Sri Lanka Institute of Information Technology

THE DEEPFAKE

Individual Assignment

IE2022 - Introduction to Cyber Security

Submitted by:

Student Registration Number	Student Name
IT[redacted]	GAMAGE G.G.I.V.M

2021.05.28
Date of submission

ABSTRACT

Deepfake images with masked faces has become a danger to Internet content. Detecting picture abnormalities, such as obvious artifacts or inconsistencies, is how most current anti-deepfake techniques work. Algorithms for deepfake may create fictitious pictures and films that seem to be hard to distinguish from genuine ones. As a result, the development of tools that can identify and analyze the integrity of digital visual material is critical. In recent months, a free software program based on machine learning has made it simple to make plausible face swaps in films that leave minimal indications of modification, dubbed "deepfake" films. Deepfake detection has been made more difficult by the ongoing evolution of video manipulation technologies and the enhancement of video quality

INSTRODUCTION

Fake news is a phenomenon that has been used for a long time. Notably, the concept predates the Internet, as publishers have long used inaccurate and deceptive facts to promote their own interests. Photographic files also became terribly standard artifacts as a Result of the widespread use of digital tools and, as a result, the growth of social networks. Deepfake was created by merging deep learning and the most effective technique for creating forged multimedia material, fake. Deepfakes are videos or pictures wherever the original person’s face is switched to some other person, for a malicious purpose. Deepfake is a concept that has gained a lot of traction and can be seen in a variety of situations. Deepfakes has image animation as well. A deepfake algorithmic rule can learn the small print of a person's face by feeding thousands of target pictures to a machine learning model. Such facial video forgeries are difficult to tell apart from the naked eye. An equivalent technique is used to train deepfake algorithms to imitate the tone, intonation, and other aspects of human speech. Deepfake technology can be thought of as a complex type of image writing software kit that makes it easy to modify images. Deepfake technology, on the other hand, goes even further in terms of how it manipulates visual and audio information. In the past, this kind of transformation needed extremely advanced video editing skills. Today, technology is accessible to anyone. It can, for example, create people that do not exist. Or that would make it seem as if actual people are speaking and doing things they did not hear. This kind of transition used to necessitate highly specialized video piece writing abilities. Technology is now available to everyone. Deepfake is a now-famous term that refers to new strategies for synthesizing or otherwise altering imagery, mostly faces in photographs, which is also the subject of this article. Fascinating false visual contents are being developed and transmitted at an astounding pace, thanks to rapid developments in computer vision and rapidly inexpensive and capable hardware. Deepfake images have recently been used to spread propaganda by showing public officials doing something they said, among other egregious and vulgar uses. As a result, deepfake identification is fast becoming a top priority for researchers, industry, and governments alike. They can be created for nefarious reasons, such as pornography recordings featuring actors, politicians, comedians, entertainers, and other public figures. Policy tensions, false news, and fake security footage, as a result, Facial video forgery identification has recently been a hot subject of discussion. When it comes to fake celebrity pornography videos or fake politician videos, deepfake videos are very dangerous. The study adds to the existing research on fake news and deepfakes by offering a thorough overview of deepfakes and establishing the emerging subject in a scholarly dialogue that also explores options for policymakers, journalists, entrepreneurs, and others to tackle deepfakes. [1]In such a perilous condition, a solution is also needed. A study has been conducted into the advancement of software capable of detecting Deep Fake created videos/animations and images. [2]

Anti-deepfake algorithms currently rely heavily on image or video anomalies such as visible objects or a lack of synchronization between lip expressions and spoken words to identify them. Many deepfake generators remove facial landmarks from videos during facial synthesis to exploit the facial areas of interest. *Figure 1* depicts few representations of the above facial artifact. Methods of facial modulation may be used on the whole face or only the parts needed for facial expressions. Following the manipulation of the targeted facial features, post-processing techniques such as resolution enhancement and color correction are used to make the distorted visualizations more accurate. Faceswap , Faceswap-GAN, Deep- FaceLab , and DFaker are the vital tools used to create videos that include face tampering. We show that fake videos can be created that are resistant to image and video compression codecs, making them a real-world problem because videos posted on social media are usually compacted. Worse, we show that in black-box conditions, where the opponent might not be aware of the classification model used by the detector, it is possible to create stable adversarial Deepfakes.

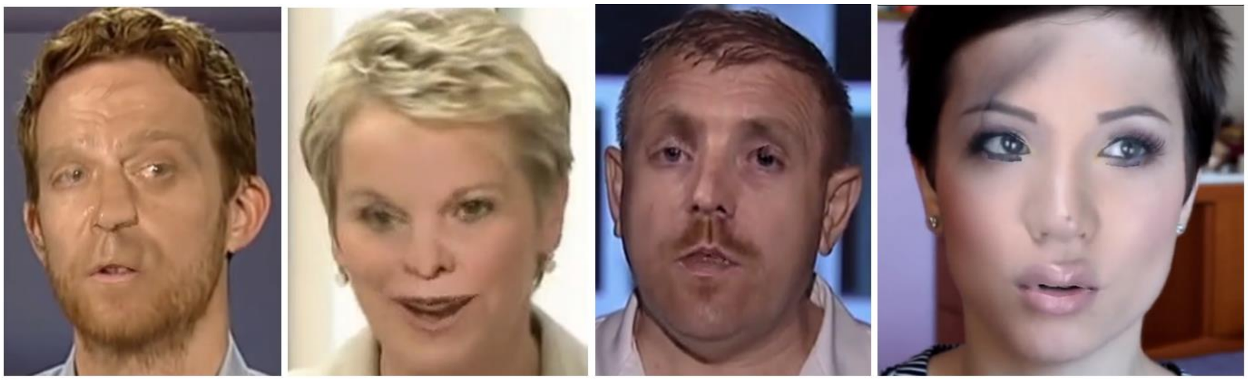


Figure 1. In the deepfake index, there are several examples of facial objects.

