Real-world scenario evaluation
19	Sasikumar et al. [39]	SIFT, MSCL, clustering	SIFT (scalar invariant features transformer), MSCL (mean shift clustering algorithm), camera motion, feature extraction, classification, segmentation, in-painting	Randomly collected data	NA	Enhance video duplicate detection security
20	Aloraini et al. [40]	Sequential and patch analysis	Patch analysis, sequential analysis, object removal video forgery, spatiotemporal analysis	SULFA [23], SYSU-OBJFORG [24]	72%	Nonadditive models are not explored
21	Hau Nguyen et al. [41]	Convolutional neural network (CNN)	Video interframe forgery detection, video authenticity, passive forensic	VFDD [42]	99%	CNN needs to be simplified for diverse forgery
22	Parveen et al. [43]	Clustering algorithm	*K*-means clustering, radix sort	Randomly collected data	NA	Limited focus on clustering algorithms
23	Hosler et al. [44]	Convolutional neural network (CNN)	Benchmark testing, video signal processing	ACID [45]	95%	Algorithm benchmark evaluations required
24	Fayyaz et al. [46]	Sensor pattern noise	Video forensics, digital forgery, sensor pattern noise, photo response nonuniformity noise (PRNU)	Dresden [47]	Not mention	Vulnerability to induced SPN attacks
25	Joshi and Jain [48]	Video tempering detection	Temporal fingerprints, optical flow	200 video clip	87.5%	Implement machine learning for classification
26	Chen et al. [49]	Scale-invariant feature transform	Invariant moment, region growing	Copy-move forgery detection (CoMoFoD) [50]	84.6%	Reduce keypoints, optimize region growing
27	Pavlović et al. [51]	Multifractal spectrum and statistic parameters	New metaheuristic and supervised learning method	CoMoFoD [50]	96%	Explore metaheuristics and multifractals further
28	Liu et al. [52]	Scale-invariant feature transform	K-means clustering	Randomly collected data	89%	Optimize parameters and explore new technologies
29	Yadav and Salmani [17]	Survey	Machine learning, deep learning, generative adversarial network, neural network	Self-collected data	NA	Limited theoretical explanation
30	Jia et al. [53]	Optical flow consistency	Coarse-to-fine detection, video passive forensic	Randomly collected data	Not mention	Enhance handling of static scenes
31	Singh and Singh [54]	Dual-clutch transmission (DCT) matrix	Region duplication, correlation coefficient, and coefficient of variation	Randomly collected data	96.6%	Struggles with subtle intensity changes
32	Afchar et al. [55]	Deep learning approach	DeepFake, Face2Face	DeepFake [34]	98% DeepFake 95% Face2Face	Limited theoretical explanation of results

Table 6: Continued.

Sr. no	Reference	Approach	Contribution	Dataset	Model accuracy	Limitation
33	Chen et al. [56]	Region based convolutional neural network	Region proposal network in faster R-CNN network	Cityscapes [57], KITTI [58], SIM10K	NA	Dependence on adversarial training techniques
34	Aneja et al. [59]	Convolutional neural network (CNN)	Recurrent neural network (RNN) powered by long-short-term-memory (LSTM)	MS COCO [60]	NA	Sequential limitations in LSTM models
35	Shou et al. [61]	Online detection of action start (ODAS)	Generative adversarial network, evaluation protocol	THUMOS'14 [62], activity net	NA	Limited practical application and evaluation
36	Nguyen et al. [63]	Convolutional neural network	Capsule network, face swap detection, facial reenactment detection	REPLAY-ATTACK [64], FaceForensics [19]	99%	Enhance resistance to adversarial attacks
37	Ulutas et al. [65]	Bag-of-words (BoW)	Scale independent features transform (SIFT)	Surrey university library for forensic analysis (SULFA) [23]	97.5%	Limited focus on real-world scenarios
38	Zhao et al. [66]	Passive blind scheme	Hue-saturation-value (HSV), speeded up robust features (SURF), fast library for approximate nearest neighbors (FLANN)	10 test shots	99.01%	Limited to interframe forgeries
39	Voronin et al. [67]	Convolutional neural network (CNN)	Spatial-temporal procedure based on statistical analysis and CNN	3000 videos	96%	Future real-time application and comparisons
40	Carreira and Zisserman [68]	Inflated 3 dimension	Two stream inflated 3D ConvNet (I3D) based on 2D ConvNet	HMDB-51, UCF-101	80.2% HMDB-51, 97.9% UCF-101	Use kinetics for comprehensive experiments
41	D'Amiano et al. [69]	Dense field algorithm	3D PatchMatch based dense field algorithm	REWIND [31]	NA	Enhance video analysis
42	D'Avino et al. [70]	Recurrent neural network	Recursive network, long short-term memory	Randomly collected data	NA	Limited theoretical explanation
43	Cozzolino et al. [71]	Convolutional neural network (CNN)	Local descriptors, bag-of-words	Synthetic [72]	94%	Explore architectural improvements for deep learning
44	Bozkurt et al. [73]	Discrete cosine transform (DCT)	Correlation image generation, coarser forgery line detection, finer forgery line localization	Randomly collected data	98%	Not mention
45	Do et al. [74]	Deep convolutional neural network (DCNN)	Generative adversarial network (GAN)	Celeb-DF [20]	80%	Limited discussion of real-world scenarios
46	Long et al. [75]	Convolutional neural network	Convolutional 3D neural network (C3D), long short-term memory (LSTM)	2394 videos, YFCC100M [76]	98%	Improve frame dropping and LSTM
47	Su et al. [77]	Region duplication	Adaptive parameter-based fast compression tracking (AFCT)	Randomly collected data	93.1%	Detect diverse video forgery types
48	Mizher et al. [78]	Spatio termporal attacks	Falsifying techniques, fingerprint framework, secure system	Self-collected data	Not mention	Neglects complex video inpainting methods
49	Zhu et al. [79]	Spatiotemporal features	Scale invariant features transformation (SIFT)	TRECVID [80], CC_WEB_VIDEO [81]	99%	Limited evaluation of real-world scenarios

Table 6: Continued.

