System for Detecting Deepfake in Videos - A Survey

Kusuma S
kusumasriram@gmail.com

Harshitha G S
harshithags2000@gmail.com

Vardhini B H
vardaiwath2007@gmail.com

Vani Shiva Bhat
bhatvani8884@gmail.com

Annapurna Ramakrishna Shanbhag
annapurnashanbhag15@gmail.com

Department of ISE, RNS Institue of Technology, Bangalore, India

***Abstract*—In recent times, freely available software grounded on Machine literacy ways has resulted in the generation of veritably realistic fake content that has counter accusations on society in an period of fake news. Software such as FaceApp are freely available and can be used by anyone to create realistic looking fake videos. Such videos if used with a bad intent also the consequences can be serious and can affect society and people. On the other hand, important exploration has been done in order to develop discovery styles to reduce the destructive consequences of deepfakes.This paper provides a review of which are used to descry similar manipulated videos. We explore several methods used to create face based manipulation videos and compare a number of deepfake discovery ways grounded on several parameters which includes generation styles, methodology used, datasets etc.**

***Index Terms*—Deepfake detection, Deep Learning, Generative Adversial Networks (GANs), Convolution Neural Networks (CNN).**

I. Introduction

Owing to the progress in deep learning technology as well as computer vision in recent years, a surge has been seen in fake face media. Every day, a large number of DF photos and videos are shared on social media platforms. DF films are spreading, feeding fake news and endangering social, national, and international ties. People are worried that what they read on the internet or watch on the internet is no longer reliable and trustworthy. In this backdrop, in January 2020, a popular social media platform announced a new policy prohibiting AI-manipulated videos that might deceive the viewers during elections. The problem is that this is dependent on the capacity to tell the difference between actual and false videos. Creation of Deepfake videos is based on the idea of replacing a person's face with somebody else's face. The requirement to achieve this is that the sufficient number of images of both the persons must be available. Research conducted in recent times has focused on how these deepfake videos are crafted and how to recognise them by analysing distinctive features closely. But with the advancement in technology, it has become increasingly challenging to tell apart fake videos from the real ones.

II. Technical Background

A. CNN

A Deep Learning advances in computer vision have been constructed and improved over time, largely through one technique – a Convolutional Neural Net-based approach work[31]

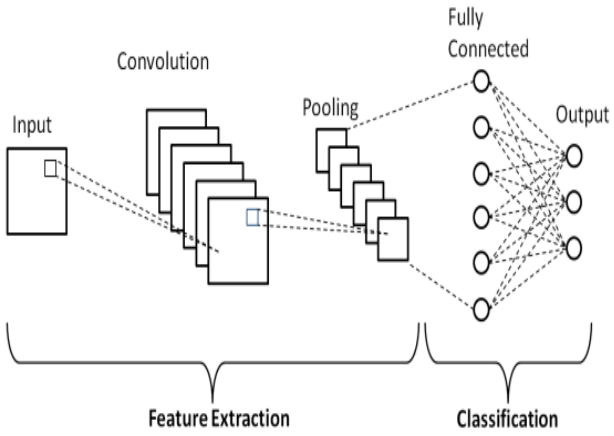

Fig. 1. Basic Architecture of CNN

CNNs are a type of deep learning network[31]. An algorithm that can take an image as input and give priority to distinct aspects/objects in the image (learnable biases and weights) while distinguishing between them. A CNN requires significantly less pre-processing than a neural network. In contrast to other classification systems, CNN has hand-engineered filters in its core techniques, which they can use with the proper training. The ability to detect these filters/features is determined by the design of the building. The study of Neurons in the Human Brain of the Organization of the Visual Cortex influenced the connectivity pattern of a CNN, which is analogous to the connectivity pattern..

The CNN's job is to condense the pictures into a format which is simpler to process while retaining important elements for accurate prediction. This is essential for designing an

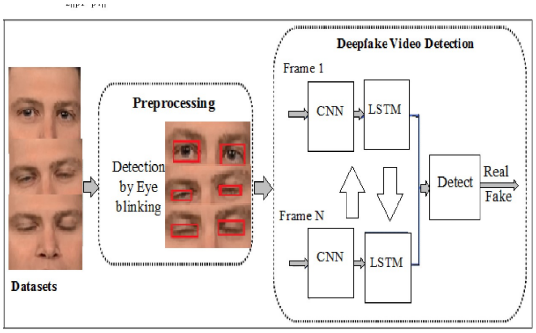

Fig. 2. DeepFake Detection using CNN and LSTM

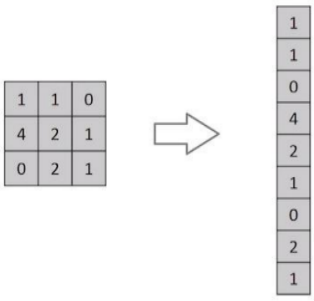

Fig. 3. Flattening of a 3x3 image matrix into a 9x1 vector

architecture that can learn features while also being scalable to large datasets

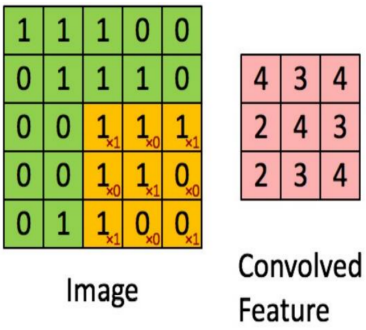

Fig. 4. Image to convoluted feature

In order to retrieve high level characteristics such as edges from the source images, CNN is used. CNN doesn't have to have a single Convolutional Layer. Traditionally, the first ConvLayer is in control of capturing Low Level information like edges, color, gradient direction, and so on. The architecture gets adjusted to the High-Level characteristics as layers are added, resulting in a network that comprehends the pictures in the dataset in the same way as we do