Though analog and digital impersonations are not recent, deepfakes to make advantage of powerful ML and artificial intelligence (AI) approaches to exploit or produce visual and audio material with a high capacity for deception. To fight Deepfakes' challenges, the machine learning group has suggested a number of countermeasures for detecting forgeries in digital media. [3]

Every day, millions of images and videos are posted to the internet, the majority of which have been manipulated using editing techniques. Most activities easily create Graphical User Interfaces

(GUI). Not only Individuals but also society are at risk because of the spread of forged recordings. Artificial Intelligence Created human-like characters will appear in the entertainment industry, who will not exist in real life but will look just like a regular human on screen and it will play roles as movie characters. Many of these developments are focused on leverage AI and machine learning breakthroughs to), Deep Learning, and Computer Vision. There is a lot of competition among researchers all over the world. In December of 2019, Facebook, Microsoft, and some of the world's most prestigious universities and partners have hosted the Deepfake Detection Challenge, which aims to encourage participants to make quick improvements in this field by encouraging them to develop novel methods for identifying and avoiding. It has a devastating impact not only on the person, but also on the society. [1] With the 2020 US election approaching, the media has expressed grave alarm over deepfake videos. People and communities as a whole are afraid how they can no longer trust what they can see online in the age of false news. [4]

***

A portrait of US President Abraham Lincoln from about 1865 is an early example of face swapping. Lincoln's head has been superimposed over an 1852 print of John Calhoun in the portrait. Late last year, an unidentified user going by the handle "deepfakes" posted pornographic videos on the popular website Reddit, saying they belonged to celebrities like Taylor Swift, Scarlett Johansson, Aubrey Plaza, Gal Gadot, and Maisie Williams [5]. While the pornographic videos were soon removed, this surprising deep learning-based facial replacement technique quickly attracted mainstream coverage and spread through many internet forums and subreddits. On February 7, 2018, almost all subreddits and web forums linked to the well-known "deepfaking" methodology were either deleted or barred. Multiple multimedia outlets, such as Discord, Gfycat, and Twitter, were also subject to the ban. When researchers at the University of Washington released a false video of former US President Barack Obama in July 2017, the public was alerted about the potentially destructive intervention of deep fake technologies [5]. Following that, in May 2018, a substandard deep fake video of President Donald Trump was posted to social media, urging Belgians to pull out of the Paris Climate Agreement. FakeApp, created by a Reddit user using an autoencoder-decoder pairing framework, is also the first effort at deepfake development. The autoencoder extracts feature vectors from face pictures, and the decoder reproduces the images in that method. Two encoder-decoder combinations are used to switch faces among origin and target photographs; each pair is used to practice on an image set, and the encoder's requirements are exchanged between network interface pairs. In other words, the encoder networks of two pairs are identical. This technique allows the typical encoder to find and learn the similarities between the two types of face pictures, which is relatively easy because faces have similar characteristics including pupils, noses, and lip positions.

In January 2020, Instagram and Facebook introduced a new policy prohibiting AI-manipulated "deepfake" videos that really are liable to mislead viewers in the run-up to the election. The problem is that this is dependent on the capacity to tell the difference between actual and false videos.

Fige 6 : **Barack Obama Deepfake Video**

Fig 2 : **Nicholas Cage Deepfake Visual**

Fig 2 and 3 shows screenshots of several well-known deepfakes, such as Barack Obama calling Trump a "dipshit" and Nicolas Cage's visage being switched into parts in films where he didn't even appear. These films attracted YouTube has millions of views, as both surprised and disturbed by how real they seemed. Because of the rapid advancement Deepfakes made up a large percentage of the audience, both academia and the technology sector have placed a significant emphasis on automatic identificaton of deepfake films, since there are more and more individuals use deepfakes to produce a variety of fake information, ranging from fake news to contents hoaxing, such as celebrity pornography. We utilized FaceForensics++ as the visual source of data and utilized this data to train two neural networks utilizing pre-processed images: Xception and MobileNet. Each network is trained to provide four models, each of which corresponds to one of four popular deepfake software products. Deepfakes, FaceSwap, Face2Face, and NeuralTextures are among them.

Deepfake is common for two reasons; the first one is because it can produce highly detailed measurement approaches, especially images, but even videos gave sufficient computation time and the second one is readily accessible, a layperson can easily access and use it. [5] FakeApp, a Reddit app that guides users through the basic steps of the deepfake algorithm, was launched. As a result, Getting today's technology - driven realm has grown easier. exceedingly important. It's much more difficult when working with deepfakes, since they're often used for nefarious reasons, and almost everyone can now build deepfakes using current deepfake software. Several techniques for detecting deepfakes have been proposed so far [6].

This demonstrated that technology is still changing and has the potential to deceive a considerable portion of the population. Image compression accuracy is improved using deep learning techniques. To construct compact representations of images, generative learning techniques and image compression autoencoders are used. Furthermore, autoencoders can extract a representation of compressed images while minimizing the loss function. Another method for creating a deepfake is to use two sets of encoders-decoders for the encoder network of separated loads [5]. Deepfake is made possible by finding a way to move both faces into the same encoder. This is conveniently achieved by having the same encoder sharing two distinct networks while using two different decoders at the same time. The potential of deep learning to describe advanced and high data is well-known [6].

Deepfakes have grown in popularity as a result of the high standard of tampered video clips and the ease of which their apps can be used by a diverse variety of users of varying computer abilities, from expert to beginner. Deep learning approaches are used to create the majority of these programs.

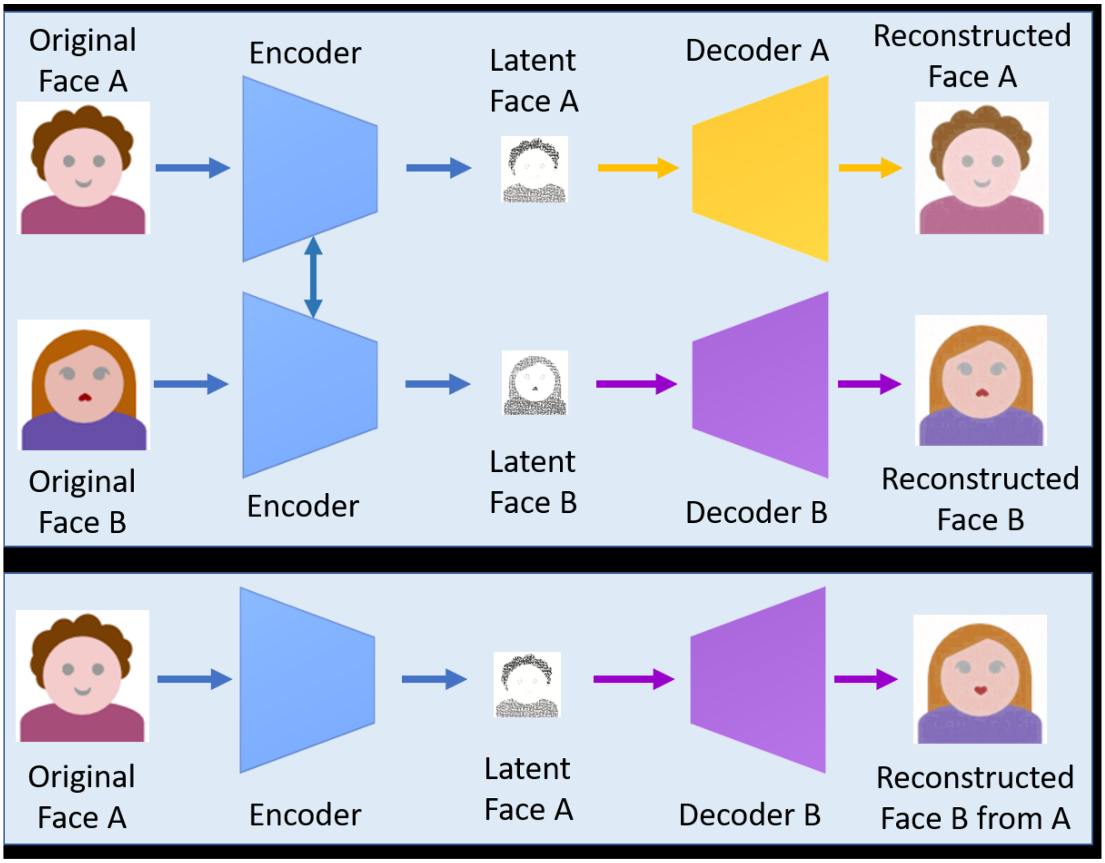


Figure 7 : Two encoder-decoder sets are used in this deepfake construction model.