Sr. no	Reference	Approach	Contribution	Dataset	Model accuracy	Limitation
50	Barhoom et al. [82]	Physical random objects	Digital tampering, digital forensics	Randomly selected data	NA	Limited theoretical explanation
51	Abbasi Aghamaleki and Behrad [83]	Passive forensics	Extract appropriate quantization error rich	MPEGx codic [84]	92.73%	Limited theoretical explanation
52	Mathai et al. [85]	Statistical moment features	Normalization cross-correlation	SULFA [23]	88%	Limited accuracy in forgery detection
53	Rao and Ni [86]	Convolutional neural network	Spatial rich model, support vector classification	CASIA v1.0 [87], CASIA v2.0, DVMM [88]	98%, 97.8%, 96%	Limited theoretical explanation
54	Rigoni et al. [89]	Video tempering detection	Quantization index modulation, watermarking	Randomly collected data	96.5%	Limited theoretical explanation

Table 7: Summary of dataset.

Sr. no	Name/Year/Ref	Total sample	Training sample	Testing sample	Format	Dimension
1	Activity net 2019 [90]	58,000	32,000	8,000	Web videos MP4	Set to 256
2	ACID 2016 [45]	21 million	16,800,000	4,200,000	MP4	8 process
3	CC_WEB_VIDEO 2018 [81]	13,129	10,503	2,625	YouTube, Google, Yahoo! (MP4)	250 or 500 frames per second
4	Cityscapes 2020 [57]	20,000 weekly video frames	16,000 weekly video frames	4,000 weekly video frames	3D videos	50 frames per second
5	Davis 2017 [25]	12 h of data are 29	12 h of training data are 23	12 h of testing data are 6	Recorded video MP4	30 frames per second and 1280 × 720 pixels
6	Deepfake 2020 [27]	100,000	83,900	17,100	MP4, JSON, CSV	128 × 128 and 256 × 256 pixels
7	Dresden 2023 [47]	13,195	10,556	2,639	MPEG-4	1920 × 1080 pixels
8	DVMM 2016 [88]	7,491 frames	5,123 frames	2,368 frames	JPEG, BMP and TIFF	240 × 160 to 900 × 600 pixels
9	FaceForensics 2019 [19]	4,000	3,200	800	MP4	320 × 240 pixels
10	GRIP 2020 [29]	4990	3,992	998	Converted into HGS format	768 × 1024 pixels
11	HMDB-51 2017 [68]	6,766 video clips, 51 action, 101 category containing	70 clips	30 clips	Web video MP4	7 × 7 × 7 pixels
12	Celeb-DF 2019 [20]	590 YouTube video, 5,639 Deepfake	4,511	1,128	MP4	256 × 256 pixels
13	MFC 2019 [27]	176,000 high-provenance (HP) images and 11,000 HP videos	14,080 images, 8,800 videos	35,200 images, 2,200 videos	JPEG, MP4, MOV	1920 × 1080 pixels and 60 fps
14	MPEG codic 2016 [84]	8,000 videos	6,400	1,600	High efficiency video coding (HEVC), AO media video 1(AV1), MP4	7680 × 4320 pixels
15	REPLAY-ATTAC 2020 [64]	89,566	71,652	17,914	MP4, high quality videos	1920 × 1080 pixels and 60 fps
16	SIM10k 2018 [56]	10,000 frames from video games	8,300	1,700	JPEG	1250 × 375 pixels
17	SULFA 2011 [23]	150 videos	120	30	MP4	320 × 240 pixels
18	TRECVID 2016 [80]	1970 videos	200	80	MP4	4096 dimension feature vector
19	VFDD 2019 [42]	50 original videos	25	25	MP4	404 × 720 pixels
20	VIRAT 2015 [28]	75 videos	60	15	MP4	1280 × 720 pixels and 30 fps

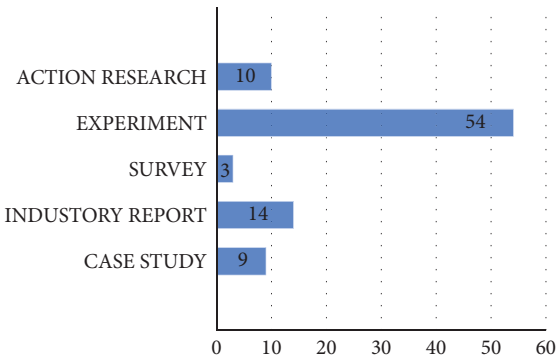

Figure 5: Publisher detail.

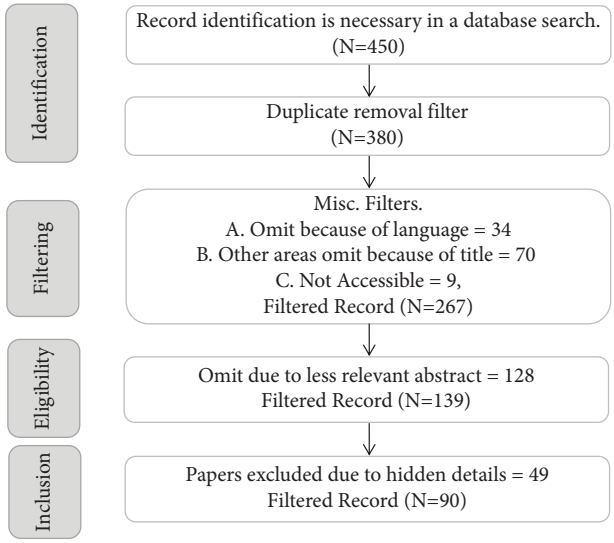

Figure 6: Primary study selection process.

elements of a convolutional neural network (CNN) and a recurrent neural network (RNN). This might be helpful in tasks like video analysis or action detection in videos that call for the processing of both visual and temporal information.

4. Categories of Forgeries

The prominent categories of forgeries in the domain are identified and discussed in this section.

4.1. Deepfake Based Forgery Detection. Wang et al. [7] worked to get a more thorough representation of the spatial and temporal features and suggested a "discrete cosine transform-based forgery clue augmentation network" (FCAN-DCT). "Compact Feature Extraction" (CFE) module and "Frequency Temporal Attention" (FTA) module are two branches of the FCAN-DCT, which also comprises of a backbone network. They thoroughly evaluated two datasets that use "visible light" "Wild Deepfake" and "Celeb-DF" (v2). They also created their own, self-created Deepfake NIR, the first video forgery dataset based on the near-infrared modality. Afchar et al. [55] propose an approach to quickly and effectively spot face tampering in films. It focuses on Deepfake and Face2Face, two current methods used to produce hyperrealistic fake videos. Use a dataset that is already available as well as one that have created using web videos to evaluate those fast networks. Do et al. [74] proposed a convolutional neural network to perform face forensic. Employ GANs to generate synthetic faces in a variety of resolutions and sizes to aid data augments. Furthermore, for strong face feature extraction, use a deep face recognition system to send weight to our system. Furthermore, the network is fine-tuned for real/fake picture categorization. I tried with the AI Challenge validation data and got decent results.