B. RNN

An RNN[32] is a neural network wherein the result of the previous step is being utilised as input in the following step. This in contrast to regular neural networks where all the inputs

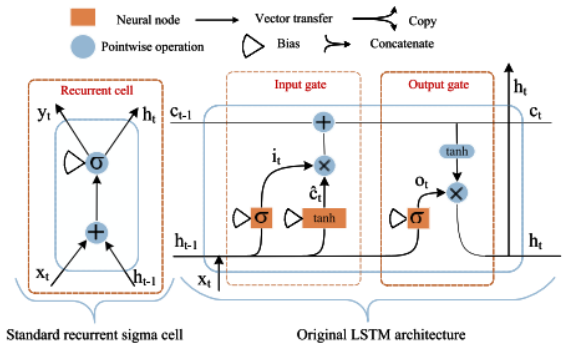

Fig. 5. RNN's fundamental architecture

and outputs are not dependent on each other. However, in few situations, for example trying to predict the following term of a phrase, the previous words are crucial, and thus the previous words must be remembered.As a result, RNNs were synthesised. They utilise the intermediate layers to solve a problem. The hidden state in RNN remembers information about the hidden sequence. RNNs have "memory" that stores all of the results of the calculations. The memory uses the same configurations for all inputs or hidden layers. Therefore, it achieves exact same outcome by performing identical work on all inputs or hidden layers.

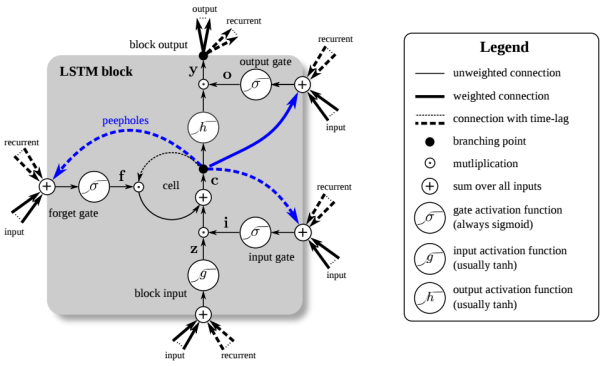

Fig. 6. LSTM Units

Hochreiter and Schmidhuber[33] LSTM was proposed to deal with long-term dependencies whenever the separation among relevant data input is great. It achieves all the intriguing result based on RNN and hence they have been the centre of deep learning. The recurrent hidden layers of RNN's,are composed of recurring cells whose conditions are controlled prior instances as well as the current sources via feedback networks.

C. GAN

Texts,images and video generation, drug discovery, and text-to-image conversion have all been employed in real-world applications.

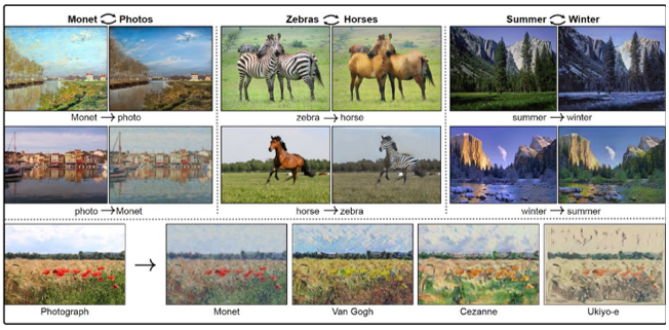

Fig. 7. Image to Image synthesis using GAN

GANs are among the most predominant Machine Learning techniques devised in recent times. In a nutshell, they are algorithms from the generative models group. These algorithms are a subcategory of the unsupervised learning discipline, that concentrates on algorithms which understand the fundamental structure of the data without defining a specific value. Generative models discover the inherent distribution of data input p(x), allowing them to generate synthesised input data x' and output data y', generally utilising hidden parameters.
Generally GAN composed of two models:

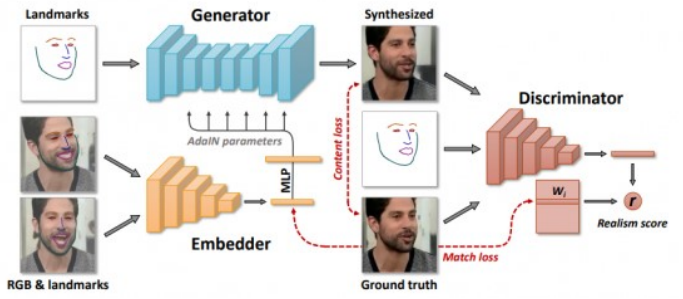

Fig. 8. DeepFake created using GAN Model

- The first model, referred to as a Generator, attempts to develop new information which is exactly equivalent to the simulated values. The Generator is like a person who creates forged artwork.
- The next model is the Discriminator. The primary objective of this model is to find whether an input data set is 'real,' indicating it relates to the original dataset, or 'fake,' meaning a forger created it was created by a forger. A Discriminator in this situation is similar to an art expert who attempts to determine not whether works of art are genuine.
- The above-mentioned Generator is modelled using a neural network G(z,θ1). Its job is to translate the input noise variables z into the required data space x. (say images). A second neural network D(x, θ2), on the other hand, models the discriminator and produces the likelihood that the data originated from the real dataset, in the range of 0% to 100%. (0,1). The weights or parameters that define each neural network are represented by $theta$i in both situations.
- As a consequence of this training, the Discriminator can accurately identify input data as true or false. This implies that its weight values are optimised to optimise the likelihood of any real input data x being classified as a part of the genuine dataset whilst also reducing the likelihood of any false picture being categorized as the real dataset. In more technical jargon, with the help of loss/error function, D(x) is optimized and D(G(z)) is minimized.
- Besides that, the Generator has been trained to misguide the Discriminator by producing data which is as real as possible, suggesting that Generator's weights have been optimized to increase the probability that any counterfeit picture will be categorized as belonging to the legitimate dataset. In mathematical terms, this means that the network's loss/error function maximizes D(G(z)).

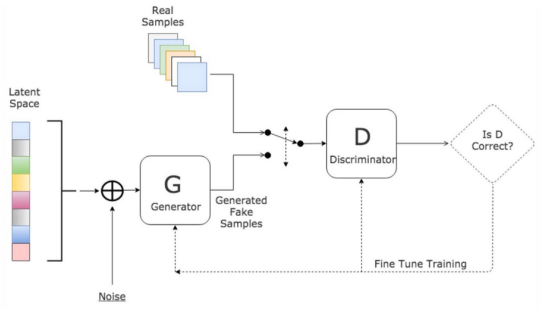

Fig. 9. Global concept of GAN

since the neural networks represent both generator and discriminator, the GAN can be trained with the aid of a gradient based optimization technique

III. Face-based Video Manipulation Methods

In the last twenty years, the prominence of simulated face tampering has shot up. Zollhofer et al. [16] provided a detailed report. Bregler et al. [17] specifically presented Video Rewrite,to create a new different video artificially of an individual with differing mouth movements by using image-based method. With video face replacement, one of the very basic automatic face swap approaches [18] were proposed by Dale et al. They use single camera movies to recreate a three - dimensional model of those two faces and then use the resulting 3D model. The first facial reenactment expression transfer was accomplished by Thies et al. [19]. A consumer-level RGB-D camera was used for recreation and tracking 3D representation of the source and target actors both. The analysed deformities in original faces are applied to model of face to be modified. To convert the graphically computed modification of faces back to their original form, Kim et al. [22] learned an image translation neural network. NeuralTextures [19] improves the texture generated using this method in agreement with the network to determine the restored output

instead of a pure translated network of image from an image.

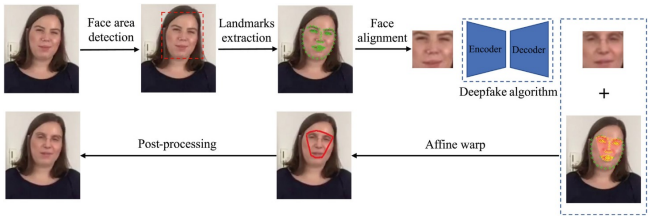

Fig. 10. Fake face Synthesis

The following table shows popular deepfake generation tools.

Fig. 11. Face Manipulation Methods

TABLE I
MANIPULATION METHODS

Methods	Deepfake Generation	Techniques used
Entire Face synthesis	This it creates non-existent faces	GAN (ex:StyleGAN)
Identity Swap	Replacing of One person's face with another person's face	FaceSwap, ZAO mobile application
Attribute Manipulation	Modification of Specific Features such as face editing, adding spectacles to a face image, and so on	GAN (ex:StarGAN)
Expression swap	modification of a person's face expression	Face2Face, Neural Textures
Picture Animation	Regardless of the initial image, it is animated into a talking head, which is quite similar to the driving video	Fake Image Animation
Image to image generation	Input image and target image will be provided to cGANs networks, and anticipated picture will be accomplished	cGANs (conditional GAN)

A. *Faceswap*

FaceSwap[9] is a method to move the facial features extracted from one video to another. The face region is extracted using sparsely identified facial markers. The method uses blend shapes to fit a 3D template model using these landmarks. The textures from the source images are used to back-project this model to the intended image, lowering the discrepancy between both the predicted shape and the localized landmarks. eventually, the image is mixed with the rendered model, and color-variance correction is applied. Until one video ends, we repeat these processes for all source and target frames in pairs. With a lesser cost of computation the execution can be run effectively on the CPU.

B. *DeepFakes*

Deepfakes[9] is a term that has come to refer to deep learning-based face substitute, but the same name is used for a deception technique that has expanded via online platforms too. A face from the origin video or photo collection is used to replace a face in a target image or video. The method uses a shared encoder to train two auto encoders to create images for training of the source and faces to be manipulated. A facial recognition system is used to crop and align the photos. The trained encoder and decoder of the source face are applied to the target face in order to create a fake image. The output of the auto encoder is then blended into the image.