For the training phase, two networks are using the encoder is still the same, however the converters are not. To make a deepfake, a picture Both of the same encoder and decoder B are used to encode and decode face A.. An enhanced a neural network-based variant of deepfakes was created by incorporating VGGFace provided adversarial loss and loss of sensation into the encoder-decoder structure (GAN). It was proposed in, i.e., faceswap-GAN. The VGGFace perceptual loss is used to make eye expressions more natural and compatible with input faces, as well as to smooth out segmentation mask glitches, resulting in output films of greater quality This approach allows for the development of resolutions in outcomes of 64x64, 128x128, and 256x256 pixels. [6] In fact, the FaceNet implementation introduces a Convolutional neural network (CNN) with several tasks to improve recognition of people's faces and alignment reliability. The CycleGAN is used to introduce generative networks. Table 1 summarizes the most common deepfake tools and their functions.

Deepfake Detection

As soon as the possibility of deepfakes was identified, methods for detecting them were suggested. Handmade characteristics obtained from items & defects in the phony visual estimation method were used in early efforts. Deep learning was used in recent methods to automatically remove salient and prejudicial characteristics in order to identify deepfakes. [6] Deepfake detection is typically thought of as a categorization that is either true or false challenge, with To differentiate among authentic and manipulated photos, classifications are utilized. To train classification models, this approach necessitates a huge archive of real and false photos. Despite the fact that the number of fraudulent videos is increasing, there are still limits in terms of developing a standard of certifying various identification algorithms.

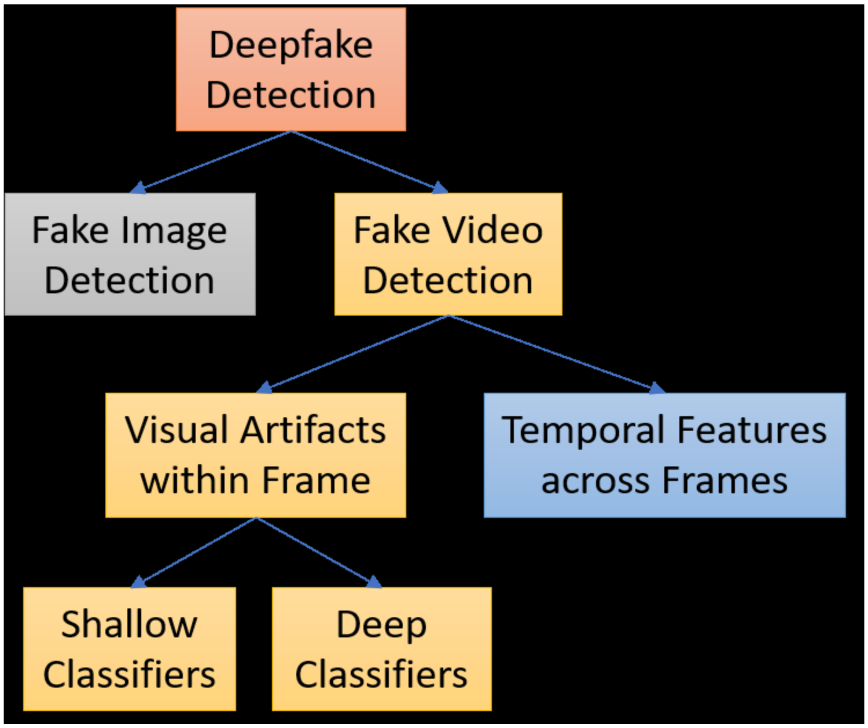


Figure 5

1.1 Detecting synthetic pictures

Head shifting has a wide range of applications in video editing, portrait transformation, most importantlyidentity theft prevention, because it may replace heads in photos with ones from either a photo collection. [6]However, a single of strategies used by Attempts by fraudsters to really get access to identifying or confirmation systems and obtain unauthorized entry. Because deep learning algorithms like CNN and GAN can maintain stance, facial expression, and lighting in pictures, swapped face photos have become more problematic for forensics models. To eliminate low level high frequency information from GAN pictures, use Gaussian blur and Gaussian noise. At the different scales, the forensic classifier must distinguish between real and fake photographs, forcing it to acquire more intrinsic and relevant properties, allowing it to generalize better than earlier image forensics approaches or image steganalysis networks. Agarwal and Varshney, on the other hand, framed GAN-based deepfake detection as a hypothesis testing issue, introducing a statistical framework based on the information-theoretic study of authenticity. The oracle error is defined as the minimal gap between genuine picture distributions and pictures generated by a certain GAN. The analytic results demonstrate that as the GAN is less precise, this distance rises, making deepfakes simpler to identify. In the case of high-resolution picture inputs, a very accurate GAN is necessary to produce difficult-to-detect false pictures. The first step is a feature extractor that uses the Siamese network design proposed in and is based on the common fake feature network (CFFN). Each unit in the CFFN is made up of many dense units.

To increase the representational capability of the false pictures, varied numbers of dense blocks were used. Depending on whether the validation data is face or generic pictures, the number of dense units is three or five, with the number of channels in each unit ranging from a few hundred to a few thousand. [6]The CFFN learning method extracts discriminative characteristics between false and genuine photos, i.e., pairwise information. These attributes are then sent into the second phase, which is a tiny CNN concatenated to the CFFN's final convolutional layer to detect misleading from real pictures. The suggested approach has been proven to detect both phony faces and false generic images. On the one hand, CelebA provided the face data set, which included 10,177 identities and 202,599 aligned face photos in diverse positions and backgrounds. Deep convolutional GAN (DCGAN), Wasserstein GAN (WGAN), WGAN with gradient penalty (WGAN-GP), least squares GAN, and progressive growth of GAN (PGGAN) are among the GAN variations used to produce false pictures with a 64x64 resolution. To validate the proposed technique, 385,198 training photos and 10,000 test photos, both genuine and fraudulent, were gathered. The broad data set, on the other hand, is taken from the ILSVRC12. To create false pictures with a size of 128x128, the large-scale GAN training model for high fidelity natural image synthesis (BIGGAN), self-attention GAN, and spectral normalization GAN are employed. The training set contains 600,000 false and genuine photos, whereas the test set has 10,000 of each type of picture. The suggested strategy outperforms competing approaches such as those described in the literature, according to experimental data.