4.2. Frame-Based Forgery Detection. Munawar and Noreen [8] presented a deep learning method to resolve the issue of frame duplication with different frame rates. They proposed a novel deep learning framework made up of "Inflated 3D" (I3D) and "Siamese-based Recurrent Neural Network" (RNN). Their proposed method first extracted the characteristics and converted the movies into frames. To find frame-to-frame duplication, an original and a fake video were fed into the I3D network. Afterwards, several frames were combined to make a sequence. Sasikumar et al. [39] explore how, to eliminate video forgeries, the suggested approach employed two deep learning-based algorithms. First, the feature extraction approach is called "Scalar Invariant Feature Transform" (SIFT). The second method is "Mean Shift Clustering Algorithms" (MSCL), which groups comparable object frames from the retrieved video. The suggested model offers information on how many and which frames in a specific movie were faked. The suggested approach employs image processing methods and is a window-based application. Fayyaz et al. [46] explain how an attacker can compensate for a forged picture by adding SPN to it. It then suggests a forgery detection method for such a situation that would rely on the correlation between noise residue and SPN as well as noise residue from prior frames. Even if the attacker adds SPN to the forged frame to make up for it, the interframe continuity of noise will be interrupted and therefore be detected.

Joshi and Jain [48] represented a passive tampering detection technique that may be used on films recorded using variable-size GOP structures is described. First, all of the video frames from a specific video sequence are retrieved. Then, a video's real and reconstructed from temporal difference is determined for each pair of adjacent frames. Frame prediction error is used in the reconstruction of video. Last but not the least, tampering is located and found using the estimated discrepancies. Nguyen et al. [63] employ a capsule network to identify several types of spoofs, such as replay attacks that leverage printed pictures or recorded movies to create computer-generated videos. It expands the use of capsule networks to address inverted graphics issues beyond their initial purpose. Singh and Singh [54] present to find frame and area duplication forgeries in films and offer a passive-blind method using two separate algorithms. Examined

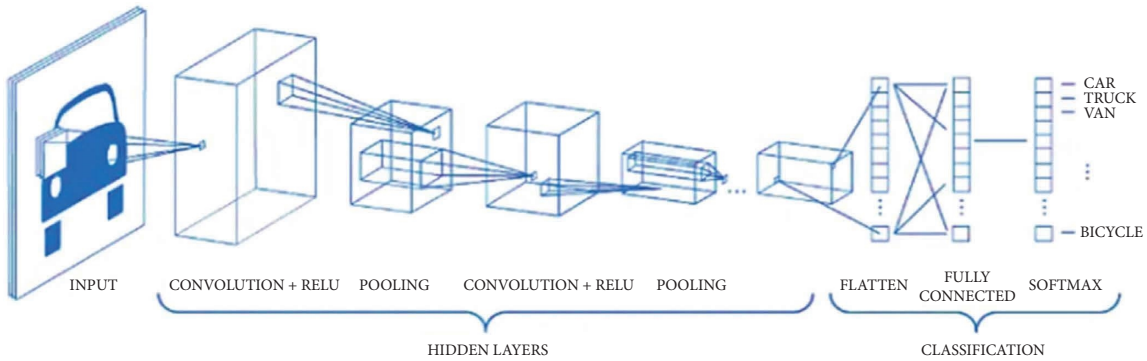

Figure 7: CNN architecture [91].

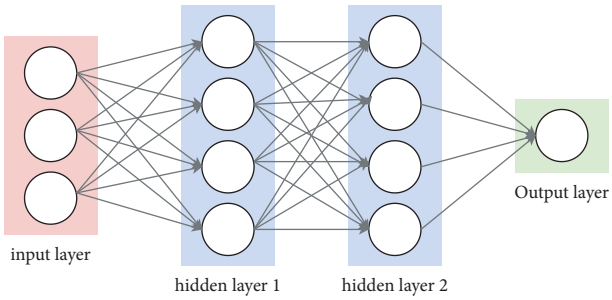

Figure 8: DNN architecture [92].

the video frame duplication forgery in three different ways, including duplication of a series of consecutive video frames at a long continue running position, duplication of numerous such sequences with various lengths at various locations, and duplication from other videos with different and identical dimensions, all of which can pose serious issues in a real-world setting. Analysed fabricated regular and irregular regions at various positions both within the same frame and from another frame to one or more sequences of subsequent frames from the same video at those same locations.

Ulutas et al. [65] propose a novel frame duplication detection approach based on the "Bag-of-Words" (BoW) model. Researchers employ the BoW model for retrieving images and videos after textual analysis. To find the order of repeated sections in the movie, frame features—visual word representations at key points—are employed. To increase efficiency and robustness, the approach computes thresholds based on the content. 31 test films from various movies and the "Surrey University Library for Forensic Analysis" (SULFA) are used to test the suggested technique. Zhao et al. [66] represent a passive-blind forensics method for video shots is suggested to identify interframe forgeries based on similarity analysis. The two components of this technique are the comparison of the "Hue, Saturation, Value" (HSV) colour histograms and the feature extraction using "Speeded Up Robust Features" (SURF) and "Fast Library for Approximate Nearest Neighbors" (FLANN) for double-checking. The forgery kinds in the tampered sites are then further confirmed using SURF feature extraction and FLANN matching.

Voronin et al. [67] accentuate the issue of detecting erased frames in movies. Frame dropping is a method of video editing in which a series of frames are dropped to jump ahead to certain parts of the source video. In digital video forensics, the automated identification of lost frames is a difficult problem. Outlines a method employing the spatial-temporal method based on the "convolutional neural network" and statistical analysis. Calculate confidence scores for each frame using the collection of several statistical procedures. Moreover, the output scores were obtained using a convolutional neural network. Bozkurt et al. [73] offer a novel video forgery detection technique with improved execution and detection capacity for detecting faked frames the frames' features are retrieved, and their correlations are shown as a correlation picture. To determine the forging operation, this approach studies a line on the correlation image. The identified line is then subjected to two new techniques (shrinking/expanding) to discover the precise position of the counterfeit. Su et al. [77] provide a rapid fraud detection technique for identifying region duplication in films based on Exponential-Fourier moments (EFMs). The system initially extracts EFMs characteristics from each block in the current frame before performing a rapid match to identify probable matching pairs. The updated areas are then located in the current frame using a postverification strategy (PVS). Finally, the "tampered areas are tracked in succeeding frames using an adaptive parameter-based fast compression tracking technique" (AFCT).