C. *Face2Face*

Face2Face [9,28] is a facial reconstructive mechanism that pass on the manifestations of a source video to an intended video without compromising the target's identity. The two video input streams were used by the initial implementation with a manually done keyframe selection. A dense face reconstruction made use of these frames to reintegrate the face under various lighting conditions and facial expressions. We use the Face2Face technique to fully automate the creation of reenactment alterations for our video database.

D. *NeuralTextures*

For their Neural Textures-based rendering approach, Thies et al. [9] used face reenactment as an example. It learns a neural texture of the target individual, including a rendering network, from the original video data. Then the models can be trained by making use of the combined losses occurred by an adversarial network and while photometric reconstruction. Tracked geometry is employed during training and testing period in the Neural Textures method. These figures are generated using Face2Face's tracking module. Only the facial emotions corresponding to the mouth region are modified;while the region of eye remains intact.

IV. DATASETS

There are a number of datasets accessible online that may be used to track the advancement of deepfake technology. New deep learning models, such as GANs, are used in the most recent advances of artificial-based video synthesis systems. The GAN model is made up of two parts, both of which are deep neural networks that have been trained in tandem. The first is the generator network intends to generate face images that are as close to real ones as possible. The second is the discriminatory section of the network aims to differentiate between them, which results in synthesised face images that look very realistic.This is the basic idea behind geneartion of deepfakes. As a result, a new video is created with the target's faces superimposed on someone from the audience. As a result, a new video is produced beginning with the target's

TABLE II
DATASET USED

	Datasets	About
1	FaceForencic++	It is a forensics dataset made up of 1000 original video sequences which are modified using 4 different face modification techniques: Deepfakes, Face2Face, FaceSwap, and Neural-Textures. The input came from 977 YouTube videos, all of which included frontal face which was clearly visible, allowing tampering methods to create plausible forgerie
2	CelebA HQ	The CelebA-HQ dataset is an enhanced version of CelebA with 30,000 images at 1024x1024 pixels.
3	Flicker	They allow us to establish a standard for image localization of textual entity mentions.
4	UADF	This dataset contains 49 medium-quality real and fake videos.
5	VoxCeleb	It is free to download and install, and it is available worldwide. After many hours of work, the dataset was created by recording interviews with celebrities and well-known people on YouTube, one of the most popular websites. i)VoxCeleb1's database contains over 100,000 samples. VoxCeleb2 has nearly a million samples.
6	CelebDF	Celeb-DeepFake DF, the DF synthesis Algorithm is used to create the videos which is critical to improving visual quality. It is divided into sections that address various visual defects found in current datasets.

face of a specific individual as an input which is swapped with an individual from the source. Zhu et al. [29] proposed Cycle-GAN as a strategy for improving GAN performance. Bansal et al. [30] developed Recycle-GAN, which extends previous work by incorporating spatial and temporal signals via conditional generative adversarial networks. The two primary types of forensics datasets. First one is Traditional and second is Deepfake. Classical DeepFake forensics datasets are the two main types. Traditional forensics datasets are made by hand with extremely controlled environments like camera artifacts, splicing, in painting, resampling, and rotation detection.IFS-TC hosted the Very first Image Forensics Challenge (2013), an international event in which attendees pictured thousands of scenes both indoors and outdoors using 25 digital cameras. There are 82 occurrences of 92 forgery variants and 101 distinct mask splice detections in the Wild Web Dataset (WWD) [9]. The WWD tries to close the evaluation gap in photo manipulation localization techniques. [3] assesses the performance. The CelebFaces Attributes Collection (CelebA) includes over 200K celebrity photos and 40 attribute annotations. CelebA contains a total of 10,177 identities, 202,599 face photos, 5 landmark locations, and 40 binary attribute annotations per image, for a total of 10,177 identities, 202,599 face photos, and rich annotations.

DeepFake datasets are the second most common type of forensics dataset. GAN-based models, which are particularly famous because of their performance, are commonly used to generate these datasets. The UADFV [3] is made up of 49 real videos and 49 DeepFake videos created with FakeAPP and the DNN model. These films are about 11:14 seconds long on average, with a resolution of 294 x 500 pixels. The DeepFake-TIMIT (DF-TIMIT) dataset [3] was created by merging the VidTIMIT dataset [3] and FaceSwap-GAN; 16 similar-looking pairs of people from VidTIMIT [3] were chosen, and the database generated approximately 10 videos for each of the 32 people using low-quality of size 64 × 64, i.e., DF-TIMIT-(LQ), and high-quality of size 128 × 128.

FaceForensics (FF) [8] is a DeepFake dataset aimed at performing forensic tasks such as facial detection and segmentation of falsified images. Over 500,000 frames, it is made up of 1004 videos (facial videos taken from YouTube). Source-to-target manipulation, in which Face2Face reenacts the facial expressions of a source video, and self-reenactment manipulation, in which Face2Face reenacts the facial expressions of a source video, are the two types of manipulation.