1.2 Fake video detection

Because of the large loss of picture information as a result of visual compression, most image detection algorithms cannot be employed for movies [6]. This section examines deepfake video identification algorithms and divides dividing them in to the 2 categories: those that use tempory information and those that look at visual artifacts inside frames. Furthermore, videos include temporal features that vary between frames, making it difficult for systems built to identify merely still fraudulent pictures to detect them.

1.2.1 Visual Frames' Timing Features

According to the discovery that temporal coherence is not adequately maintained in the synthesis method of deepfakes, video manipulation is done frame by frame such that low-level errors caused by face manipulations present appear as periodic aberrations and frame-to-frame discrepancies.

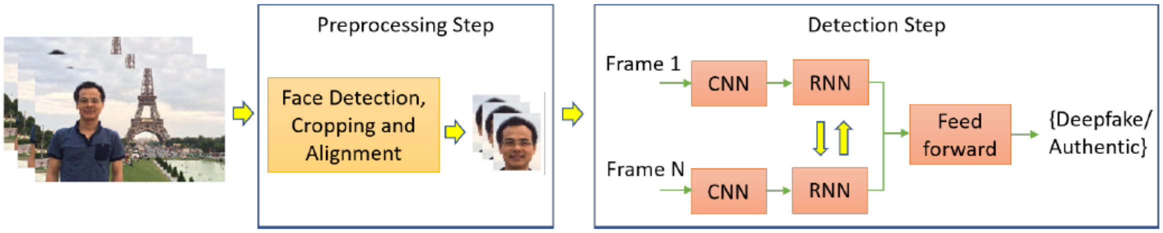


Figure 11

Figure 6: A two-stage procedure for detecting face manipulation, with the first stage detecting, cropping, and aligning the phase includes putting heads on a series of frames, and the third stage involves putting heads on a series of frames. using a convolutional neural network (CNN) and a recurrent neural network to discriminate between manipulated and legitimate face pictures (RNN).

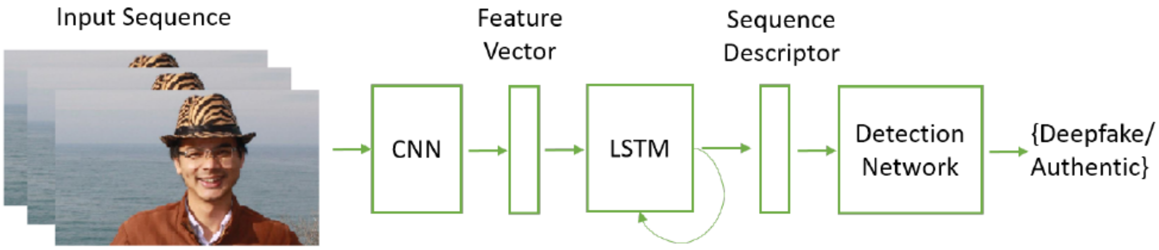


Figure 12

Figure 7. shows a deepfake detection approach that uses a To extract temporal information, a convolutional neural network (CNN) and a long short-term memory (LSTM) were used information from a sequence of visual and express them using a sequence descriptor. The sequence descriptor is utilized as a parameter, as well as the odds of the frame sequence being legitimate or deepfake are determined using just a detection network with convolution layers.

On the contrary, advocated using a physical indication, eye blinking, to detect deepfakes, based on the finding that a person in deepfakes blinks far less frequently than in untampered films. A healthy adult person blinks between 2 and 10 times per second, with each blink lasting between 0.1 and 0.4 seconds. Deepfake algorithms, on the other hand, frequently employ Internet face photos for training, which often show individuals with open eyes (very few photographs on the internet show persons with closed eyes). As a result, deepfake algorithms are unable to build fake faces that blink normally without access to photos of individuals blinking. In other words, deepfakes have substantially lower blinking rates than typical videos. The installation of LSTM aims to identify these temporal patterns efficiently since eye blinking has high temporal connections. A blink is defined as a peak over the threshold of 0.5 with a length of fewer than 7 frames, and the blinking rate is determined based on the prediction findings. This approach is tested using a web-based data set that includes 49 interview and presentation videos, as well as the deepfake algorithms' fake versions of those films. The experimental findings show that the suggested technique has a promising performance in identifying fraudulent videos, which may be further enhanced by considering the dynamic pattern of blinking. High-frequency blinking, for example, might be a symptom of manipulation.

1.2.2 Visual Artifacts within Video Frame

As mentioned in the preceding section, the approaches for detecting deepfake films that use temporal patterns across video frames are generally based on deep recurrent network models. This section looks at the alternative method of obtaining discriminant characteristics by decomposing films into frames and looking at visual artifacts inside single frames. To distinguish between fraudulent and legitimate films, these characteristics are distributed into either a deep or shallow classifier. As a result, the approaches in this area are divided into two categories: deep and shallow classifiers.

1.2.2.1 Deep classifiers.

Deepfake movies are typically made with low resolutions, necessitating an affine face warping strategy (i.e., scaling, rotation, and shearing) to match the originals' configuration.

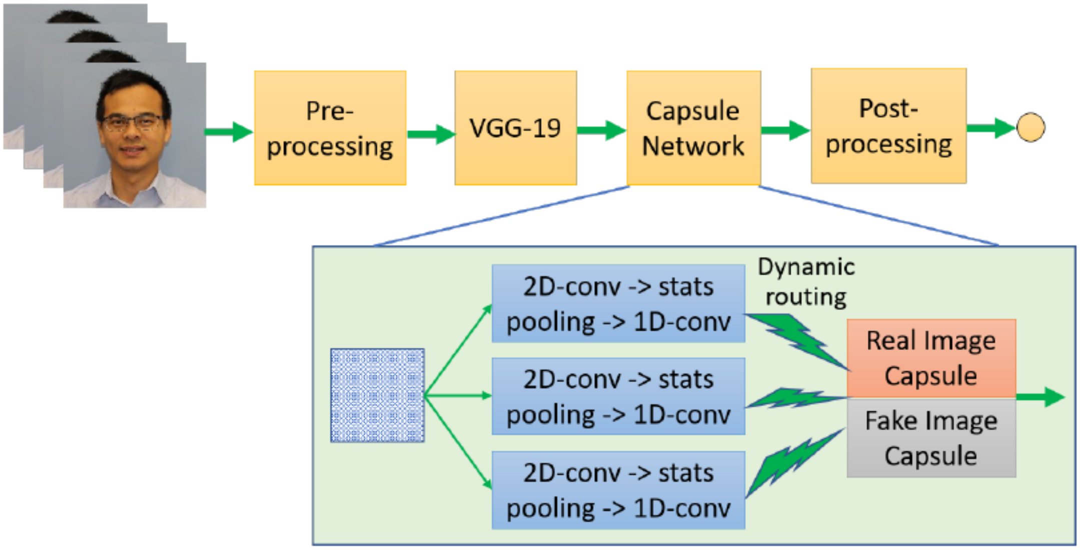


Figure 13

Figure 9: The VGG-19 network's properties are used by the capsule network to differentiate fraudulent photos or videos from authentic ones (top). Before VGG-19 is used to extract latent features for the capsule network, which consists of three primary capsules and two output capsules, one for genuine and one for false pictures, the pre-processing stage recognizes the face region and scales it to 128x128 pixels (bottom). The statistical pooling is an important component of the capsule network that detects forgeries.