Barhoom et al. [82] suggested a novel reverse method to identify frame duplication in order to prevent theft by blocking the ip camera in particular locations. Abbasi Aghamaleki and Behrad [83] suggest a novel passive technique for tampering detection and localization in MPEGx-coded films. The suggested approach is capable of detecting double compression and frame insertion or deletion with various GOP formats and lengths. The quantization error traces on the residual errors of P frames are theoretically investigated in order to develop the suggested technique. Rigoni et al. [89] outlines a methodology for finding altered data in digital audio-visual material. The suggested architecture combines temporal and spatial watermarks without lowering the calibre of the host videos. Watermarks are embedded using a modified version of the "Quantization

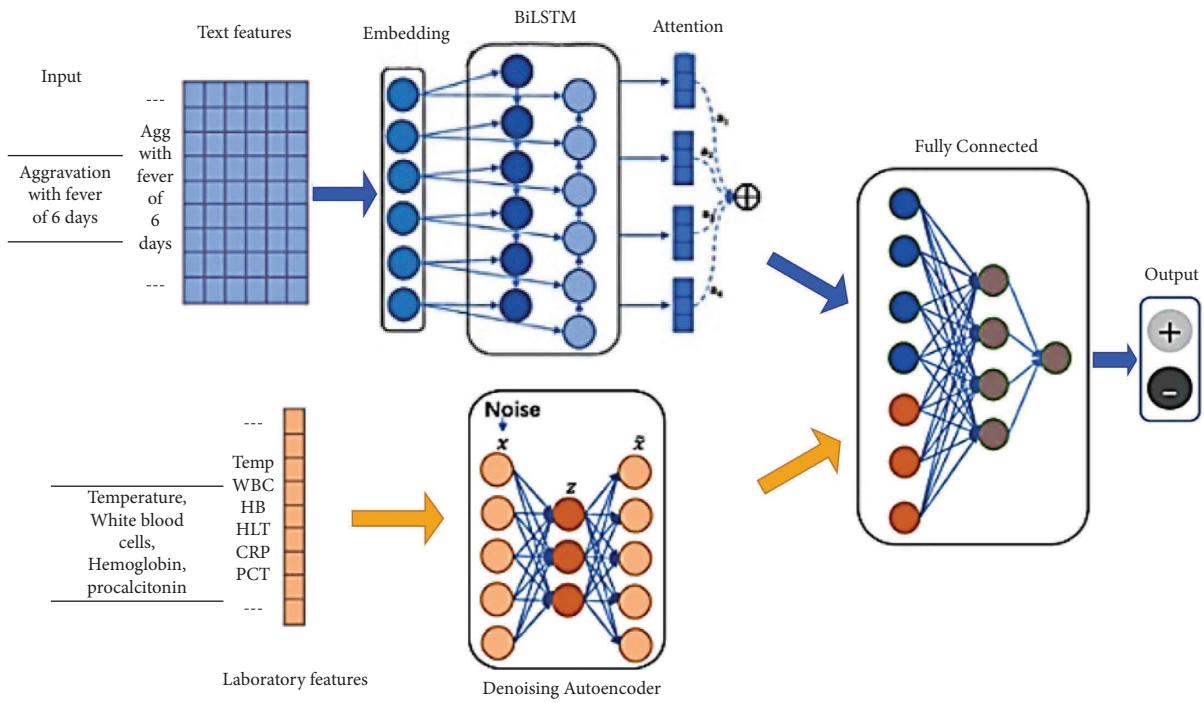

Figure 9: Hybrid neural network [93].

Index Modulation" (QIM) technique. The QIM watermarking algorithm's vulnerability enables pixel-level detection of local, global, and temporal-altering assaults. The method can also determine what kind of tampering assault it is. The framework is quick, reliable, and precise.

4.3. Copy-Move Forgery Detection. Parveen et al. [43] presented a pixel-based copy-move picture fraud detection technique is presented to validate the validity of digital images. The steps in the proposed technique are as follows: The steps are as follows: (1) convert the colour picture to grayscale; (2) split the grayscale image into overlapping blocks of size 88; (3) extract features using DCT based on several feature sets; (4) cluster blocks using the K-means method; and (5) use radix sort for feature matching. Experimental findings show that the suggested approach may successfully identify the fake part from digital photos. Pavlović et al. [51] represent a novel approach to the identification of such changes, employing both common statistical statistics and specific multifractal parameters as defining characteristics. Images are split into non-overlapping, fixed-size pieces before to processing. The typical characteristics are computed for each block. Employed a metaheuristic strategy to categorize the observed blocks and put forth a brand-new semimetric function for comparing the similarities between blocks. Simulation demonstrates that the suggested strategy delivers high accuracy and recall with minimal computing cost.

Liu et al. [52] show that the suggested technique, the superpixel of the picture is divided into complicated parts and smooth regions using *K*-means clustering technology. "Scale-Invariant Feature Transform" (SIFT) features are employed in complicated regions to find tampering. The sector mask feature and RGB colour feature are suggested as ways to identify tampering in smooth zones. For the copy-move detection, filtering out incorrect matching is done to these two categories of areas. Jia et al. [53] consider the three requirements and provide an innovative method to identify frame copy-move forgeries. Based on "optical flow" (OF) and stable parameters, a coarse-to-fine detection approach is developed. To locate potentially manipulated locations, coarse detection specifically examines the consistency of the OF sum. Following that, fine detection is carried out to determine the exact position of the forgery. To further limit false detection, duplicated frame pairs are matched using OF correlation. Voronin et al. [67] accentuate the issue of detecting erased frames in movies. Frame dropping is a method of video editing in which a series of frames are dropped to jump ahead to certain parts of the source video. In digital video forensics, the automated identification of lost frames is a difficult problem. Outlines a method employing the spatial-temporal method based on the "convolutional neural network" and statistical analysis. Calculate confidence scores for each frame using the collection of several statistical procedures. Moreover, the output scores were obtained using a convolutional neural network.