The FaceForensics++ (FF++) [15] dataset has 1,000 actual YouTube videos were gathered, and 1,000 DeepFake videos were created using each of the four face methods: DeepFake, Face2Face, FaceSwap, and Neural Texture are some of the tools available.

Dataset	# Real		# DeepFake		Release Date
	Video	Frame	Video	Frame	
UADFV	49	17.3k	49	17.3k	2018.11
DF-TIMIT-LQ	320*	34.0k	320	34.0k	2018.12
DF-TIMIT-HQ			320	34.0k	
FF-DF	1,000	509.9k	1,000	509.9k	2019.01
DFD	363	315.4k	3,068	2,242.7k	2019.09
DFDC	1,131	488.4k	4,113	1,783.3k	2019.10
Celeb-DF	590	225.4k	**5,639**	2,116.8k	2019.11

Fig. 12. Basic Information of Various DeepFake Video Dataset

V. METHODOLOGY

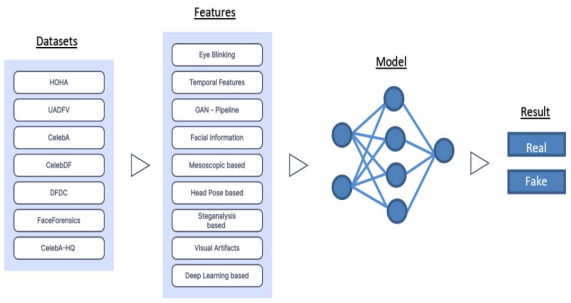

Fig. 13. Some important models and features used for detection

Methodology is a systematic, theoretical evaluation of the procedures utilized in an area of study. It involves a conceptual investigation of a collection of methods and principles related to a specific field of study. It typically includes terminology such as mode of thinking, theoretical model, stages, and quantitative and qualitative methodologies.

TABLE III
DEEP FAKE DETECTION METHODS

No	Title of the paper	Techniques used	Dataset used
1	Deepfake Video Detection Using Recurrent Neural Networks	employs LSTM and RNN, CNN for feature extraction, LSTM for sequence processing	HOHA.
2	Detection of DeepFakes in Videos Utilizing Feature engineering in Deep Learning CNN Frameworks	DWT, CNN+SIFT	False texts, voices, movies, and images are all examples of fraudulent media..
3	Deep Learning and Super Resolution Algorithms Combined for Deep Fake Detection	CNN Resnet50 Model with Super Resolution Algorithms	CelebA and UADFV
4	Deepfake Detection Using Clustering-based Embedding	Regularization Meso4 algorithm, FWA algorithm, EVA algorithm, Multitask algorithm, and the final Xception-c23	UADFV, CelebDF, and DeepFakeDetection algorithms
5	Deepfake Detection Using SVM	DeepFake, Image Processing, SVM, GAN, DFT	CelebA dataset
6	DeepFakes and Beyond: A Survey of Fake and Detection and Face Manipulation Databases	CNN, RNN	DFFD
7	Fighting deepfake with residual noise utilising CNN	InceptionResNet V2, CNN Deep Learning algorithm.	DFDC FaceForensics dataset
8	Deepfake creation and detection using Deep Learning	MesoNet CNN,	Deepfake Detection Challenge
9	Detecting CNN Generated Facial Images in Real-World Scenarios	Using Generative Adversarial Networks (GANs) CelebA-HQ (CAHQ)	Flickr-FacesHQ (FFHQ)
10	FaceForensics++: Attempting to Detect Manipulated Facial Images	Dete Using Steganalysis Method	FaceSwap, DeepFakes, Face2Face,Neural Textures,Post Processing Video Quality

These methodologies define the means or modes of data gathering, or, in some cases, how a specific outcome is to be calculated. Although great importance is given to the nature of the operations and types of processes to be followed in order to achieve a target, methodology does not define specific procedures.

[1] proposes a temporal-aware pipeline for detecting deepfake videos automatically. To extract frame-level features, it employs a convolutional neural network (CNN). These features are then used to train a recurrent neural network (RNN) to detect whether or not a video has been manipulated. The following are the primary contributions of this work: The analysis is presented in two stages, with a CNN extracting frame-level features and a temporally-aware RNN network capturing temporal discrepancies caused by the face-swapping procedure. 600 films were used to test the proposed method, with half of them being deepfakes obtained from various video hosting websites. The usefulness of the given approach is empirically demonstrated in a balanced setting, allowing it to determine whether a suspect video is genuine or not.

Due to the significant loss of frame data following video compression, most image detection methods cannot be employed for videos alone. Implementing an extraction of the videos into frames was a good technique to study this media. DWT is used for filtering in [2]: DWT divides an image into four components. DWT outputs images as Approximation, Horizontal Detail, Vertical Detail, and Diagonal Detail subband images called low-low (LL), low-high (LH), high-low (HL), and high-high (HH) when using the Python library PyWavelets (HH). Vertical Detail, often known as high-low, HL, is the specific output that will be used in this. This image decomposition can be thought of as a frame's underlying high frequency filter. The feeding dataset for the CNN will be the freshly filtered frames. A CNN is used for categorization. This will aid in the detection of a fault or abnormality within the Deep Fake. The collection of frames is separated into 90 percent training frames and 10 percent testing frames before being fed into the CNN model. This is a crucial phase in the training of a model. For improved accuracy, the training procedure is set to roughly 50 epochs. The training rate for the films is not the same; in fact, higher the resolution, longer will be the time taken to train the model.