1.2.2.2 Shallow classifiers

The majority of deepfake detection algorithms focus on artifacts or inconsistencies in inherent properties between fake and actual photos or videos. The 3D head postures are analyzed since the deepfake face generating pipeline has a flaw. To get the detection results, the retrieved features are input into an SVM classifier. Experiments on two data sets demonstrate the suggested approach's superior performance over competing approaches. The first data set, UADFV, has 49 deep fake videos and their actual counterparts. The second data set, that is a subset of the data used in the DARPA MediFor GAN Image/Video Challenge, has 241 genuine photos and 252 deep fake pictures. A strategy for exploiting deepfake and face alteration artifacts based on visual aspects like eyes, teeth, lips, and facial contours was also investigated.

A list of the most popular deepfake detection techniques

Methods	Classifiers / Techniques	Dealing with	Features	Data Sets Used
Eye blinking	LRCN	Videos	-Use LRCN to learn the spatial variations of eye blinking. -Based on the fact that deepfakes blink at a significantly lower frequency than usual.	Contains 49 interview and presentation videos, as well as the deepfakes that match to them.
Using spatiotemporal features	RCN	Videos	RCN, which combines the convolutional network DenseNet with the gated recurrent unit cells, is used to investigate temporal differences between frames.	1,000 movies are included in the FaceForensics++ data set.
Intra-frame and temporal inconsistencies	CNN and LSTM	Videos	CNN is used to extract frame-level features, which are then distributed to LSTM, which creates sequence descriptors that may be used for classification.	A compilation of 600 videos taken from numerous sources.
Using face warping artifacts	VGG16, ResNET50, 101 or 152	Videos	CNN models are used to detect artifacts based on resolution inconsistencies between the distorted face	- UADFV [83], which has 49 actual and 49 fraudulent videos totaling 32752 frames.

			region and the surrounding environment.	- DeepfakeTIMIT
MesoNet	CNN	Videos	- At the mesoscopic analysis level, two deep networks, Meso-4 and MesoInception-4, are introduced to analyse deepfake movies. - Deepfake and Face Forensics data sets have 98 percent accuracy.	Two data sets: deepfake, which is based on web videos, and FaceForensics, which is based on the Face2Face method.
Capsule forensics	Capsule networks	Videos/ Images	- The VGG-19 network extracts latent properties, which are then input into the capsule network for classification. - Through a series of rounds, a dynamic routing method is utilized to route the outputs of three convolutional capsules to two output capsules, one for false pictures and the other for genuine photos.	The replay-attack by the Idiap Research Institute, deepfake face swapping by , facial reenactment FaceForensics, and a totally computer-generated picture set employing are the four data sets.
Head poses	SVM	Videos/ Images	- The facial region's 68 markers are used	- UADFV is made up of 49 deep fake videos

			to extract features. - Using the retrieved features, classify with SVM.	and their actual counterparts. - DARPA MediFor GAN Image/Video Challenge: 241 actual pictures and 252 deep fake images
The roughness of the eye, the teach, and the face	Logistic regression and neural network	Videos	- Use face texture variances, as well as missing reflections and features in the eye and teeth areas of deepfakes, to your advantage. - For classification, logistic regression and neural networks are utilized.	A YouTube video data collection was obtained.
PRNU Analysis	PRNU	Videos	- Analysis of noise patterns caused by manufacturer flaws in light-sensitive sensors in digital cameras. - Investigate the changes in PRNU patterns between real and deepfake films, as face swapping is thought to change local PRNU patterns.	The authors used DeepFaceLab to create 10 real and 16 deepfake films.

Preprocessing Combined with deep network	DCGAN, WGAN-GP and PGGAN	Images	- Improve models' generalization ability to recognize GAN produced pictures. - Remove phony pictures' low-level attributes. - To increase generalization abilities, force deep networks to focus more on pixel level similarities between false and actual images.	- Real data set: CelebA-HQ, which includes high-resolution 1024x1024 facial photos. - Fake data sets: DCGAN, WGAN-GP, and PGGAN all produce fake data sets.
Bag of words and shallow classifiers	SVM, RF, MLP	Images	Using the speech to text approach, extract discriminant features and input them into SVM, RF, and MLP for binary classification: innocent vs manufactured.	The well-known LFW face database, contains 13,223 photos with resolution\s of 250x250.
Pairwise learning	CNN concatenated to CFFN	Images	Feature extraction with CFFN based on the Siamese network architecture, followed by classification with CNN.	- Face pictures: authentic CelebA pictures and false DCGAN, WGAN, WGAN-GP, least squares GAN, and PGGAN pictures -General pictures: actual ILSVRC12

				pictures and false ILSVRC12 pictures created using BIGGAN, self-attention GAN, and spectral normalization GAN.

Table 2

Whenever the recreated sound becomes loud, picture is consistent together with original's rumble pictures, iterative image building is ideal. The noise in the bogus videos isn't the same as it is in the actual ones. The distinction between denoised and source footage, known as residual noise, has been shown to be a discriminative characteristic in deepfake detection.

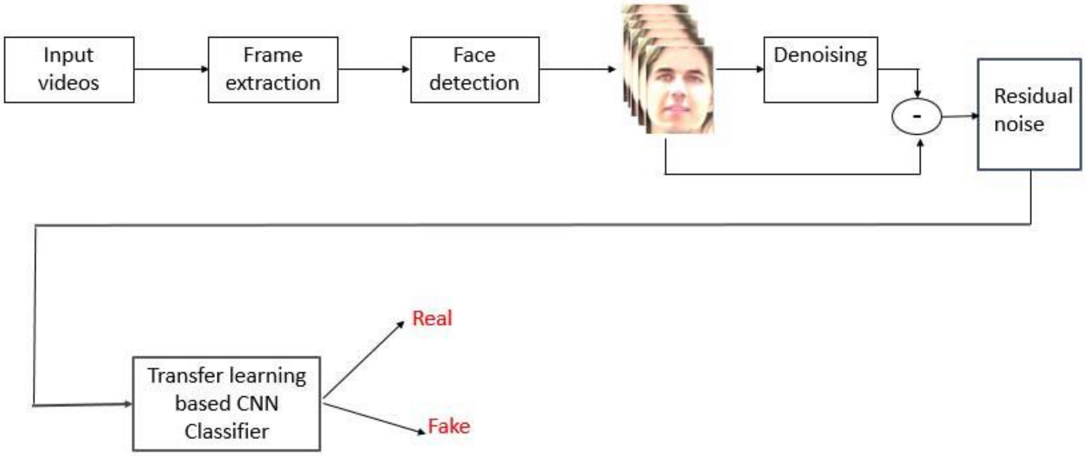


Figure 14 : The DeepFake detection process in action.