D'Amiano et al. [69] present a novel method for detecting and localising video copy-move frauds. It can be challenging to find well-crafted video copy-moves, especially when some uniform backdrop is duplicated to occlude foreground items. Employ a dense-field technique with invariant characteristics that provide resilience to multiple postprocessing processes to identify both additive and

occlusive copy moves. Mizher et al. [78] represent a several forms of video falsifying, video forgery detection approaches are researched and characterized, difficulties to current forgery detection systems are discussed, and a conclusion of proposed ideas is made. Zhu et al. [79] presented by monitoring the SIFT to create temporal concentration SIFT (TCSIFT), which substantially compresses the number of local features to eliminate visual redundancy while maintaining as many of the benefits of SIFT at the same time, SIFT features are stably stored with temporal information. Results from experiments on the two distinct datasets CC WEB VIDEO and TRECVID show that our technique can produce equivalent accuracy, a smaller storage size, a faster execution time, and the ability to adapt to varied video transformations.

4.4. Object-Based Forgery Detection. Kim et al. [6] proposed a brand-new object-based frame identification network. Their suggested technique leveraged symmetrically overlapping motion residuals to improve the ability to distinguish between video frames. Because their suggested motion residual characteristics were produced using overlapped temporal frames, the deep neural network took advantage of temporal fluctuations in the video stream. Alsakar et al. [9] showed the introduction and discussion of a newly created passive video forgery system. For the purpose of finding and detecting the two significant forms of forgeries, insertion and deletion, an arbitrary number of core tensors are chosen. To accomplish greater data reductions and to give useful characteristics to track forgeries throughout the whole film, these tensor data are orthogonally modified. Experimental findings and comparisons demonstrate the superiority of the suggested method, which can identify and locate both forms of assaults with a precision value of up to 99%.

Jin et al. [10] presented a dual-stream framework for object-based video forgery detection. First, discriminative characteristics are extracted using two distinct kinds of branches. A "Conditional Random Field" (CRF) layer is then applied to the segmentation results following the dual-stream feature fusion. Finally, the video tracking approach is incorporated to identify temporal consistency. To improve the localization outcomes, depth information is used. Aloraini et al. [40] explore how to identify forged regions in films and identify object removal forgery and provides a unique method based on sequential and patch analysis. By modelling video sequences as stochastic processes, sequential analysis may be used to identify fake videos by monitoring changes in their properties. By employing anomalous patches to visualize the movement of deleted items, pinpoint fabricated areas.

Chen et al. [56] enhance object detection's resilience across domains. Approach the domain shift from two angles: (1) the instance-level shift, such as object appearance, size, etc., and (2) the image-level change, such as picture style, lighting, etc. To lessen the domain difference, propose two domain adaptation components at the picture and instance levels based on the most recent state-of-the-art Faster R-CNN model. Based on the H-divergence theory, the two domain adaptation components are accomplished through the adversarial training of a domain classifier. Aneja et al. [59] create a convolutional captioning method for images. On the difficult MSCOCO dataset, show its effectiveness and show performance comparable to the LSTM baseline with a shorter training time per parameter. Provide convincing arguments in support of convolutional language generation techniques by performing a thorough examination. Mathai et al. [85] suggested using statistical moment characteristics and a normalised cross-correlation factor to detect and locate video forgeries. For each frame block, the characteristics from the prediction-error array are computed (set of a certain number of continuous frames in the video). When compared to other nonduplicated frame blocks, the normalised cross-correlation of those attributes will be higher between duplicated frame blocks. The duplication is verified by utilising a determined threshold based on the mean squared error. Using the method, the position of the duplicated block is also discovered.

5. Discussion of Challenges

There are several challenges concluded from the discussion in previous sections. The reliance on static images rather than video sequences for Deepfake detection [2] is one of many. Furthermore, lack of evaluation on real-world scenario video forgery scenarios [5], large-scale datasets for comprehensive performance assessment [3], absence of extensive evaluation on diverse video datasets, and limiting the generalizability of the proposed framework [4] are the key challenges.

The reliance on motion residual-based analysis is not effective in detecting tampering techniques which does not significantly alter the motion patterns [6]. Existing face forgery detection methods based on frequency domain only focus on single frames and overlook the discriminative part and temporal frequency clues among different frames in synthesized videos [7].

Identification of forged duplication frames in large videos with variant frame rates in real time is not feasible due to computational limitations, lack of generalization, and low-performance accuracy [8]. Scheme based on tensor representation and orthogonal tracing feature algorithms may have limitations in detecting and locating insertion and deletion forgery in videos for more types of attacks beyond the ones tested [9]. Performance in complex or crowded environments, robustness to variations in Wi-Fi signal strength or scalability to larger surveillance networks are also the areas of concern [35].

System's ability to detect multiple interframe forgeries within a single video is an issue that is not explored explicitly [32]. The scalability of the proposed framework and detection of complex types of forgeries may pose challenges when adding more models and coding parameters, requiring further research and exploration [36]. The specific types or variations of video face forgery detection that may not be effectively addressed by the temporal cues are also an area that requires investigation [33]. The approaches have limitations in terms of efficiency and computational

requirements, especially for machines with low memory [37]. Performance and accuracy when faced with more complex and advanced types of video forgeries beyond the ones mentioned (such as content-aware manipulations or Deepfake videos) becomes even worse [38]. System's security and performance in the presence of large geometric attacks like editing or embedding logos is also a hot topic to investigate deeper [39]. A lot of work is done on detecting forged videos with object removal and moving backgrounds; however, further investigation is needed to improve detection performance at the pixel level [40].

6. Discussion for Key Questions

6.1. In the Literature, Which Technologies, Models, Methods, and Practices Are Discussed? To address the challenges, a novel deep learning framework made up of "Inflated 3D" (I3D) and "Siamese-based Recurrent Neural Network" (RNN) is suggested. The extraction of characteristics and conversion of movies into frames is the first stage in the suggested framework. To find frame-to-frame duplication, an original and a fake video are sent to the I3D network. Incredibly effective way using a large convolutional neural network to reveal interframe manipulation in videos. A decoder that forecasts the tampering mask and a 3D-UNet-based watermark embedding network. Clustering methods to expedite the block matching approach throughout the process of detecting image fraud. To create a more thorough spatial-temporal feature representation, a "Forgery Clue Augmentation Network" (FCAN-DCT) based on the discrete cosine transform is used. "Compact Feature Extraction" (CFE) module and "Frequency Temporal Attention" (FTA) module are two branches of the FCAN-DCT, which also comprises of a backbone network.