In [5], a new method is proposed which includes clustering-based embedding regularization. To create films that can replicate unique artefacts in deepfake videos, open-source algorithms are used. A clustering-based embedding regularisation is incorporated into the classification objective to increase local smoothness of the representation space, resulting in a model that learns to withstand hostile cases. Three recent deepfake datasets are used to test the approach. The method's efficiency is demonstrated by the outcomes of the experiments. The Xception network is trained for categorization using positive and negative samples. The class number is set to 3 during the training phase and the generated samples are also classed as negative samples during testing process to increase the classification effect. During the training phase, a regularisation loss is also used to assure the embedding space's inter-class distance and intra-class smoothness.

[8] proposes a method that makes use of residual noise, which would be the difference between the initial image and its denoised form. Because of its distinct and discriminatory features, residual noise has been shown to be effective in detecting deepfakes, which can be successfully captured by convolutional neural networks with transfer learning. This technique is tested on two datasets: low resolution FaceForensics++ video sequences and high resolution videos from the Kaggle Deepfake Detection challenge (DFDC). The article proposes a classification strategy based on transfer learning utilising Convolutional Neural Networks to investigate the residual noise of actual and false videos. On two datasets, FaceForensics++ and DFDC, the suggested method's performance is demonstrated (DFDC). The purpose is to determine

whether the residual noise acquired from an authentic video differs from that obtained from false films, given a video V with frames F of h × w dimension. The denoised version of the frame is subtracted from the frame itself to extract the residual noise. The frames are denoised using the wavelet transform function WF. Each frame's residual noise is calculated. The backbone is the deep learning technique InceptionResNetv2, which is a 164-layer CNN trained on over a million photos from the ImageNet dataset.

In [14], a method for detecting faked movies of faces is proposed, with the technology being used at a mesoscopic level of analysis. Indeed, in a situation where a video after undergoing video compression possess a severe degradation of the picture noise,these microscopic investigations based on image noise become ineffective. Similarly, the human eye struggles to classify counterfeit images at a higher semantic level, especially when the image illustrate a human face. As a result, it recommends using a deep neural network with a adequate number of layers as an intermediary approach. With a minimal level of representation and a surprisingly low number of parameters, the two designs below have produced the more accurate scores of classification across all of the testing. They are based on image classification networks with well-performing convoluted layers and are pooled for feature extraction and are classified using a densely connected network. Meso-4: This network starts with four layers of sequential convolutions and pooling, then adds a dense network with one hidden layer. The fully connected layers employ dropout to regularise and improve their resilience, while the convoluted layers use ReLU functions of activation to induce non-linearities and Batch Normalization to regularise their output and prevent the vanishing gradient effect. This network has a total of 27,977 trainable parameters. MesoInception-4: A different structure is to use a variation of Szegedy et al inception's module to replace Meso4's first two convolutional layers. The module's goal is to stack the output of many convolutional layers with varied kernel shapes to expand the function space in which the model can be tuned. Instead of the 5 × 5 convolutions of the original module of Deepfake Detection which uses Neural Networks , a 3 × 3 dilated convolutions are used in order to avoid high semantic.This notion of using dilated convolutions with the inception module may be found in, but we've added 11 convolutions before the dilated convolutions for dimension reduction and an extra 11 convolution in parallel to operate as a skip-connection between succeeding modules.

The deepfake detection algorithms Xception and MobileNet are discussed in [15] as two stratergies for classifying tasks to identify deepfake movies automatically. FaceForensics++ training and evaluation datasets, which include four datasets developed using four distinct and famous deepfake technologies, are used. The outcomes demonstrate good accuracy across datasets, with applied accuracy ranging from 91 to 98 percent. A voting mechanism is also being created that can detect bogus videos by combining all four approaches rather than just one. Deep learning algorithms for implicitly classifying and therefore detecting deep fake videos are discussed in this research. FaceForensics++ was used asa source of raw video data, and this data was used to train two neural networks utilising pre-processed images: Xception and MobileNet. Each network is trained to provide four models, each of which corresponds to one among four of the most popular deepfake software platforms. Deepfakes, Face2Face, FaceSwap, and NeuralTextures are among them. The model's evaluations show that it can distinguish between real and fake films with a high level of precision However, this performance is largely dependent on the deep fake platform employed. To solve this, a voting process is proposed that takes advantage of the results of the multiple models to produce a more effective solution.

VI. Challenges

Lately, many DeepFake tools with incredibly realistic performance levels have become available, and many are still under development. On the contrary, the development of the DeepFake creation model is posing significant challenges to forensics and deep learning professionals in terms of combating it. GANs, a prominent AI approach, are made of two discriminative and generative models that compete to generate convincing fakes. These imitations of actual people are commonly quite famous and quickly spread all over social media sites, making it a great propaganda tool. As a matter of fact, a skilled attacker who masters the basic ideas of forensic tools can use a variety of counter-forensic methods to avoid detection. As a result, the detecting models must be able to figure out deceiving data and the real-world situations that decrease precision. As a result, numerous counter-forensics methodologies aimed at puzzling existing sensors are critical to the progress of multimedia forensics, as they expose loopholes in existing solutions and inspire further studies into more effective solutions. Many models exist to date for creating or detecting deepfakes, however they are still lacking in specifics and have limitations.