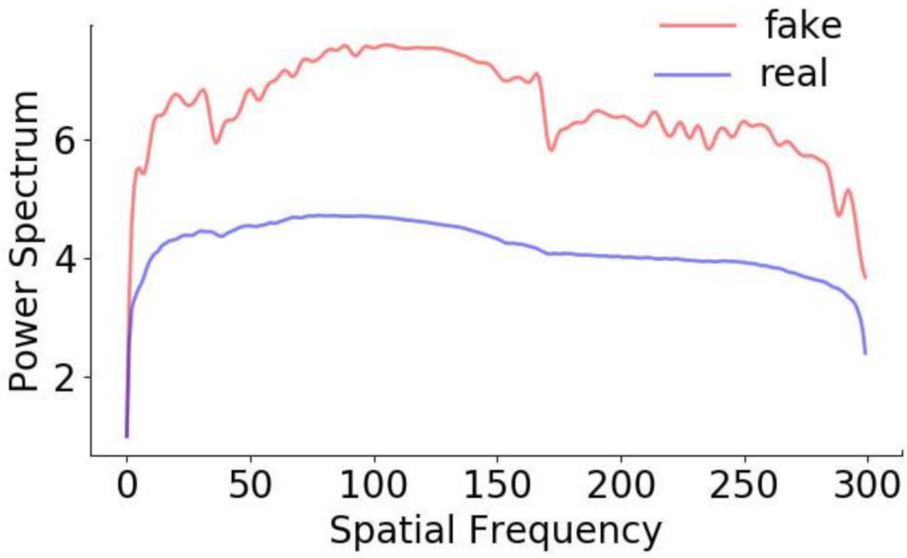


Figure 15 : The residual noise's 1D power spectrum. There is a distinct difference in behavior between actual and false frames.

Figure 1 illustrates the fundamental procedure. The processes in the data preparation procedure take images of movies then there are the faces from either the images. As mentioned in, extracting facial areas enhances detection accuracy. After that, the residual noise is calculated. As a pre-trained model, the deep network InceptionResNetV2 [7] is utilized for deepfake categorization.

2.1 Residual Noise

Given a video V of h * w dimensions and frames F. Our aim is to see if the sound that is still present extracted from an authentic video differs from that extracted from the false ones. The mean of 1000 residual noise power spectral is depicted in Figure 7.

The disparity between actual and false images encourages us to examine residual noise as a potential deepfake detection method. The denoised version of the frame is subtracted from the frame itself to extract the remaining noise. The frames are denoised using the wavelet transform function W_F. Each frame's residual noise is calculated as follows:

$$R_{N} = F- W_F(.) \qquad (1)$$

The magnitude of the 2D Discrete Fourier transform 2DFT is then calculated as follows:

$$R_{NF} = 2DFT|R_N| \qquad (2)$$

The 2D FFT of the image $|R_{NF}|$ is subjected to an adaptive Wiener filter as follows:

$$[\Theta_{xy}] = [R_{NF}]\ .\sigma^2{}_Z /\{ [Z_{RN}]_{xy} + \sigma^2{}_Z\} \qquad (3)$$

With

$$\sigma^2{}_Z= \delta.\sigma^2 R_N\ F \qquad (4)$$

Where x = 1,...,h, y= 1,...,w, Z_{RN} is the matrix containing the variance of R_{NF}'s energy, and $\sigma^2 R_N F$ is R_{NF}'s variance.

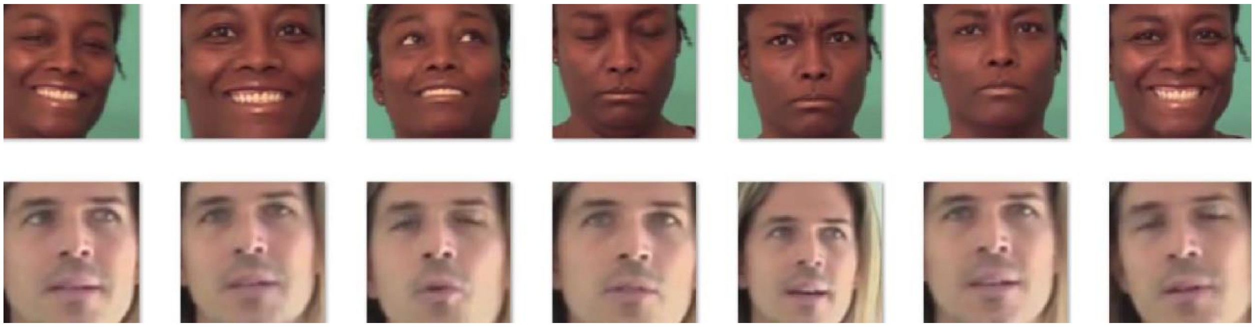


Figure 16 : ***From the DFCD Dataset, three examples of successive actual frames.***

Datasets: 1000 original films make up the FaceForensics dataset (Massive Video Dataset for Forgery Detection in Humans Face). Four face modification algorithms are used to create 4000 fake videos from the source footage: Deepfakes, Face2Face, FaceSwap, and neural textures. All original and synthetic Faceforensics videos have a pair of videos. The Deepfake Detection Challenge DFDC, which was launched in December 2019, is the second dataset. It's the initial subset of 100,000 fake and 20,000 actual films (size = 470GB), which includes a wide range of videos made specifically for deepfakes study [9]. Figure 8 show actual and false frames from the DFDC dataset in succession.