6.2. When Implementing Forged Video Detection, What Techniques/Methods/Models Are Developed for Metric Selection? Frame duplication is recognized using a "Siamese (twin) RNN" integrated with I3D from a collection of forged sequence frames. These two forms of video forgeries are easily found and located using SVD tube-fibber tensor construction. To improve segmentation results, dual-stream feature fusion and a "Conditional Random Field" (CRF) layer are used. Deep automated feature extraction is performed using a "2D convolution neural network" (2D-CNN) of spatiotemporal information and fusion, and classification is performed using a Gaussian RBF multiclass support vector machine (RBF-MSVM). A network of "fully temporal convolutions network" (FTCN). The major finding of FTCN is to keep the temporal convolution kernel size constant while reducing the size of the spatial convolution kernel. Video interframe manipulation using a "deep convolutional neural network" (DCNN).

6.3. What Mitigation Techniques Are Advised to be Used to Recognize Forged Videos? CNN's core model is based on the Xception model. Deep learning architectures are appropriate for object-based forgery detection in a more sophisticated H.265 encoding format and in more realistic real-world scenarios. Examine how successfully video forgery detection can be used in the context of increasingly hostile and real-world disturbances. Improving the process for discovering and recognizing additional sorts of assaults. Technology can identify several interframe forgeries in a single video.

7. Conclusions

This article presents a comprehensive overview of significant research endeavours aimed at identifying counterfeit videos, a pressing issue exacerbated by the widespread accessibility of image and video editing tools in today's technology-driven era. As digital manipulation capabilities have become increasingly accessible, the potential for misuse and harm has grown in tandem with the creative possibilities. The malicious use of facial alteration raises concerns regarding real-world injustices inflicted upon unsuspecting individuals. The emergence of potent computer software and mobile applications has ushered in a new era of visual manipulation, allowing virtually anyone to manipulate images and videos effortlessly. The diligent systematic literature review (SLR) conducted for this study has encompassed pivotal research articles, revealing various methods for identifying forged videos. These methods leverage cutting-edge technologies, including deep neural networks, convolutional neural networks, Deepfake analysis, watermarking networks, and clustering, among others. The synergy of success factors and mitigation strategies forms a cohesive framework of solutions to address the multifaceted challenges associated with counterfeit video content. The study concludes that major challenges include frame duplication, deletion, and insertion. Furthermore, frame rate fluctuation and loop detection are also important problems from the standpoint of duplicated frames. Due to computational restrictions, a lack of generalization, poor performance accuracy, and real-time identification of counterfeit duplicating frames for big films with different frame rates, this is not feasible.

Data Availability

The data used to support the findings of this study are included within this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. Tyagi and D. Yadav, "A detailed analysis of image and video forgery detection techniques," *The Visual Computer*, vol. 39, no. 3, pp. 813–833, 2023.

[2] S. Ganguly, S. Mohiuddin, S. Malakar, E. Cuevas, and R. Sarkar, "Visual attention-based Deepfake video forgery detection," *Pattern Analysis & Applications*, vol. 25, no. 4, pp. 981–992, 2022.

[3] V. Kumar, M. Gaur, and V. kansal, "Deep feature based forgery detection in video using parallel convolutional neural network: VFID-Net," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 42223–42240, 2022.

[4] S. Tan, B. Chen, J. Zeng, B. Li, and J. Huang, "Hybrid deep-learning framework for object-based forgery detection in video," *Signal Processing: Image Communication*, vol. 105, pp. 116695–116710, 2022.

[5] Y. Zhou, Q. Ying, X. Zhang, Z. Qian, S. Li, and X. Zhang, "Robust watermarking for video forgery detection with improved imperceptibility and robustness," 2022, https://arxiv.org/abs/2207.03409.

[6] T. H. Kim, C. W. Park, and I. K. Eom, "Frame identification of object-based video tampering using symmetrically overlapped motion residual," *Symmetry*, vol. 14, no. 2, pp. 364–415, 2022.

[7] Y. Wang, C. Peng, D. Liu, N. Wang, and X. Gao, "Spatial-temporal frequency forgery clue for video forgery detection in VIS and NIR scenario," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, p. 1, 2023.

[8] M. Munawar and I. Noreen, "Duplicate frame video forgery detection using Siamese-based RNN," *Intelligent Automation & Soft Computing*, vol. 29, no. 3, pp. 927–937, 2021.

[9] Y. M. Alsakar, N. E. Mekky, and N. A. Hikal, "Detecting and locating passive video forgery based on low computational complexity third-order tensor representation," *Journal of Imaging*, vol. 7, no. 3, pp. 47–25, 2021.

[10] X. Jin, Z. He, Y. Wang, J. Yu, and J. Xu, "Towards general object-based video forgery detection via dual-stream networks and depth information embedding," *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 35733–35749, 2022.

[11] T. Nayerifard, H. Amintoosi, A. G. Bafghi, and A. Dehghantanha, "Machine learning in digital forensics: a systematic literature review," 2023, http://arxiv.org/abs/2306.04965.

[12] L. Stroebel, M. Llewellyn, T. Hartley, T. S. Ip, and M. Ahmed, "A systematic literature review on the effectiveness of Deepfake detection techniques," *Journal of Cyber Security Technology*, vol. 7, no. 2, pp. 83–113, 2023.

[13] R. Chauhan, R. Popli, and I. Kansal, "A comprehensive review on fake images/videos detection techniques," in *Proceedings of the 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 1–6, Noida, India, October 2022.

[14] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: a systematic literature review," *IEEE Access*, vol. 10, pp. 25494–25513, 2022.

[15] H. F. Shahzad, F. Rustam, E. S. Flores, J. Luís Vidal Mazón, I. de la Torre Diez, and I. Ashraf, "A review of image processing techniques for Deepfakes," *Sensors*, vol. 22, no. 12, pp. 4556–4628, 2022.

[16] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: a Survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131–148, 2020.

[17] D. Yadav and S. Salmani, "Deepfake: a survey on facial forgery technique using generative adversarial network," in *Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 852–857, IEEE, Madurai, India, May 2019.

[18] Z. Wang, J. Bao, W. Zhou, W. Wang, and H. Li, "AltFreezing for more general video face forgery detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4129–4138, Vancouver, Canada, June 2023.

[19] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, "FaceForensics++: learning to detect manipulated facial images," 2019, https://arxiv.org/abs/1901.08971.

[20] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: a large-scale challenging dataset for DeepFake forensics," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3204–3213, Vancouver, Canada, June 2020.

[21] D. Liu, Z. Zheng, C. Peng et al., "Hierarchical forgery classifier on multi-modality face forgery clues," *IEEE Transactions on Multimedia*, vol. 14, no. 8, pp. 1–12, 2023.

[22] V. Kumar and M. Gaur, "Multiple forgery detection in video using inter-frame correlation distance with dual-threshold," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 43979–43998, 2022.