CHALLENGES IN DEEPFAKE CREATION

Despite major efforts to better the visual quality of generated DeepFakes, there are still a few obstacles to solve. Generalization, temporal coherence, lighting constraints, lack of realism in eyes and lips, hand movement behaviour, and identity leak age are some of the issues associated with making DeepFakes.

- **Generalization**:The properties of generative models are determined by the dataset used to train them. As a result, after completing training on a certain dataset, the model's output reflects the learned properties (fingerprint). Furthermore, the output quality is influenced by the amount of the dataset used during training. As a result, in order for the model to provide high-quality output, it must be given a dataset large enough as input to attain a specific sort of feature.
- **Temporal coherence**: Other flaws involve visible fluttering and juddering among frames, as well as a dearth of temporal coherence. Deepfake creation methods which

operate on every frame do not take into account temporal inconsistency which leads to several issues.

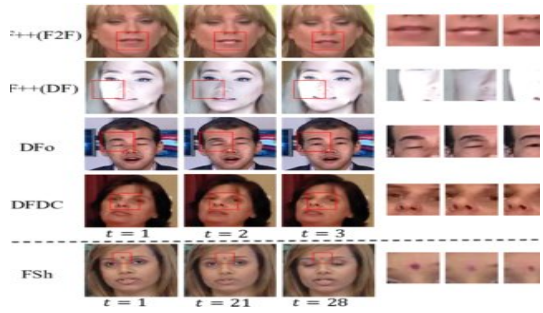

Fig. 14. Abnormalities of temporal coherence

- **Differences in illumination**: DeepFake datasets are created in a controlled environment with consistent lighting and background. In indoor/outdoor settings, however, a rapid change in lighting conditions results in colour inconsistencies and strange irregularities in the output.
- **Lack of realistic emotions**: The main challenges of DeepFake generation based on eye and lip synchronisation are an absence of emotions, disruptions, and the target's communication tempo.

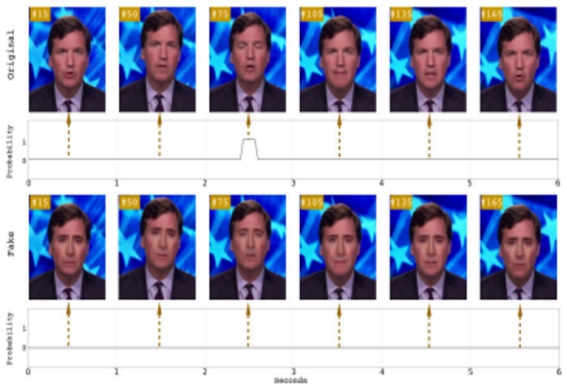

Fig. 15. Abnormalities of eye blinking

- **Hand movements**:DeepFake developing models struggle to match hand movements to emotions communicated through gestures when the target displays emotions through hand movements. Furthermore, because this type of expression dataset is limited, creating DeepFakes of this type is difficult.

CHALLENGES IN DEEP FAKE DETECTION

Although DeepFake detectors have improved dramatically in terms of performance, there are still a few flaws in the current detection algorithms that must be addressed. DeepFake detection systems face several difficulties out of which prominent ones are explained below.

- **Lack of DF datasets**: The effectiveness of a DF detection model is determined by the diversity of large datasets used during training. It is challenging to develop a model which can detect an unknown type of manipulation if it is tested on downloaded media that contains the unknown type of manipulation. Due to the prevalence of web-based platforms, post processing techniques are applied to DF multimedia in order to mislead the DeepFake detector. Such manipulation might include eliminating temporal artefacts, blurring, smoothing, cropping, and so on.
- **Unfamiliar attack type**:Another difficult task is developing a solid DF detection model against obscure sorts of assaults. These techniques are utilized to trick classifiers in their result.
- **Temporal Aggregation**: Current DF detection methods use binary frame-level classification to determine if a video frame is legitimate or fraudulent. These approaches, however, may encounter challenges like temporal anomalies and real/artificial frames appearing at regular intervals because they do not account for interframe temporal consistency.
- **Unlabeled data**:DeepFake detection methods are typically developed using massive datasets. At times, for example, in journalism or policing, only a limited number of dataset is available. Therefore, fostering a DeepFake detection model, unlabeled dataset is difficult.

VII. CONCLUSION

This paper takes a close look into DeepFake, a relatively new and This paper examines DeepFake, a relatively new and well-known method. It describes the DeepFake principles, GAN-based DeepFake applications, and their benefits and drawbacks. Models for DeepFake detection are also addressed. Most modern deep learning-based detection algorithms are incapable of transferring and generalising, implying that multimedia forensics is still in its infancy. Several well-known organisations and experts who work to advance applied approaches have expressed strong interest. However, maintaining data integrity requires significant effort, necessitating the implementation of additional security measures.