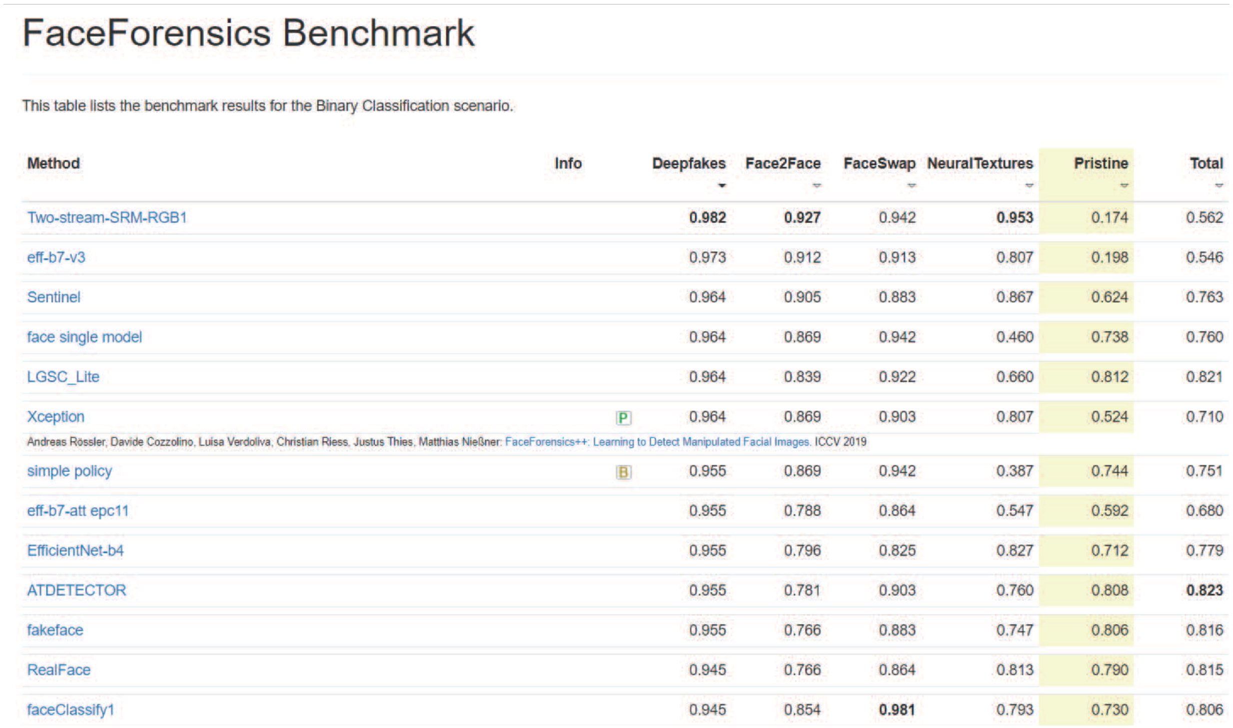
3.1 Deep Learning

There are several deep learning models and tools available today. To begin with, Xception has a strong performance when measured against the FaceForensics test environment. FaceForensics provides a platform for academics to put their training set to the test. [8]Figure 6 shows the performance of models trained by several teams, as well as the methodologies they used. Xception outperformed the other approaches on four distinct datasets and, more crucially, it is open source and comes with copious documentation, making model training and tuning easy. MobileNet was selected because its structure is comparable to that of Xception. They both use depthwise and pointwise convolution layer and are based on convolutional neural networks (CNNs). The only difference is that MobileNets contains less features to improve the model’s efficiency.

3.1.1Deep Learning in Face Synthesis

There are three types of synthesis.

- Face-Reenactment: When we transfer the expression of a target face to a source face, we receive a result video in which the source face appears with the expression of the target face. The source face and mouth expression are provided by the target face.
- Face-Swap: When we use Face-Swap, we preserve the source actor's expression while swapping his face with the targeted actor's. The source face receives face identity from the target face, as depicted.
- Face-Generation(video): Face-Generation is the process of creating a speaking video using a picture of one's face and audio data.



FaceForensics Benchmark

This table lists the benchmark results for the Binary Classification scenario.

Method	Info	Deepfakes	Face2Face	FaceSwap	NeuralTextures	Pristine	Total
Two-stream-SRM-RGB1		**0.982**	**0.927**	0.942	**0.953**	0.174	0.562
eff-b7-v3		0.973	0.912	0.913	0.807	0.198	0.546
Sentinel		0.964	0.905	0.883	0.867	0.624	0.763
face single model		0.964	0.869	0.942	0.460	0.738	0.760
LGSC_Lite		0.964	0.839	0.922	0.660	0.812	0.821
Xception	P	0.964	0.869	0.903	0.807	0.524	0.710
Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner: FaceForensics++: Learning to Detect Manipulated Facial Images. ICCV 2019							
simple policy	B	0.955	0.869	0.942	0.387	0.744	0.751
eff-b7-att epc11		0.955	0.788	0.864	0.547	0.592	0.680
EfficientNet-b4		0.955	0.796	0.825	0.827	0.712	0.779
ATDETECTOR		0.955	0.781	0.903	0.760	0.808	**0.823**
fakeface		0.955	0.766	0.883	0.747	0.806	0.816
RealFace		0.945	0.766	0.864	0.813	0.790	0.815
faceClassify1		0.945	0.854	**0.981**	0.793	0.730	0.806

Figure 17 : FaceForensics Benchmark

4.1 Xception

In 2017, Google researchers presented Xception, which stands for Extreme Inception. It uses Depthwise Separable Convolutions to create a deep convolutional neural network architecture. [8]The Depthwise Separable Convolutions deal with both spatial and depth dimensions (image width and height). By packaging the Xception model as a highlevel package with pre-trained weights, the Keras framework has made utilizing Xception simple. It's simple to use, stable, and quick to implement. This implies that rather than focusing on specific model development, the focus may be on reliable extracting features and identifying deepfake movies.

5.1 MobileNets

MobileNets are effective neural network topologies that are well-suited for mobile and embedded vision-based applications with little processing resources, as their name indicates. A depthwise separable convolution underpins this design. [4]Convolution that decomposes conventional convolutions into 1*1 pointwise convolution is referred to as depthwise separable convolution. The inputs are filtered and merged in a new set of outputs using standard convolutional layer. Filter functions based on the convolution kernel using standard convolution techniques, then combine these functions to generate a new representation. Separate layers for filtering and merging are used to partition the depthwise separable convolution into two layers. There are two levels to deep and separable convolution: depthwise convolution and pointwise convolution. To apply a single filter to each input channel, depthwise convolution is utilized (input depth). Then, using a basic 1*1 convolution, pointwise convolution creates a linear combination of depth layer output. MobileNets contains a total of 28 layers. Figure 7 shows a comparison between a regular convolutional layer and a depthwise convolutional layer. MobileNets does not, for example, deal with photos of the sides of heads. It also lowers picture distortion by restricting the size of pictures that may be used, as smaller models are less prone to overfitting. Other training parameters, however, stay constant regardless of model size.