[23] G. Qadir, S. Yahaya, and A. T. S. Ho, "Surrey university library for forensic analysis (SULFA) of video content," *IET Conference on Image Processing*, vol. 600, pp. 1–5, 2012.

[24] A. Kohli, A. Gupta, and D. Singhal, "CNN based localisation of forged region in object-based forgery for HD videos," *IET Image Processing*, vol. 14, no. 5, pp. 947–958, 2020.

[25] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "DDD17: end-to-end DAVIS driving dataset," 2017, http://arxiv.org/abs/1711.01458.

[26] B. Zi, M. Chang, J. Chen, X. Ma, and Y. G. Jiang, "WildDeepfake: a challenging real-world dataset for Deepfake detection," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2382–2390, Ottawa, Canada, November 2020.

[27] W. Guan, A. K. Venkatesh, X. Bai et al., "Time to hospital arrival among patients with acute myocardial infarction in China: a report from China PEACE prospective study," *European heart journal. Quality of care & clinical outcomes*, vol. 5, no. 1, pp. 63–71, 2019.

[28] J. Moon, Y. Kwon, K. Kang, and J. Park, "ActionNet-VE dataset: a dataset for describing visual events by extending virat ground 2.0," *Image Processing and Pattern Recognition (SIP)*, vol. 9, pp. 1–4, 2015.

[29] R. Malhotra, M. I. Tareque, N. C. Tan, and S. Ma, "Association of baseline hand grip strength and annual change in hand grip strength with mortality among older people," *Archives of Gerontology and Geriatrics*, vol. 86, Article ID 103961, 2020.

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

[31] H. Zhang, M. Wang, R. Hong, and T. S. Chua, "Play and rewind: optimizing binary representations of videos by self-supervised temporal hashing," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 781–790, New York, NY, USA, October 2016.

[32] S. Fadl, Q. Han, and Q. Li, "CNN spatiotemporal features and fusion for surveillance video forgery detection," *Signal Processing: Image Communication*, vol. 90, pp. 116066–116132, 2021.

[33] Y. Zheng, J. Bao, D. Chen, M. Zeng, and F. Wen, "Exploring temporal coherence for more general video face forgery detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 15024–15034, Montreal, Canada, December 2021.

[34] B. Dolhansky, J. Bitton, B. Pflaum et al., "The DeepFake detection challenge (DFDC) dataset," 2020, http://arxiv.org/abs/2006.07397.

[35] Y. Huang, X. Li, W. Wang, T. Jiang, and Q. Zhang, "Towards cross-modal forgery detection and localization on live surveillance videos," in *Proceedings of the IEEE Conference on Computer Communications*, pp. 1–10, Vancouver, Canada, May 2021.

[36] S. Verde, E. D. Cannas, P. Bestagini, S. Milani, G. Calvagno, and S. Tubaro, "FOCAL: a forgery localization framework based on video coding self-consistency," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 217–229, 2021.

[37] H. Kaur and N. Jindal, "Deep convolutional neural network for graphics forgery detection in video," *Wireless Personal Communications*, vol. 112, no. 3, pp. 1763–1781, 2020.

[38] J. L. Zhong, C. M. Pun, and Y. F. Gan, "Dense moment feature index and best match algorithms for video copy-move forgery detection," *Information Sciences*, vol. 537, pp. 184–202, 2020.

[39] R. Sasikumar, K. T. Nasreen, and C. Jeganathan, "Video forgery detection using deep learning techniques and clustering algorithms," *Studia Rosenthaliana*, vol. 12, no. 207, pp. 207–216, 2020.

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

[41] X. Hau Nguyen, Y. Hu, M. Ahmad Amin, K. Gohar Hayat, V. Thinh Le, and D.-T. Truong, "Detecting video inter-frame forgeries based on convolutional neural network model," *International Journal of Image, Graphics and Signal Processing*, vol. 12, no. 3, pp. 1–12, 2020.

[42] Z. Qi, R. Zhu, Z. Fu, W. Chai, and V. Kindratenko, "Weakly supervised two-stage training scheme for deep video fight detection model," in *Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 677–685, Macau, China, October 2022.

[43] A. Parveen, Z. H. Khan, and S. N. Ahmad, "Block-based copy–move image forgery detection using DCT," *Iran Journal of Computer Science*, vol. 2, pp. 89–99, 2019.

[44] B. C. Hosler, X. Zhao, O. Mayer, C. Chen, J. A. Shackleford, and M. C. Stamm, "The video authentication and camera identification database: a new database for video forensics," *IEEE Access*, vol. 7, pp. 76937–76948, 2019.

[45] S. Kamal, S. H. Ripon, N. Dey, A. S. Ashour, and V. Santhi, "A MapReduce approach to diminish imbalance parameters for big deoxyribonucleic acid dataset," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 191–206, 2016.

[46] M. A. Fayyaz, A. Anjum, S. Ziauddin, A. Khan, and A. Sarfaraz, "An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues," *Multimedia Tools and Applications*, vol. 79, no. 9-10, pp. 5767–5788, 2020.

[47] M. Carstens, F. M. Rinner, S. Bodenstedt et al., "The dresden surgical anatomy dataset for abdominal organ segmentation in surgical data science," *Scientific Data*, vol. 10, no. 1, pp. 3–8, 2023.

[48] V. Joshi and S. Jain, "Tampering detection and localization in digital video using temporal difference between adjacent frames of actual and reconstructed video clip," *International Journal on Information Technology*, vol. 12, no. 1, pp. 273–282, 2020.

[49] C. C. Chen, W. Y. Lu, and C. H. Chou, "Rotational copy-move forgery detection using SIFT and region growing strategies," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18293–18308, 2019.

[50] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD-new database for copy-move forgery detection," in *Proceedings of the ELMAR-2013*, pp. 49–54, Zadar, Croatia, September 2013.

[51] A. Pavlović, N. Glišović, A. Gavrovska, and I. Reljin, "Copy-move forgery detection based on multifractals," *Multimedia Tools and Applications*, vol. 78, no. 15, pp. 20655–20678, 2019.

[52] Y. Liu, H. Wang, Y. Chen, H. Wu, and H. Wang, "A passive forensic scheme for copy-move forgery based on superpixel segmentation and K-means clustering," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 477–500, 2020.

[53] S. Jia, Z. Xu, H. Wang, C. Feng, and T. Wang, "Coarse-to-Fine copy-move forgery detection for video forensics," *IEEE Access*, vol. 6, pp. 25323–25335, 2018.