REFERENCES

[1] David Guera, Edward J. Delp. Deepfake Video Detection Using Recurrent Neural Networks.
[2] Sonya J. Burroughs, Balakrishna Gokaraju, Kaushik Roy and Luu Khoa. DeepFakes Detection in Videos using Feature Engineering Techniques in Deep Learning Convolution Neural Network Frameworks. 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR).
[3] Yuezun Li , Xin Yang , Pu Sun , Honggang Qi and Siwei Lyu Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. arXiv:1909.12962v4 [cs.CR] 16 Mar 2020.
[4] Nikita S. Ivanov, Anton V. Arzhskov, Vitaliy G. Ivanenko. Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection.
[5] Kui Zhu, Bin Wu and Bai Wang. Deepfake Detection with Clustering-based Embedding Regularization in 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC).
[6] Harsh Agarwal, Ankur Singh and Rajeswari D. Deepfake Detection Using SVM in Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems (ICESC-2021).
[7] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales and Javier Ortega-Garcia. DeepFakes and Beyond: A Survey of Face Manipulation and FakeDetection. arXiv:2001.00179v3 [cs.CV] 18 Jun 2020

[8] Marwa Chendeb El Rai ,Hussain Al Ahmad, Omar Gouda, Dina Jamal, Manar Abu Talib and Qassim Nasir. Fighting Deepfake By Residual Noise using Convolution Neural Networks in 2020 3rd International Conference on Signal Processing and Information Security(ICSPIS).

[9] Andreas Rossler, Davide Cozzolino, Justus Thies, Luisa Verdoliva, Matthias Nießner, Christian Riess. FaceForensics++: Learning to Detect Manipulated Facial Images in arXiv:1901.08971v3 [cs.CV] 26 Aug 2019.

[10] Nils Hulzebosch, Sarah Ibrahimi, Marcel Worring. Detecting CNN-Generated Facial Images in Real-World Scenarios.

[11] Hady A. Khalil, Shady A. Maged. Deepfakes Creation and Detection Using Deep Learning in 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC).

[12] Artem A. Maksutov, Viacheslav O. Morozov, Aleksander A. Lavrenov and Alexander S. Smirnov Methods of Deepfake Detection Based on Machine Learning.

[13] Luisa Verdoliva Media Forensics and DeepFakes: an overview in arXiv:2001.06564v1 [cs.CV] 18 Jan 2020.

[14] Darius Afchar, Vincent Nozick, Junichi Yamagishi and Isao Echizen. MesoNet: a Compact Facial Video Forgery Detection Network in arXiv:1809.00888v1 [cs.CV] 4 Sep 2018.

[15] Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang and Richard O. Sinnott. Deepfake Detection through Deep Learning in 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT).

[16] Michael Zollhofer, Justus Thies, Darek Bradley, Pablo ¨ Garrido, Thabo Beeler, Patrick Peerez, Marc Stamminger, ´ Matthias Nießner, and Christian Theobalt. State of the art on monocular 3d face reconstruction, tracking, and applications. Computer Graphics Forum, 37(2):523–550, 2018.

[17] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97, pages 353–360, 1997.

[18] Kevin Dale, Kalyan Sunkavalli, Micah K. Johnson, Daniel Vlasic, Wojciech Matusik, and Hanspeter Pfister. Video face replacement. ACM Trans. Graph., 30(6):130:1–130:10, Dec. 2011.

[19] Justus Thies, Michael Zollhofer, Matthias Nießner, Levi Val- ¨ gaerts, Marc Stamminger, and Christian Theobalt. Real-time expression transfer for facial reenactment. ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia 2015, 34(6):Art. No. 183, 2015.

[20] Justus Thies, Michael Zollhofer, Marc Stamminger, Chris- ¨ tian Theobalt, and Matthias Nießner. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016.

[21] Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation for image content changes. In IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[22] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Nießner, Patrick Perez, Chris- tian Richardt, Michael Zollhofer, and Christian Theobalt. Deep Video Portraits. ACM Transactions on Graphics 2018 (TOG), 2018.

[23] Zhihe Lu, Zhihang Li, Jie Cao, Ran He, and Zhenan Sun. Recent progress of face image synthesis. In IAPR Asian Conference on Pattern Recognition, 2017.

[24] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In IEEE International Conference on Image Processing, 2017.

[25] David Guera and Edward J. Delp. Deepfake video detection ¨ using recurrent neural networks. In IEEE International Conference on Advanced Video and Signal Based Surveillance, 2018.

[26] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In International Conference on Learning Representations, 2018.

[27] Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. Conditional cyclegan for attribute guided face image generation. In European Conference on Computer Vision, 2018.

[28] Justus Thies, Michael Zollhofer, Marc Stamminger, Chris- ¨ tian Theobalt, and Matthias Nießner. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016.

[29] Mohammed Akram Younus, Taha Mohammed Hasan. Effective and Fast DeepFake Detection Method Based on Haar Wavelet Transform in 2020 International Conference on Computer Science and Software Engineering (CSASE), Duhok, Kurdistan Region – Iraq.

[30] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-gan: Unsupervised video retargeting," in Proceedings of the European Conference on Computer Vision (ECCV), Washington, 2018, pp. 119-135.

[31] https://en.wikipedia.org/wiki

[32] A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.