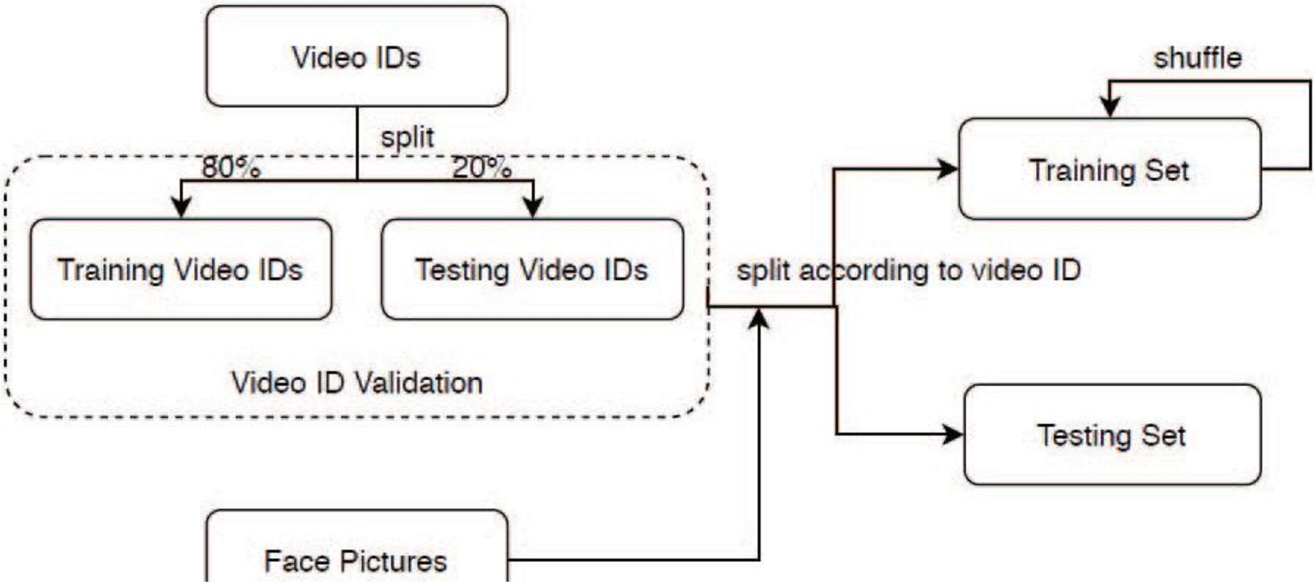


Figure 18 : Flowchart for Data Preprocessing

Based on the video ID of the image route, it was added to a list of training set pathways. For example, if 00132 is a video ID for training, the training routes list will include all pathways from this movie. The training routes list was randomized before each session. Instead of reading all of the photographs into memory, a generator was constructed to generate samples. When creating samples, each image was turned into a 299*299*3 numpy array. For false photographs (1) and non-fake photographs (2), labels were employed (0).

Twenty percent of the films from each dataset associated to the deepfake video producing algorithms, such as Deepfakes, Face2Face, FaceSwap, and NeuralTextures, were labeled and saved for the final assessment [13]. In addition, the model trained with the Deepfakes dataset was tested with the other three datasets to see if the model's detection performance was affected by the fake video generating method's overfitting. Instead of concentrating just on the model's correctness as a binary classification, the assessment was focused on the classification result's true positive rate (TPR) and true negative rate (TNR).

Real Video Detection Accuracy = TN/(TN+FN) * 100%

Fake Video Detection Accuracy = TP/(TP+FP) * 100%

Overall Detection Accuracy = (TN+TP)/(TP+TN+FP+FN) * 100%

This following formular were used to calculate the detection accuracy of actual and false video, as well as the total detection accuracy.

***

DeepFake detection strategies which will receive additional attention in the returning years. Other sorts of DeepFakes. though face swapping is currently the foremost wide renowned sort of DeepFake videos, it is by no suggests that the foremost effective. above all, for the purpose of impersonating somebody, face swapping DeepFake videos have many limitations. Psychological studies show that face recognition for the most part relied on data gleaned from face form and hairstyle. As such, to create the convincing impersonating result, the person whose face is to get replaced (the target) should have an identical face form and hairstyle to the person whose face is employed for swapping. Second, because the synthesized faces got to be spliced into the first video frame, the inconsistencies between the synthesized region and the remainder of the first frame may be severe and troublesome to hide. In these respects, the opposite sorts of DeepFake videos, namely, head puppetry and lip-syncing, square measure more practical and therefore ought to become the main target of ulterior analysis in DeepFake detection. strategies learning whole face synthesis or reenactment has experienced a quick development in recent years. though there have not been as several easy-to-use and ASCII text file computer code tools generating these styles of DeepFake videos as for the face-swapping videos, the continued the sophistication of the generation algorithms can amendment the situation future. because of the synthesized region is totally different from face swapping DeepFake videos (the whole face within the former and lip space within the latter), detection methods designed supported artifacts specific to face swapping are unlikely to be effective for these videos. Correspondingly, we should develop effective detection strategies to these styles of DeepFake videos. Audio DeepFakes. AI-based impersonation is not restricted to imagery, recent AI-synthesized content-generation square measure leading to the creation of extremely realistic audios. Using synthesized audios of the impersonating target will considerably make the DeepFake videos additional convincing and compounds its negative impact. As audio signals square measure 1D signals and have a awfully totally different nature from pictures and videos, different strategies got to be developed to specifically targeting such forgeries.

CONCLUSION

For widespread practical application, the overall running efficiency, detection accuracy, and, most crucially, the false positive rate must all be improved. The detection approaches must also be more resistant to post-processing stages in real life, as well as social media washing and counter-forensic technology. The forgeries makers and digital experts are always competing in terms of technology, know-how, and capabilities.

References

[1] R. H. B. P. N. B. C. C. F. B Dolhansky, "The Deepfake Detection Challenge (DFDC) Preview Dataset," C C Ferrer, 2019.

[2] Y. S. Malik, "DeepFakes and Detecting DeepFakes," IEEE, Lahore,Pakistan, 2021.

[3] P. N. M. J. F. K. J. M. Shehzeen Hussain, "Adversarial Deepfakes," IEEE, California San Diego, 2021.

[4] D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deepfake Detection through Deep Learning," IEEE, Leicester, UK, 2020.

[5] J. A. M. Albahar, "DEEPFAKES: THREATS AND COUNTERMEASURES," www.jatit.org, Saudi Arabia, 2019.

[6] Q. V. H. N. C. M. N. D. N. D. T. N. S. N. Thanh Thi Nguyen, "Deep Learning for Deepfakes Creation and Detection," IEEE, Leicester, UK, 2020.

[7] M. C. E. Rai, H. A. Ahmad, O. Gouda, D. Jamal and M. A. Talib, "FIGHTING DEEPFAKE BY RESIDUAL NOISE USING CONVOLUTIONAL NEURAL," IEEE, DUBAI, United Arab Emirates, 2020.

[8] D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deepfake Detection through Deep Learning," IEEE, Leicester, UK, 2020.