[54] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11527–11562, 2019.

[55] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a compact facial video forgery detection network," in *Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, Hong Kong, China, December 2018.

[56] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3339–3348, Salt Lake City, UT, USA, June 2018.

[57] M. Cordts, M. Omran, S. Ramos et al., "The cityscapes dataset," in *Proceedings of the Computer Vision Foundation*, pp. 1–4, Salt Lake City, UT, USA, June 2015.

[58] J.-E. Deschaud, "KITTI-CARLA: a KITTI-like dataset generated by CARLA Simulator," 2021, http://arxiv.org/abs/2109.00892.

[59] J. Aneja, A. Deshpande, and A. G. Schwing, "Convolutional image captioning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 5561–5570, 2018.

[60] T. Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Proceedings of the Computer Vision Foundation*, vol. 8693, no. 5, pp. 740–755, New York, NY, USA, June 2014.

[61] Z. Shou, J. Pan, J. Chan et al., "Online action detection in untrimmed, streaming videos," eccv, arXiv prepr," 2018, http://arxiv.org/abs/1802.06822.

[62] J. Gao, Z. Yang, and C. Sun, "Turn tap: temporal unit regression network for temporal action proposals supplementary materials university of southern California," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV 2017)*, vol. 880, pp. 3628–3636, Venice, Italy, October 2017.

[63] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: using capsule networks to detect forged images and videos," in *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2307–2311, Brighton, Great Britain, May 2019.

[64] A. A. Elsaeidy, N. Jagannath, A. G. Sanchis, A. Jamalipour, and K. S. Munasinghe, "Replay attack detection in smart cities using deep learning," *IEEE Access*, vol. 8, pp. 137825–137837, 2020.

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

[66] D. N. Zhao, R. K. Wang, and Z. M. Lu, "Inter-frame passive-blind forgery detection for video shot based on similarity

analysis," *Multimedia Tools and Applications*, vol. 77, no. 19, pp. 25389–25408, 2018.

[67] V. V. Voronin, R. Sizyakin, I. Svirin, A. Zelensky, and A. Nadykto, "Detection of deleted frames on videos using a 3D convolutional neural network," *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II*, vol. 10802, pp. 239–246, 2018.

[68] J. Carreira and A. Zisserman, "Quo vadis, action recognition," in *Proceedings of the The Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6299–6308, New Orleans, LA, USA, June 2022.

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

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

[71] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security Workshop--IH&MMSec'17*, pp. 159–164, Philadelphia, PA, USA, June 2017.

[72] T. E. Raghunathan, "Synthetic data," *Annual Review of Statistics and Its Application*, vol. 8, no. 1, pp. 129–140, 2021.

[73] I. Bozkurt, M. H. Bozkurt, and G. Ulutaş, "A new video forgery detection approach based on forgery line," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 25, no. 6, pp. 4558–4574, 2017.

[74] N. Do, I. Na, and S. Kim, *Forensics Face Detection From Gans Using Convolutional Neural Network*, ISITC, Bridgewater, NJ, USA, 2018.

[75] C. Long, E. Smith, A. Basharat, and A. Hoogs, "A C3D-based convolutional neural network for frame dropping detection in a single video shot," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1898–1906, Honolulu, HI, USA, July 2017.

[76] B. Thomee, D. A. Shamma, G. Friedland et al., "YFCC100M: the new data in multimedia research," *Communications of the ACM*, vol. 59, no. 2, pp. 64–73, 2016.

[77] L. Su, C. Li, Y. Lai, and J. Yang, "A fast forgery detection algorithm based on exponential-Fourier moments for video region duplication," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 825–840, 2018.

[78] M. A. Mizher, M. C. Ang, A. A. Mazhar, and M. A. Mizher, "A review of video falsifying techniques and video forgery detection techniques," *International Journal of Electronic Security and Digital Forensics*, vol. 9, no. 3, pp. 191–208, 2017.

[79] Y. Zhu, X. Huang, Q. Huang, and Q. Tian, "Large-scale video copy retrieval with temporal-concentration SIFT," *Neurocomputing*, vol. 187, pp. 83–91, 2016.

[80] D. D. Le, S. Phan, V. T. Nguyen et al., "NiI-Hitachi-Uit at TRECVID 2016," 2016, https://www-nlpir.nist.gov/projects/tvpubs/tv16.papers/nii-hitachi-uit.pdf.

[81] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras, and Y. Kompatsiaris, "Near-duplicate video retrieval with deep metric learning," in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 347–356, Venice, Italy, October 2017.

[82] Academia, "Frame duplication forgery detection using physical random," *Saba Journal of Information Technology And Networking (SJITN)*, vol. 4, no. 1, pp. 13–19, 2016.

[83] J. Abbasi Aghamaleki and A. Behrad, "Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding," *Signal Processing: Image Communication*, vol. 47, pp. 289–302, 2016.

[84] B. Taraghi, H. Amirpour, and C. Timmerer, "Multi-codec ultra high definition 8K MPEG-DASH dataset," in *Proceedings of the 13th ACM Multimedia Systems Conference*, pp. 216–220, Athlone, Ireland, June 2022.

[85] M. Mathai, D. Rajan, and S. Emmanuel, "Video forgery detection and localization using normalized cross-correlation of moment features," in *Proceedings of the 2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pp. 149–152, Santa Fe, NM, USA, March 2016.

[86] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, New York, NY, USA, December 2016.

[87] T. T. Jing Dong, "Casia image tampering detection evaluation database," in *Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing*, pp. 422–426, Beijing, China, July 2013.

[88] T.-T. Ng and S. Chang, *A Data Set of Authentic and Spliced Image Blocks*, Columbia University, New York, NY, USA, 2004.

[89] R. Rigoni, P. G. Freitas, and M. C. Q. Farias, "Detecting tampering in audio-visual content using QIM watermarking," *Information Sciences*, vol. 328, pp. 127–143, 2016.

[90] Z. Yu, D. Xu, J. Yu et al., "ActivityNet-QA: a dataset for understanding complex web videos via question answering," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 9127–9134, 2019.

[91] P. A. Gutiérrez, "Convolutional neural networks, explained," 2020, https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939.

[92] J. Johnson, "What's a deep neural network? Deep nets explained," 2020, https://www.bmc.com/blogs/deep-neural-network/.

[93] P. A. Gutiérrez and C. Hervás-Martínez, "Hybrid artificial neural networks: models, algorithms and data," *Advances in Computational Intelligence*, vol. 6692, pp. 177–184, 2011.